Young Researchers Seminar

2021

Seminar proceedings

University of Ljubljana
Faculty of Maritime Studies and Transport

15 – 17 September 2021

Portoroz, Slovenia

European Conference of Transport Research Institutes (ECTRI)
Forum of European National Highway Research Laboratories (FEHRL)
Forum of European Road Safety Research Institutes (FERSI)
And
University of Ljubljana, Faculty of maritime studies and transport

10th YOUNG RESEARCHERS SEMINAR

YRS 2021

SEMINAR PROCEEDINGS

15. – 17. September 2021

Portoroz, Slovenia

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 74498819
ISBN 978-961-7041-09-5 (PDF)

The papers are peer-reviewed by international experts
Publisher: Faculty of Maritime Studies and Transport, Portoroz, 2021
Editors: Marina Zanne, Patricija Bajec, Elen Twrdy
Electronic publication: https://yrs.sdzp.org/wp/wp-content/uploads/2021/09/YRS2021-Proceedings.pdf

INDEX (in alphabetical order by the family name of the first author)

DEFINING ENERGY (IN)EFFICIENCY OF LAST-MILE POSTAL DELIVERY: A REAL CASE STUDY IN THE MADRID REGION

Alessandra Boggio-Marzet, Researcher, PhD Student
Andres Monzon, Full Professor, PhD Supervisor
TRANSyT-Transport Research Centre, Universidad Politécnica de Madrid
C/Profesor Aranguren nº3, 28040 Madrid, Spain
alessandra.boggiomarzet@upm.es

ABSTRACT

Cities are undergoing a process of urbanisation and urban freight transport had to deal with the boom of e-commerce, which is accelerated by COVID-19 pandemic and produces a sharp increase in freight activities, bringing additional challenges to city logistics. Furthermore, the lack of real data concerning logistics companies makes the research of efficient solutions for urban logistics more complicated.

Based on results from a real case study, this paper identifies the main (in)efficiencies of last-mile postal deliveries(LMD) under real traffic conditions, in terms of both energy consumption and operative features. During one month, thirteen drivers of the same logistic company drove light duty vehicles for their normal deliveries. Meanwhile, GPS position, speed and other parameters were instantaneously recorded:242 delivery routes were alnalyzed in the city center and the peri-urban area of Madrid, corresponding to 7,262km travelled. After each delivery route drivers fulfilled a short questionnaire and leave a feedback to report any incidents encountered while delivering.

According to our results, clear inefficiencies dominate the current delivery performance: a large proportion of deliveries should be shifted to non-motorized transport modes especially in city center, where 40% of LMD is characterized by average speed lower than 10km/h. This aspect, coupled with the long waiting times necessary for delivering, the lack of parking space and traffic congestion, and the low percentage of first right time delivery reaching only 75%, emphasizes the urgency of implementing sustainable strategies of LMD (parcel lockers, UCC, cargo-bikes among others) to improve efficiency and reduce their negative environmental impact.

Keywords:
Last-mile delivery, real data, operative features, energy inefficiency

1. ENERGY EFFICIENCY IN TRANSPORT: AN UNFILLED GOAL

The great transformation we are experiencing within the current process of globalisation causes an increase in the mobility of people and goods at an astonishing rate. Indeed, in recent decades, different lifestyles and the development of new services have changed our habits and our mobility, and, as a consequence, have influenced the liveability of our cities, town centres and suburbs (Ministero delle Infrastrutture e dei Trasporti, 2020).

As known, the transport sector enables the movement of people, animals and goods from one place to another. Therefore, if on the one hand it is of foundamental importance for the prosperity of a country and its citizens, on the other hand it gives rise to side effects that can become costs when they relate to congestion, accidents, air pollution, noise, impact on climate change, fragmentation and so on (Ecotale, 2014).

According to the European Environment Agency, transport sector causes a quarter of global Greenhouse Gas (GHG) emissions in the European Union (EU) and contributes significantly to the climate change (EEA, 2019). While other economic sectors, such as households and industry, have drastically reduced their emissions compared to the 1990 level (Kyoto Protocol), transport emissions have increased, being 29% above 1990 levels in 2018. Moreover, within the sector, road transport is the main culprit, with higher growth than all other modes, both for passengers and freights (Eurostat 2020).

In Spain, transport sector is responsible for 24% of greenhouse gas emissions (in particular CO_2), and the Country is facing a major challenge to ensure its compliance with its international commitments (IDAE, 2018), aware that there is still a long way to go. The dominant mode is still road, and the effect of all the improvements - vehicles and fuels - have been offset by growth in demand and journeys: more trips and over longer distances have more than offset the limited improvements.

In this context, the EU has launched in 2019 the so-called "Green Deal". It proposes a Climate Pact, combining a growth strategy with the objective of preserving the planet for future generations. One of the key action elements is "accelerating the shift towards smart and sustainable mobility", which aims to reduce current transport emissions by 90% by 2050 (EC, 2019).

The objective of energy efficiency in road transport can be approached from different perspectives, both in terms of passengers and freight: shifting to other vehicle or fuel typologies, improving vehicles and fuels efficiencies, increasing occupancy, and making driving more efficient, both in terms of eco-driving or eco-routing. Over the years, it has been seen that strategies applied to urban logistics can be more or less efficient, but it is necessary to differentiate between different geographical, social and economic contexts, as well as by type of transport (goods or passengers) at the time of applying one strategy or another (Marcucci et al., 2017).

Figure 1 shows that a large proportion of road transport emissions are concentrated in urban areas: 25% of total emissions are directly urban, and a large proportion of interurban movements originates in or are destined for cities: movement of people and goods.

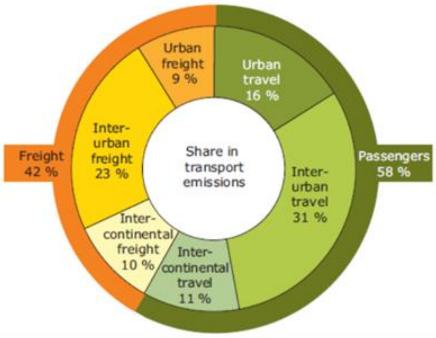


Figure 1: Share of transport emissions in the European Union (EEA, 2019)

According to Figure 1, 42% of transport emissions is caused by freight transport. Considering both urban and interurban freight transport, the share is equal to 32%, a figure that is noteworthy and requires proper reflection. Indeed, the inner nature of urban freight transport is characterized by frequent starts and stops; several studies have already demonstrated that vehicles pollute more in ignition, acceleration and braking (Garcia-Castro et al., 2018; Lois et al., 2019).

1.1. Transportation in changing society

The World Social Report 2020 affirms that for the first time in history, more people live now in urban than in rural areas (United Nations, 2020). More and more, indeed, there is a displacement trend of population from the countryside to the city. Today, 55% of the world's population lives in urban areas, a proportion that is expected to increase to 68% by 2050: the urbanization, combined with the overall growth of the world's population could add another 2.5 billion people to urban areas by 2050 (United Nations, 2020).

In this context, a paradigm shift will have occurred over the course of a century: the minor living environment has become the dominant one. Consequently, the transport system must keep pace with this urbanization trend; it must adapt and try to move towards the application of sustainable mobility strategies. In fact, although there are many environmental problems concentrated in cities, they are still the engine of the economy and the centre of business. Thus, efficient and well-planned freight transport may improve a city's functioning and enhance its competitiveness.

According to the United Nation Development Programme, cities account for 60 to 80 percent of energy consumption and at least 70 percent of carbon emissions, even though they occupy just 3% of the Earth's land (UNDP 2019). The evolution of urban mobility turns out to be a fundamental key to change the current reality.

1.2. Current trends in freight transport

Within urban mobility, Last-mile delivery (LMD) is defined as the movement of goods from a transport hub to the final destination. It is the last leg of the logistic chain directly to consumer, and it is usually one of the more expensive, least efficient and most polluting logistic chapters: it can reach 60% of the total logistics costs of a firm (Lafkihi et al., 2019). Indeed, in addition to the externalities proper of urban mobility, i.e. traffic congestion, occupancy of urban land, GHG emissions and noise, LMD problems are also related to the collection of goods and mails, and to those parcels that are subject to complaints or delivered to a wrong end customer.

Previous studies estimate urban freight transport counting up to 30% of total traffic emissions depending on the local context (Marcucci et al., 2017). In larger European urban areas, freight vehicles are responsible for half of particulate matter (PM) and a third of transport-related NO_X emissions (Macharis and Melo, 2011). Nonetheless, the urban freight transport system has to content with the current boom in e-commerce, which has dramatically increased the number of parcels delivered each day, and raised customer expectations, which now include not just fast but also free deliveries. In Spain alone, e-commerce exceeded 48.8 billion of euros in 2019, up almost 25% since the previous year, with a steady increase year-on-year (Comisión Nacional de los Mercados y la Competencia, 2020). Certainly, this increment further strengthened by the spread of the COVID-19 pandemic.

The current pandemic, in addition to its known effects on people's health, has had important social and cultural effects over society indeed, such as an increase in work from home, the redefinition of priority spaces, and the boom of digital channels among others. Specifically, the lockdown following the spread of the COVID-19 pandemic results in more and more products that, instead of being purchased by the consumer from the shop shelves, were placed in virtual shopping trolleys and will therefore have to be available in logistics centres and delivered directly to our homes.

1.3. Last-mile deliveries: challenges and features

According to Ragás Prat (2018), the “Boom” of e-commerce leads to two main consequences: first, the number of deliveries and the mobility of goods vehicles circulating in the city is multiplied. Second, any address becomes a potential delivery point. In other words, the urban distribution of goods is carried out in residential areas not designed to host this type of operation on a massive scale. In this context, the urban distribution of goods is being rapidly transformed and requires urgent responses from operators and administrations.

Several strategies and initiatives are promoted to increase energy and operative efficiency and thus mitigate negative impacts of LMD:

- vehicle-routing improvements (Musolino Et al., 2019; Croce et al., 2020),
- green and clean logistic fleets (Schliwa et al., 2015; Albergaria et al., 2018; Giordano et al., 2018),
- technological vehicle innovations (Alho et al., 2018; Taniguchi et al., 2020) and
- competitors’ cooperation (Savelsberg et al., 2015).

Additionally, several authors have identified trends and advances in technology leading to a new paradigm of urban logistics. According to Sevelsbergh and Woansel (2016), trends in population growth and urbanization, coupled with the desire for speed, the sharing economy and the rise of e-commerce, all increase the complexity of city logistics, and exacerbate its negative impacts on congestion, safety and environment. On the other hand, new and emerging technologies could drive innovation in city logistics and potentially decrease these effects. 5G technologies, drones and autonomous cargo vehicles are only few examples of insight into the future that have good potential for improving the efficiency and sustainability of urban freight system (Taniguchi et al., 2020).

However, those initiatives encounter various difficulties to implement in practice, such as real-time information is missing to provide dynamic vehicle route service; the investment costs are very high for carriers; and freight competitors are not willing to share valuable information to cooperate, since it is part of their business. It requires more feasible and cost-effective strategies or solutions to aid the implementation in urban freight deliveries.

Ranieri et al. (2018) grouped different last-mile solutions for freight transport according to four main aspects: innovative vehicles, proximity stations or points, collaborative and cooperative solutions and innovation in public policies and infrastructures. In the same line, Lafkihi et al. (2019) and Mangiaracina et al. (2019) conducted a comprehensive and extensive literature review of LMD challenges and opportunities and identified several effective last-mile solutions including night deliveries, which can shift traffic to off-peak hours, thereby reducing congestion by day and hence the related GHG emissions. According to Olsson et al. (2019), in order to optimize LMD operations, companies can act on several levels: (1) strategic level, such as location of distribution centres, (2) tactical level, such as fleet size and night deliveries, and (3) operational level, such as vehicle routing.

Literature pointed out the importance of realizing that geographic, economic, social and cultural circumstances affect city logistics and people's perception of critical issues related to the city logistics (Rai et al., 2021). Depending on the area of implementation and intensity, same freight initiatives may lead different results on reducing social or environmental impacts. Furthermore, most of the scientific publications about the theme they address transport policy with a theoretical approach without presenting real data; others simulate the solutions in a specific area, without taking into account different surroundings to apply. Real data concerning the operative features of logistic companies are missing, and the lack of real data makes the research of sustainable solutions for urban logistic more complicated.

In order to discover effective measure for urban freight transport (especially LMD), it is essential to properly characterize urban freight trips in both city centre and peri-urban delivery area, where have higher demand on goods movements, considering that it still lacks a full understanding of the issue. It is crucial to analyse and identify the main energy and operational inefficiencies of LMD to figure out which strategy shall be implemented, taking into account geographical, social and environmental contexts.

To our knowledge, there is a real scarcity of studies presenting an analysis of the inefficiencies of LMD based on a real field-test data. Thus, we performed a real data collection campaign to identify the main (in)efficiencies of last-mile postal deliveries (LMD) under real traffic conditions, in terms of both energy consumption and operative features.

After introducing the research context and the motivation of the paper, Section 2 presents the methodology used to carry out the investigation, including the data collection campaign and the data processing to compile the dataset. The analysis of the results and discussion are presented in Section 3. Finally, the main findings and policy recommendations are described in Section 4.

2. METHODOLOGY

This research analyses data collected in a field test of last-mile deliveries carried out by the countrywide Spanish logistic company, with the aim of characterizing the main patterns of current distribution operations, identifying the main operational inefficiencies, and specifying to what extent these operations vary in two different delivery areas (city centre and peri-urban area).

The methodology developed to carried out the research is composed by three main working packages, briefly described below.

First, we defined the case study, which consisted of a data collection campaign in real traffic conditions carried out in the Madrid Region, with the collaboration of the main Spanish Logistic Company. The field test involved 13 drivers from two distribution depots of the same company. They were responsible for completing their normal delivery route. Meanwhile, all vehicles taking part in the field test were monitored on a second-by-second basis to obtain instantaneous GPS position, speed, trip time and other engine parameters. After each delivery route, drivers were required to fullfil a short questionnaire in which they could report any incidences encountered.

A systematic data processing approach was then applied to integrate, filter and process all the recorded data, considering that a delivery route contains a certain number of parcel deliveries (micro-trips), whose characteristics vary according to the operating conditions. During the data processing, we thus defined two different databases: the first composed by delivery routes, and the second composed by each single parcel delivery micro-trips.

Finally, Once data processing has been completed, all data were analysed to deepening knowledge about the main inefficiencies of LMD. We developed a data analysis method, meaning the correct manipulation of items, to generate and address meaningful results regarding the research objectives settled (Section 3).

2.1. The case study

The case study takes place within the Metropolitan Area of Madrid. Madrid is the Main city of Spain, capital of the Country and of the same Region of Madrid. The associated metropolitan area counts with 6.6 millions of inhabitants, making it one of the most populous in the EU; of these, half of the population is concentrated in the city of Madrid (Monzón et al., 2017).

The Metropolitan Area of Madrid covers 8,030 km^2, with a ratio of urbanised surface of 11%. Three orbital motorways encircle the city (the M-30, M-40 and M-50) which are accessed from seven radial motorways. The terrain around the town is relatively flat, except for the mountain range, which is relative remoteness and borderline, located in the NW province.

The study is based on data derived from a real data collection campaign carried out in 2018 in cooperation with Correos, which is the Spanish national postal service (state-owned). It has 51,000 employees and distributes more than 3,600 million shipments per year. Within the Madrid Region, Correos owns 27 logistic depots that carry out distribution by Light Duty Vehicles (LDV): 15 depots are located within the city of Madrid and 12 depots are located outside the city. Two of these, one located in the city centre, and the other in the peri-urban area of Madrid, were involved in the field test,providing us with enough data for in depth analysis of the LMD in this region.

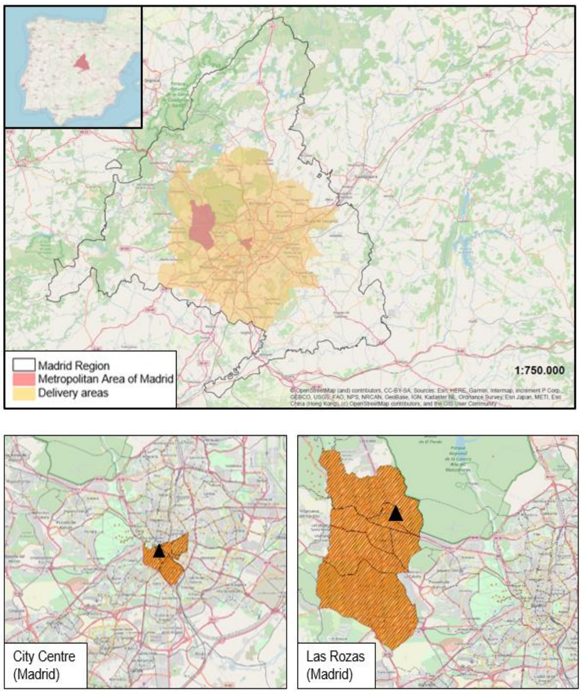


Figure 2: The Metropolitan Area of Madrid and the field-test delivery areas.

Figure 2 shows the main difference between the two delivery areas involved in the experiment, which are a key point to understand the operative features of LMD. One of the two depots provided by Correos is located in the municipality of Las Rozas, located in the NW fringe of the Metropolitan Area of Madrid, and it covers a peri-urban delivery area of over 130 km^2. The other depot is located in the Madrid city centre, and serves a delivery area ten times smaller (approx. 12 km^2), but with much higher population density (27 times) than Las Rozas.

The data collection campaign took place during four weeks. For one month, 13 professional drivers (11 males and 2 females, aged between 33 and 62) were involved in the field test. They drove electric and diesel LDV along their normal working routes, while GPS position, speed and other parameters were collected instantaneously each second. In total, 242 valid delivery-working periods (each one corresponding to a delivery route) of about four hours were recorded, covering 7,262 km of route. The routes were not predefined; rather, they were defined each day differently by the driver to optimize all deliveries during the working day.

In addition, after each delivery route, we collected feedback from each driver through a short questionnaire in which, among others items required, they could report any incidences encountered, such as traffic congestion or lack of parking. This information was very useful to understand real problems related to the distribution of parcels according to different delivery areas, based on real data, under real conditions, without simulating or hypothesizing.

2.2. Dataset definition

To perform the data analysis, it was necessary to define the dataset in two different dimensions, namely delivery routes and single parcel delivery micro-trips. Each delivery route contains a certain number of parcel deliveries (micro-trips) whose characteristics vary according to different operating conditions. Thus, we created two different databases: the first takes the form of delivery route, and describes the main aspects of last-mile postal delivery in terms of general operating features; the second allows the characterization of the driving patterns proper of single parcel delivery micro-trips and the understanding of energy inefficiency while delivering. Figure 3 shows the concept of delivery route ad single parcel delivery micro-trip just presented.

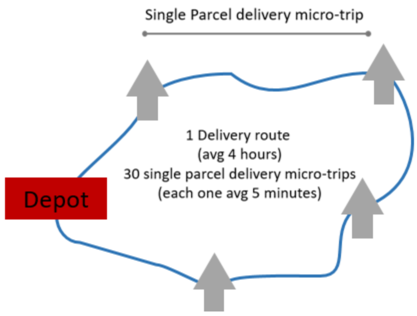


Figure 3: The concept of delivery route and single parcel delivery micro-trip.

We used all recorded data to manually define the delivery micro-trips within each delivery route, based on the instantaneous recordings. We selected one-second time intervals with null speed recorded, and if the consecutive time with null speed was greater than 2 minutes (± 30 seconds depending on the GPS positioning) we considered it as a delivery stop among subsequent micro-trips. Although this was not an absolute criterion, and we had to manually check with GPS data. We thus created two datasets designed as the basis for the analysis by programming in R, a free well-known software environment for statistical computing and graphics. The first database (Delivery route) comprises 242 delivery routes, each lasting about four hours of driving. The second database (Micro-trips) consists of 4,479 single parcel delivery micro-trips. This research considers 1,965 km of route covered in Madrid city centre and 5,297 km in the peri-urban area of Madrid.

The aim of the research is to identify the main inefficiencies of LMD in terms of both energy consumption and operative features. In the case of the vehicle's energy inefficiency analysis, we have considered the database composed by single parcel delivery micro-trips, considering several driving parameters that have been shown to have a significant influence on the energy consumption of the vehicle. For the case of operational inefficiencies, we rely on statistical analysis based on the two databases created, and on the feedbacks from drivers.

3. RESULTS AND DISCUSSION

We adopt statistical methods to characterize the operating performance of LMD, and to present the extent to which distribution operations vary for the same company in two delivery areas studied (city centre and peri-urban area), as well as their inefficiencies. Thus, we first defined the main operative features of LMD to improve knowledge on the delivery activity within the two areas considered. Then, we accordingly identified their main inefficiencies: both in terms of energy and operative features, both in city centre and in peri-urban delivery areas.

3.1. Energy inefficiencies of LMD

In order to have a concrete idea of the main aspects of delivery activities, Table 1 is presented. Distinguishing between peri-urban delivery area or city centre, Table 1 shows the average values, as well as minimum, maximum, and standard deviation, of the main operative features according to delivery area.

Table 1: Operative features according to delivery area

Operative features	Peri-urban area				Urban area			
	Avg	Sd	Max	Min	Avg	Sd	Max	Min
Daily trip distance [km]	50.93	17.75	93.67	17.38	14.34	4.58	55.44	5.64
Daily trip duration [hh:mm]	04:24	01:01	06:22	02:27	03:57	00:50	05.58	02:02
Micro-trip distance [km]	2.25	3.24	19.22	0.02	0.77	0.89	6.82	0.01
Median micro speed [km/h]	24.8	9.1	94.9	0	12.8	7.61	44.02	0
%Time micro-trip speed <3km/h	9%	8%	70%	0%	21%	13%	87%	0%

Despite similar delivery trip time of both depots implicated in the field test (average 4 hours), in city centre the average distance covered during a delivery route is 14.3 km, which is 3.5 times lower than in peri-urban delivery area. In city centre, 21% of the trip time is spent at speed below 3 km/h. Moreover, 50% of single parcel delivery micro-trips in the area is shorter than 477 meters and is performed with a median speed below 13 km/h.

The peri-urban case is quite different in terms of operating features. The average distance covered by each driver during the delivery route is 50.9 km, and the average distance to deliver a single parcel is 2.25 km (SD: 3.24), implying a rather disperse area. There, 50% of micro-trips are more than 1.1 km long. The median speed is average 25 km/h and the percentage of time at operating speeds below 3 km/h is 9%, less than half the percentage proper of the city centre.

Besides that, the operating speed of the vehicle is a decisive factor when it comes to define the energy efficiency of LMD. Figure 4 shows the instantaneous speed proper of a vehicle during two typical delivery routes in city centre and peri-urban delivery area.

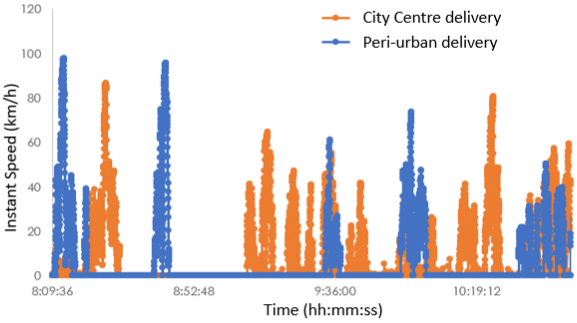


Figure 4: Instantaneous speed during two typical delivery routes

In city centre, drivers drove almost three time slower than in peri-urban area, and, on average, 20% of trip time is at stop state owing to road congestion, traffic light and out-door delivery. Nonetheless, the very low operating speed can be caused by several factors. According to our data, in city centre, between 16% and 27% of the trip time is spent at operating speeds of less than 3 km/h. In particular, 35% of delivery micro-trips in city centre have median speed of 7 km/h, and speeds of over 19 km/h is attained only 5% of the trip time. These deliveries could therefore be performed by other means of transport than duty vehicles, since the median distance for single deliveries is less than 480 m, a longitude that can easily be covered on foot or by bicycle, for example. In addition, the covariance of speed (meaning the speed variability while driving) during these deliveries is the highest recorded, up to 0.8 in most cases, highlighting a driving style characterized by high accelerations and decelerations which, partly due to the road network and traffic flows, increase fuel consumption and the resulting emissions (Lois et al., 2019). There, even if the reduced speed is maintained, the high level of covariance causes the consumption to increase markedly, and consequently the emissions. In order to reduce km travelled by LDV a sensible approach might be to implement an urban consolidation centre (UCC) from which different cargo bikes can delivery nearby without polluting. Indeed, it has been demonstred that the implementation of a UCC reduce up to 36% of the total emissions released to the environment (Firdausiya et al., 2019).

To make the concept clearer, Figure 5 shows the instantaneous positioning and operating speed of a vehicle during a typical city centre delivery route. The colours are consistent with the operating speed of the vehicle: red for speed lower than 10 km/h, yellow for speed between 10 km/h and 30 km/h and green for speed greater than 30 km/h.

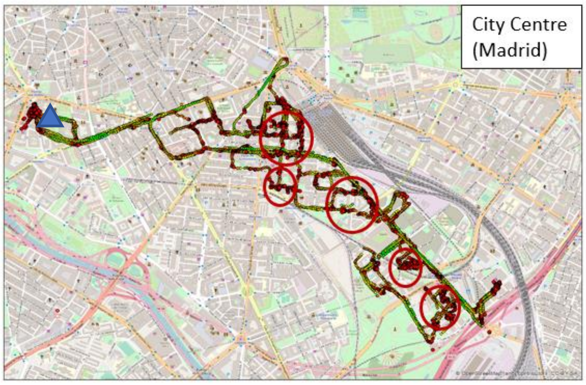


Figure 5: instantaneous speed and positioning of a delivery route in city centre.

As shown in Figure 5, in a delivery route there are some micro-zones within which the driver doesn't reach operating speeds higher than 10 km/h. In general, in 40% of cases, in city centre, the average speed of a single parcel delivery micro-trip is less than 10 km/h. This aspect should make us reflect on the real need to transfer LMD operations from vehicles to non-motorised mean, especially in city centre zones. Moreover, as shown in Figure 5, there are some streets covered several times during the same delivery route, due to the one-way streets in the city centre, as well as the lack of parking spaces. This aspect can in part justify the very low operating speed proper of the sample. Indeed the driver is obliged to go along the same street several times during different micro-trips of the same delivery route, because of the network layout of the area: one-way streets characterize the area, and in city centre it is quite typical the difficult of parking because of the surrounding traffic.

On the subject of the energy efficiency of vehicles, it has already been proven by literature that heavy accelerations and decelerations cause a considerable increase in vehicle's energy consumption (Garcia-Castro et al., 2018; Wang & Boggio-Marzet, 2018). Thus, we want to analyse the proximity of the stops that the driver makes during his journey. Apart from those stops caused by traffic, traffic lights and pedestrian crossings, the driver must adapt his driving to a sort of intermittency of situations characterised by frequent start and stop and consequently producing high GHG and pollutants emissions. Figure 6 shows the delivery points covered by two drivers, one responsible to deliver in the the peri-urban area and the other in the city centre, differentiating colours according to the day of delivery.

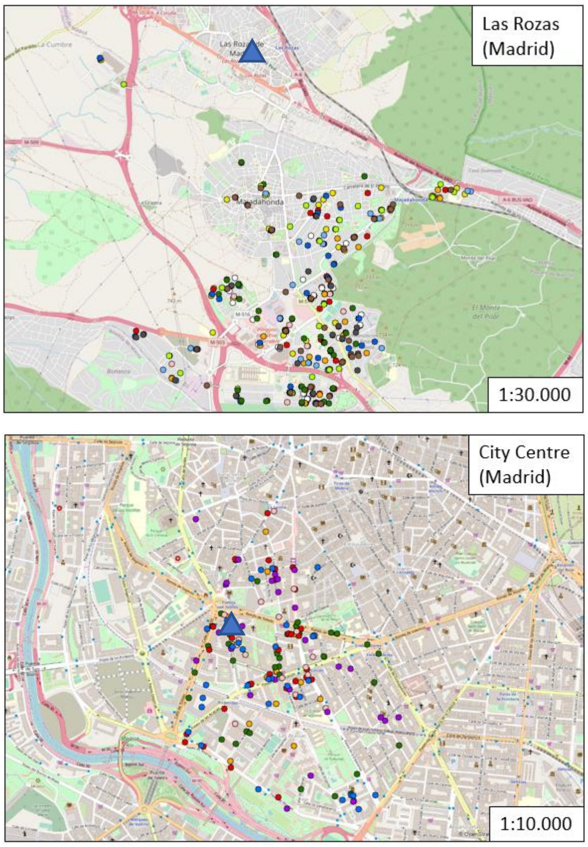


Figure 6: Delivery points covered by a single driver according to the day of delivery

From Figure 6, it emerges that the deliveries are almost the same from day to day. Indeed, the delivery route of a specific driver is quite homogeneous each day, since each driver is responsible of a certain zone of the corresponding delivery area. Minor variations are due to a single end customer located in a nearby road, in a particular building or in a nearby building, but in the same "micro-zone" for which each driver is responsible. In this context, it would be useful to perform a deep analysis of the geolocation of delivery points and determine the optimal location to install a set of parcel lockers, since 18% of deliveries in the peri-urban area were made in less than a minute and half, a few tens of meters apart. These deliveries refer to those made along broad roads flanked by large houses, one next to the other, with low traffic, where the driver can stop the vehicle right in front of the door, and deliver the parcel to the concierge, then repeat the process along the whole avenue. This process produces deliveries with an extremely high environmental impact. Given that vehicles pollute more in ignition, acceleration and braking, it would be advisable in these road contexts to implement another delivery system that is not based on LDVs.

3.2. Operative inefficiency of LMD

For this analysis, we also took into account the questionnaires that all drivers fulfilled after each delivery route. Thus, the main aspects with regard to operational inefficiency of LMD have been drown both from statistical analysis and, for the most part, directly from drivers.

The first inefficiency identified in both delivery areas is the long waiting times needed to carry out certain deliveries. Although the average delivery time is 5 minutes, there are some deliveries that require a waiting times of more than an hour. On average, the delivery is carried out in 04:17 minutes in the city centre, with standard deviation equal to 06:14, which means that around 70% of deliveries are made in less than 10 minutes. In the peri-urban area, the delivery is made in average 07:25 minutes with standard deviation equal to 09:41, which means that around 70% of deliveries are made in less than 15 minutes. Although, the maximum delivery time recorded during the data collection campaign is respectively 52 minutes in city centre and 96 in peri-urban setting, regardless of the driver. This is due to the fact that, the final customer may be an industrial estate or a large company with extensive premises, so the driver must be identified and undergoes several security checks before making the delivery, and then deliver to several buildings on the estate. Unfortunately, there is not a single driver responsible for all these kind of deliveries, and consequently there is no fixed route covering these delivery points; thus, operative efficiency is lost, increasing the percentage of long waiting times for all the drivers involved in LDV deliveries.

Another interesting aspect concerns those deliveries that fail, for example because the customer is not at home when the driver comes or because the address of the delivery point is wrong. If the driver does not find the customer at home, he/she must return a second time, at the end of his/her delivery route. If the customer continues to be absent, the driver transfers the package to the driver responsible of the next delivery route, who will have to change his route on purpose. This obviously implies a loss of time and reduce the operational efficiency of LMD, but it also generates a waste of energy consumption for the vehicle, as well as more kilometers are travelled, implying that more traffic is created, more pollutant are emitted etc.

The percentage of failed deliveries at the first intent is around 25% in both delivery areas, and it is reduced to 11% in the case of peri-urban deliveries and to 8% in the city centre by means of the further deliveries. It is interesting that the same poor result is reported in the work of van Duin et al. (2015). In fact, in their study they affirm that First right time delivery is just 75%. This phenomenon has an extraordinary effect on the environment in our cities, whether in terms of congestion, pollution, noise, etc. Surely, it would be interesting to create an instant geolocation system, even if, in this case, we would certainly be confronted with privacy issues.

Finally there are two main concerns from the feedbacks reported in the questionnaires by drivers: the lack of parking space and the congestion related problems. The lack of parking space is one of the main reasons resulting in longer trip distance and lower speed circulating (searching for free parking plot), consequently increases the use of energy and produces higher emissions and air pollutants. In this case, real-time parking information is required and other modes such as cargo motorcycles or cargo bycicles, which take up less space than cars should be considered to replace LDVs in future. It is worth pointing out that lack of parking space represents 29% of the problems detected while delivering in city centre, but, according to the feedback of drivers, it is not relevant in the peri-urban case: deliveries by foot in city centre should be promoted, as they would totally avoid the problem of lack of parking space.

The other key problem during LMD is the road congestion. Around 40% of delivery routes carried out report traffic jams, regardless of the delivery area considered. This aspect should be taken into account by policymakers, since the deep nature of LMD is characterized by short length, so that traffic jams can have a much greater influence on the operative efficiency of LMD than in long distance deliveries. Another time it seems an optimal solution the hypothesis of introducing a fleet of cargo motorcycles, that occupy less space than the LDV, contributing to reduce the loss of time caused by search of parking and traffic congestion.

4. CONCLUSIONS AND POLICY RECOMMENDATIONS

Nowadays, cities are undergoing a considerable process of urbanisation, and urban freights transport has to deal with the current boom in e-commerce, which produces a sharp increase in freight transport activity and brings additional challenges to city logistics. Through this research, we characterized the main patterns of last-mile postal deliveries in terms of energy consumption and operative features, both in city centre and peri-urban delivery area. With the purpose of changing the current emissions trends produced by last-mile delivery over cities and citizens, we thus aimed to study in quantitative terms, which are the most evident inefficiencies while delivering within different delivery areas of the same Madrid Region.

To this aim, 242 delivery routes were analysed, corresponding to 4,479 single parcel delivery micro-trips and covering a total of 7,262 km of route: 1,965 km in Madrid city centre and 5,297 km in the peri-urban area of Madrid. The main findings emerged from our results are the following:

- There are clear inefficiencies in the current delivery performance, in terms of both energy consumption and operative features: a large proportion of deliveries currently made by LDV should be shifted to non-motorised transport modes, especially in city centre. Concretely, 40% of deliveries in city centre are characterized by average speed lower than 10 km/h, and, being shorter than 400 m, these deliveries could be easily covered by foot, cargobike or cargo motorcycle.
- The road network and the population density are quite etherogenous within the same delivery area, mainly in peri-urban area: the distance between subsequent deliveries can be of some tens of meters along certain roads, or it can be of more kilometers in other micro-zones. Sustainable LMD strategies according to micro-zones should be implemented.
- The lack of parking space, as well as the failed deliveries (25% at the first time) and the frequent traffic jams imply a loss of time for the driver and reduce the operational efficiency of LMD; they also generates a waste of energy consumption for the vehicle, as well as more kilometers are travelled. Thus, an interesting solution is the implementation of a urban consolidation centre (UCC) from which deliveries could be done by cargo motorcycle or cargo bike which occupy less space than cars and do not imply a loss of time searching for parking, nor do they contribute to increase traffic congestion and pollutants.
- From an energetic point of view, it would be useful to perform a deep analysis of the geolocation of delivery points and determine the optimal location to install a set of parcel lockers in those kind of roads characterized by high population density, since most of the time the delivery points reached by drivers are almost the same, day by day. Moreover, the proximity of the delivery points one to each other results in additional fuel wastage due to the frequent start end stop of the vehicle added to those produced by the road context, i.e. due to traffic jams, traffic lights or pedestrian crossings.
- In terms of operational efficiency, there is a clear lack of homogeneity with regard to the time needed for delivery. Although on average it takes 5 minutes to make a delivery, it can take up to one hour. It could be useful to define a fixed route covering deliveries to large companies, avoiding that all drivers have to experience the same long waiting times.

Through this research, we pointed out that most of the operating inefficiencies of last-mile deliveries are similar regardless of the delivery area, both in terms of operative features and energy consumption. Nonetheless, it emerges that within the same delivery area, the optimal solution could vary according to the micro-zone.

Due to the large amount of data analysed, this research provides a broader view for transport planners and policymakers in terms of LMD management, operation and usage to design emissions reduction strategies for LMD. Also, the results and conclusions emerged in this study shall be implemented into the SUMPs – sustainable urban mobility plans - into the parts addressing city logistcs.

Further research should expand the study area and focus on measuring and quantifying the impact of different last-mile solutions according to socio-demographic circumstances. Furthermore, it wouold be interesting to investigate the final consumer perspective, both in terms of usage of the urban freight delivery service, and in terms of acceptability towards certain solutions (for example, it is known that in order to reduce traffic impacts, freight delivery could be carried out at night, but this would imply an increase in the noise produced by vehicles).

Acknowledgements

The authors would like to warmly thank the willingness shown by Dr. Ing. Libor Krejcì in providing valuable suggestions and ideas that contributed to the improvement of the paper.

REFERENCES

[1] Alho, A. R., e Silva, J. D. A., de Sousa, J. P., & Blanco, E. (2018). Improving mobility by optimizing the number, location and usage of loading/unloading bays for urban freight vehicles. Transportation Research Part D: Transport and Environment, 61, 3-18.

[2] Comisión Nacional de los Mercados y la Competencia (2020). http://data.cnmc.es/datagraph/ Accessed 10 February 2021

[3] Croce, A. I., Musolino, G., Rindone, C., & Vitetta, A. (2020). Route and Path Choices of Freight Vehicles: A Case Study with Floating Car Data. Sustainability, 12(20), 8557.

[4] de Mello Bandeira, R. A., Goes, G. V., Gonçalves, D. N. S., Márcio de Almeida, D. A., & de Oliveira, C. M. (2019). Electric vehicles in the last mile of urban freight transportation: A sustainability assessment of postal deliveries in Rio de Janeiro-Brazil. Transportation Research Part D: Transport and Environment, 67, 491-502.

[5] Ecotale, 2014. External costs of transport and land equalization. http://www.ecotale.eu/files/it/ecotale_it.pdf

[6] EEA, European Environment Agency, “Annual European Union greenhouse gas inventory 1990-2017 and inventory report 2019” EEA/PUBL/2019/051. Disponible en: https://www.eea.europa.es/publications/european-union-greenhouse-gas-inventory-2019

[7] EEA, European Environment Agency, “National action across all sectors needed to reach greenhouse gas Effort Sharing targets”, 2019. Disponible en: https://www.eea.europa.eu/publications/national-action-across-all-sectors

[8] European Commission (2019). The European Green Deal. Disponible online en: https://ec.europa.eu/info/sites/info/files/european-green-deal-communication_en.pdf

[9] EUROSTAT, 2020. Disponible online en: https://ec.europa.eu/eurostat/statistics-explained/pdfscache/1180.pdf

[10] Firdausiyah, N., Taniguchi, E., & Qureshi, A. G. (2019). Impacts of Urban Consolidation Centres for Sustainable City Logistics Using Adaptive Dynamic Programming Based Multi-Agent Simulation. IOP Conference Series: Earth and Environmental Science (Vol. 328, No. 1, p. 012071). IOP Publishing.

[11] Garcia-Castro, A., Monzon, A., Valdes, C., & Romana, M. (2017). Modeling different penetration rates of eco-driving in urban areas: Impacts on traffic flow and emissions. International Journal of Sustainable Transportation, 11(4), 282-294.

[12] Giordano, A., Fischbeck, P., & Matthews, H. S. (2018). Environmental and economic comparison of diesel and battery electric delivery vans to inform city logistics fleet replacement strategies. Transportation Research Part D: Transport and Environment, 64, 216-229.

[13] IDAE. Second Action Plan of Spanish Strategy of Energy Saving and Efficiency 2011–2020. Madrid, 2011. Available online: https://www.idae.es/uploads/documentos/documentos_11905_PAEE_2011_2020._A2011_A_a1e6383b.pdf

[14] Lafkihi, M., Pan, S., & Ballot, E. (2019). Freight transportation service procurement: A literature review and future research opportunities in omnichannel E-commerce. Transportation Research Part E: Logistics and Transportation Review, 125, 348-365.

[15] Lois, D., Wang, Y., Boggio-Marzet, A., & Monzon, A. (2019). Multivariate analysis of fuel consumption related to eco-driving: Interaction of driving patterns and external factors. Transportation Research Part D: Transport and Environment, 72, 232-242.

[16] Macharis, C., & Melo, S. (2011). Introduction-city distribution: challenges for cities and researchers. City distribution and urban freight transport: multiples perspectives, Edward Elgar Publishing, Northampton, 1-9.

[17] Mangiaracina, R., Perego, A., Seghezzi, A., & Tumino, A. (2019). Innovative solutions to increase last-mile delivery efficiency in B2C e-commerce: a literature review. International Journal of Physical Distribution & Logistics Management.

[18] Marcucci, E., Le Pira, M., Gatta, V., Inturri, G., Ignaccolo, M., & Pluchino, A. (2017). Simulating participatory urban freight transport policy-making: Accounting for heterogeneous stakeholders’ preferences and interaction effects. Transportation Research Part E: Logistics and Transportation Review, 103, 69-86.

[19] Ministero delle Infrastrutture e dei Trasporti (2020). La logistica urbana in una visione integrata. https://mit.gov.it/sites/default/files/media/pubblicazioni/2021-01/La%20logistica%20urbana%20-%20n.1%20dicembre%202020.pdf

[20] Monzón, A., Cascajo, R., Pieren, G., Romero, C., Delso, J.: Informe del Observatorio de la Movilidad Metropolitana 2015. (2017). Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente.

[21] Musolino, G., Rindone, C., Polimeni, A., & Vitetta, A. (2019). Planning urban distribution center location with variable restocking demand scenarios: General methodology and testing in a medium-size town. Transport policy, 80, 157-166.

[22] Olsson, J., Hellström, D., & Pålsson, H. (2019). Framework of last mile logistics research: A systematic review of the literature. Sustainability, 11(24), 7131.

[23] Ragàs Prat, I.: Logística urbana Manual para operadores logísticos y administraciones públicas. Marge books (2018).

[24] Rai, Heleen Buldeo, Sara Verlinde, and Cathy Macharis. "Who is interested in a crowdsourced last mile? A segmentation of attitudinal profiles." Travel Behaviour and Society 22 (2021): 22-31.

[25] Ranieri, L., Digiesi, S., Silvestri, B., & Roccotelli, M. (2018). A review of last mile logistics innovations in an externalities cost reduction vision. Sustainability, 10(3), 782.

[26] Savelsbergh, M., & Van Woensel, T. (2016). 50th anniversary invited article—city logistics: Challenges and opportunities. Transportation Science, 50(2), 579-590.

[27] Schliwa, G., Armitage, R., Aziz, S., Evans, J., & Rhoades, J. (2015). Sustainable city logistics—Making cargo cycles viable for urban freight transport. Research in Transportation Business & Management, 15, 50-57.

[28] Simoni, M. D., Marcucci, E., Gatta, V., & Claudel, C. G. (2019). Potential last-mile impacts of crowdshipping services: a simulation-based evaluation. Transportation, 1-22.

[29] Taniguchi, E., Thompson, R., & Qureshi, A. (2020). Modelling city logistics using recent innovative technologies. Transportation Research Procedia 46, 3-12.

[30] United Nations Development Program, Human Development Report 2019: Beyond income, beyond averages, beyond today: inequalities in human development in the 21st century. http://hdr.undp.org/en/2019-report (2019). Accessed 10 February 2021.

[31] United Nations, Department of Economic and Social Affairs: World Social Report 2020 – Inequality in a rapidly changing world. https://www.un.org/development/desa/publications/world-social-report-2020.html (2020). Accesse 10 February 2021.

[32] Van Duin, J. H. R., et al. "Improving home delivery efficiency by using principles of address intelligence for B2C deliveries." Transportation Research Procedia 12 (2016): 14-25.

[33] Wang, Y., & Boggio-Marzet, A. (2018). Evaluation of eco-driving training for fuel efficiency and emissions reduction according to road type. Sustainability, 10(11), 3891.

METHODOLOGY FOR DETERMINING SCENARIOS FOR COLLISION AVOIDANCE SYSTEM TESTING

Victor Bolbot, Dr
Gerasimos Theotokatos, Prof
Evangelos Boulougouris, Dr
Dracos Vassalos, Prof
Maritime Safety Research Centre
Department of Naval Architecture, Ocean and Marine Engineering
University of Strathclyde
100 Montrose street, G4 0LZ, Glasgow, United Kingdom

ABSTRACT

The introduction of autonomous ships is becoming a more tangible reality as a number of projects are dealing with developing next-generation ships operating in the European waters. One of such projects is AUTOSHIP aiming at developing two demonstrator ships with autonomous capabilities, one for short sea operations and another for inland waterways operations, funded by the European Commission. One of the key enabling technologies to be developed for this purpose is the collision avoidance system, whose safe performance is critical for autonomous ships. This can be ensured by proper testing of the collision scenarios in a simulated environment and during sea trials. Still, the quality of testing results will be depending on the number of considered scenarios. The aim of this study is therefore to suggest a systematic way to identify potential collision conditions which can be used as input for testing of the collision avoidance system. In our systematic approach, the potential situations that collision avoidance will have to handle are generated in a semi-automatic way by defining potential values for environmental features, objects' parameters, and intelligent agents' behavior. The generated hazardous scenarios are used to demonstrate the applicability of the developed approach.

Keywords:
Autonomous ships; Collision avoidance; Safety; Testing scenarios

1. INTRODUCTION

We live in an era when novel systems are being introduced, including marine autonomous surface ships (MASS) [1]. Major industrial initiatives, such as Yara Birkeland [2] and ASKO [3], have already launched commercial autonomous ships projects along by projects such as MUNIN [4], AAWA [5], SISU, SVAN [6], AUTOSHIP [1] and many others.

The collision avoidance system can be considered as a critical system on the autonomous ships as it will be making decisions [7] affecting the safety of the own and surrounding ships. A number of accidents have occurred in the other industries involving similar functions as referred in the literature since the design of such systems is associated with a number of challenges[8, 9]. One of the challenges is ensuring adequate situational coverage of the potential conditions that may be encountered by the ship associated with environmental complexity [10].

This problem can be tackled through systematic identification and analysis of the potential collision and interaction scenarios and by developing testing and verification techniques which would allow earning sufficient confidence in the collision avoidance functionalities. In other words, only once an adequate number of scenarios has been tested in a virtual or real environment can we be assured that such a system will not jeopardise safety, once put into operation.

The navigation of ships is primarily regulated by COLREGs [11]. However, the COLREG requirements have been designed having crew in mind and not the automatic systems. They do not provide numerical criteria for crew actions and their implementation rely on the crew judgement, so it can not be used to develop a comprehensive set of testing scenarios. They do not include information on the frequency of encountered situations and do not contain a comprehensive list of objects with which the MASSs will be intracting.

Automatic Identification System (AIS) data from ship traffic systems can be a valuable data source for identifying such situations. It has been widedly used for the analysis of traffic conditions and identification of the most probable collision situations (see for instance [12-15]). However, the quality of AIS data is under question, as the ships which have switched off their transponder and small recreational ships, which do not require to have AIS, are not visible on this data [16]. Also objects other than ships and buoys are not included in this data.

Several previous works have suggested the development of collision avoidance system on ships e.g. [17-19], but very few focused on the testing and testing scenarios generation [20, 21]. A number of other studies focused on testing scenarios generation such as in automotive [22, 23], aviation [24], etc.

The present study aims at developing testing scenarios for ship collision avoidance system. A number of sampling techniques is employed to generate various encountering conditions, but only the hazardous situations are selected for testing. The hazardous situations are identified using some criteria. In the study one Own Ship (OS) and two Target Ships (TS) are considered. The ships are considered to operate in an open ocean. More details are provided in the next sections.

2. METHODOLOGY

The methodology is presented using pseudocode in Table 1.

Table 1: The pseudocode used for scenarios generation

Algorithm Testing scenarios generation
Procedure: Pseudocode for testing scenario generation
Input: speed, weather, currents, ships size, number of test points (n), number of ships, etc.
Generate potential situations using SB, LH, RMC sampling
For i=1:n % for all the sample points
Estimate D_1 %distance between MASS and high speed craft $TCPA_1$ %time of closest approach $DCPA_1$ %distance of closest approach --- D_2 %distance between MASS and sailboat $TCPA_2$ %time of closest approach $DCPA_2$ %distance of closest approach --- a_1 %estimation of ellipse primary axis for OS b_1 %estimation of ellipse secondary axis for OS
If D_1<D_2 **then**

If $TCPA_1$>0 & $DCPA_1$<a_1 & D_1 <1nm **then**
Situation should be considered
Elseif $TCPA_2$>0 & $DCPA_2$<a_1 & D_2 <1nm **then**
Situation should be considered
End if

Elseif D_2<D_1 **then**

If $TCPA_2$>0 & $DCPA_2$<a_1 & D_2 <1nm **then**
Situation should be considered
Elseif $TCPA_1$>0 & $DCPA_1$<a_1 & D_1 <1nm **then**
Situation should be considered
End if

End if
Calculate the MSRE, mean for sampling methods
End procedure

First the input parameters are identified. This can be implemented with the assistance of the dedicated questionnaires and with the support of the AIS and meteorological data.

Then, during the second step, the encountering situations are generated using one of the sampling techniques: Sobol sequences (SB)[25], random Monte Carlo (RMC) sampling [26] or Latin hypercube (LH) [27]. Uniform distribution is assumed for the sampled parameters.

The SB are a quasi-random low-discrepancy sequence that uses two as basis for successively finer uniform partition of the unit in multidimensional space [25]. Example of SB in one dimension is the following sequence:

$$SB_1 = [1, 1/2, ¼, ¾, \ldots] \quad (1)$$

As it does not include stochastic calculations, every call of the SB sequence will generate the same result, in comparison to the RMC and LH sampling techniques.

The RMC generates samples based on repeated random sampling according to a predefined probability distribution [26]. The RMC are widely used in a number of application areas and scientific fields. The problem of random sampling can be likened to the picking up a ball with a number ranging from 1 to 100 from a Bingo machine.

Lastly, the LH is also method for generating a near-random samples [27]. In terms of sampling, it splits the sample space into a square grid and takes only one sample in each row and in each column. The sampling can be likened to rook chess problem, where the rooks must be set on such squares that they cannot be taken by the other rook, with the constrain that there must be a rook on each row and column.

For each of the encountering situation generated using the sampling techniques (SB, RMC, LH), geometric distance between the OS and TS_i (D_i), time to the closest point of approach between the OS and TS_i ($TCPA_i$) and distance at the closest point of approach between the OS and TS_i ($DCPA_i$) are estimated. The equations for these metrics are provided below (eq. 2-6), where (x, y) is location for each ship, u is speed, φ is ship speed direction, 0 is used to denote OS and i TS_i based on information presented in [18, 20, 21]. The equations 3 and 4 present the intermediate steps of the calculations.

$$D_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} \quad (2)$$

$$k_{2,i} = {u_0}^2 - 2u_0u_i\,cos(\varphi_0 - \varphi_i) + {u_i}^2 \quad (3)$$

$$k_{1,i} = 2\,(u_0y_0\,cos(\varphi_0) - u_0y_i\,cos(\varphi_0) - u_iy_0\,cos(\varphi_i) + u_iy_i\,cos(\varphi_i) + u_0x_0\,sin(\varphi_0) - u_0x_i\,sin(\varphi_0) - u_0x_i\,sin(\varphi_0) - u_ix_0\,sin(\varphi_i) + u_ix_i\,sin(\varphi_i)) \quad (4)$$

$$TCPA_i = -\frac{k_{1,i}}{2 \cdot k_{2,i}} \quad (5)$$

$$DCPA_i = \sqrt{k_{2,i} \cdot TCPA_i^2 + k_{2,i} \cdot TCPA_i + D_i^2} \quad (6)$$

Several criteria were considered for triggering the critical situations identification, namely:

a) Firstly, $TCPA_i > 0$, as we are not interested in the situation where the closest encounter happened in the past.
b) Secondarily, the current distance between ships should be equal or less than 1 nm (D_i <1nm). COLREGs do not specify any specific distance, and the 1nm has been set in line with other publications [18, 28]. It is the distance at which it can be considered that OS should take action to avoid the collision.
c) The third is criteria is related to the dangerous zone. The dangerous zone is defined using an ellipse set at ship location with axis dependent on the OS length and speed according to equations provided in [18], were a_1 denotes the primariy axis, b_1 the secondary axis of ellipse and L the ship length (shown also in Figure 1). This is expressed mathematically as $DCPA_2 < a_1$.

$$a_1 = \begin{cases} (4 - 0.3(10 - V_1))L, & u \leq 10\ kn \\ (4 + 0.3(V_1 - 10))L, & u > 10\ kn \end{cases} \quad (7)$$

$$b_1 = \begin{cases} (1.6 - 0.14(10 - V_1))L, & u \leq 10\ kn \\ (1.6 + 0.14(V_1 - 10))L, & u > 10\ kn \end{cases} \quad (8)$$

The criteria must concurrently be satisfied for a situation to be considered as hazardous. The analysis of criteria starts with the ship that is the closest to the OS.

The performance of sampling techniques is assessed using the following criteria a) the number of identified collision scenarios; b) the difference between the anticipated and actual mean for one of the independent parameters that are not used as input to the metrics and criteria in eq. (2) to (8) and c) the mean square root error (MSRE) of each sample for the independent parameter.

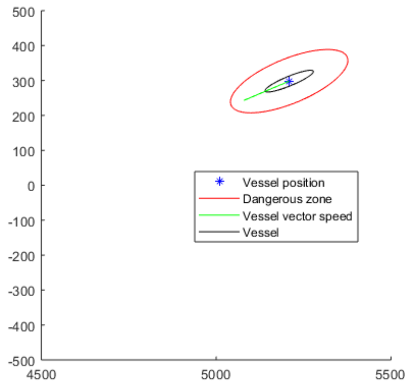


Figure 1: The demonstration of dangerous zone around ship

3. INVESTIGATED CASES

For the analysis we have employed a small short-sea ship (OS) from the AUTOSHIP project [1] which is operating outside coasts of Norway and is interacting with a sailing boat (TS1) and a high speed craft (TS2). The input parameters of the investigated situations are provided in Table 2. The sampled parameters with their ranges are provided in Table 3. These 18 parameters are varying from 0 to their maximum value by using the sampling techniques which have been referred previously (SB, RMC, LH). For comparing sampling techniques performance, 100 up to 100,000 samples are generated. Wave height is used as an independent parameter to estimate the mean and MSRE, since it is not used as input to eq. (2)-(8). The test area is set to [0 3nm] x [0 3nm] in line with [18].

Table 2: The input parameters.

	Fish feeding ship	Sailboat	High speed craft
Length	74.7m	6m	12m
Beam	13.6m	2m	2.5m
Max speed	15kn	10kn	40kn
Max current	3m/s		
Max waves height	2m		
Max wind speed	14 kn		

Table 3: The sampled parameters.

Random parameters	Range
Fish feeding ship speed	[0 max]
Fish feeding ship speed direction	[0 2pi]
Fish feeding ship location	[0 3nm] x [0 3nm]
Sail boat speed	[0 max]
Sail boat speed direction	[0 2pi]
Sail boat location	[0 3nm] x [0 3nm]
High speed craft speed	[0 max]
High speed craft direction	[0 2pi]
High speed craft location	[0 3nm] x [0 3nm]
Current speed	[0 max]
Current direction	[0 2pi]
Waves height	[0 max]
Waves direction	[0 2pi]
Wind speed	[0 max]
Wind direction	[0 2pi]

4. RESULTS AND DISCUSSION

4.1. Encountering situations generated using Sobol

The generated encountering conditions for SB number = 100 are provided in Figure 2. As it can be observed, the algorithm is effective in generating and identifying hazardous situations that may occur between the OS and the other two TS. Still, it can be also observed that out of the 100 potential encountering situations only 4 were classified as critical and important for testing. It practically indicates that during automatic scenarios generation a lot of 'non-hazardous' encountering situations will be produced. Still, this is dependent on the selected criteria and metrics.

4.2. Comparison between sampling techniques

The coverage of the 3nmx3nm test area using the three sampling techniques in the present paper is provided in Figure 3. It can be visually observed that the coverage of the area using SB sequences is more uniform than the other sampling techniques. For RMC sampling, clustering can be observed e.g. at x=2000m and y=5000m. At the same coordinates at LH sampling a void can be observed.

Further conclusions can be drawn through the metrics estimated for the different sampling techniques in Table 4 and Table 5. The results in Table 4 are generated repeating the sampling 10 times for RMC and LH sampling technique and 5 times in Table 5, as RMC and LH include stochastic calculations. Only for N=10^5 it was repeated 3 times. No repetition was required for SB samples, as it is a quasi-random technique, so the calculations are deterministic. It can be observed that SB samples in this algorithm identify similar number of scenarios with other techniques with the effectiveness at 4.5%. High discrepancy between the estimated mean waves height value and some samples mean was observed as demonstrated in Table 4 for RMC and LH compared to SB. The stability of the results is the significant advantage of the SB sequences due to their quasi-random nature. As a consequence, there is no need to repeat the results saving computation time when applying SB sequences. This is in line with the findings for SB in other engineering or mathematical problems [29-31].

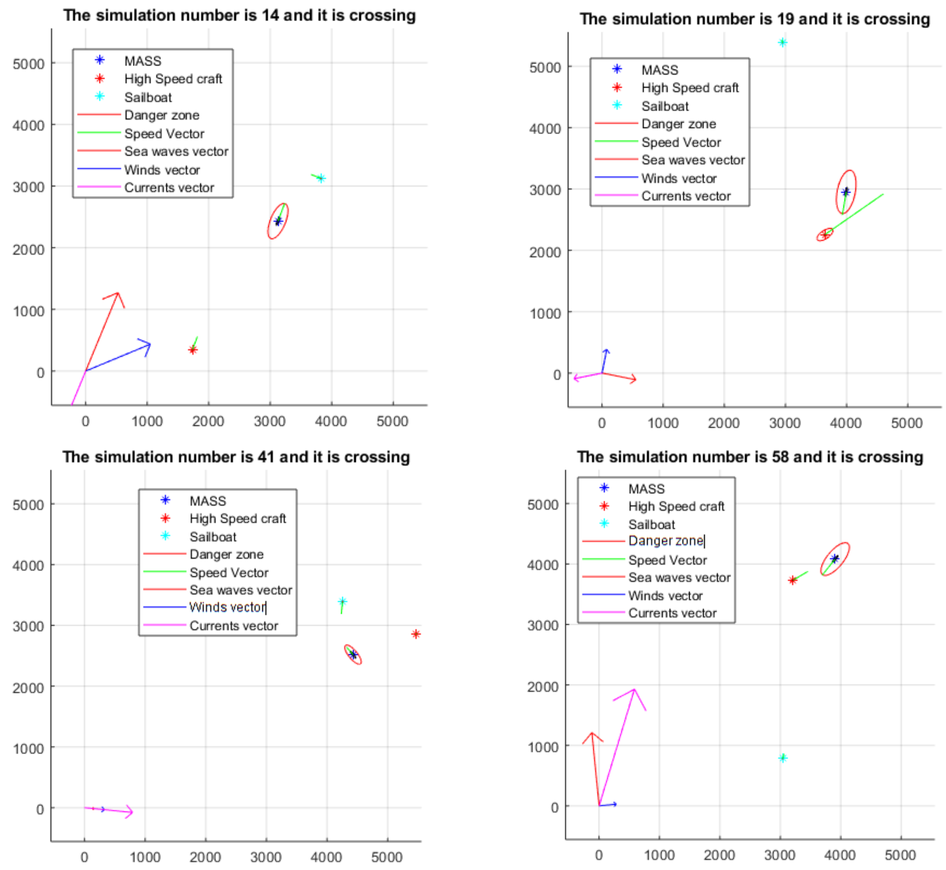


Figure 2: The hazardous encountering situations generated using Sobol sequences

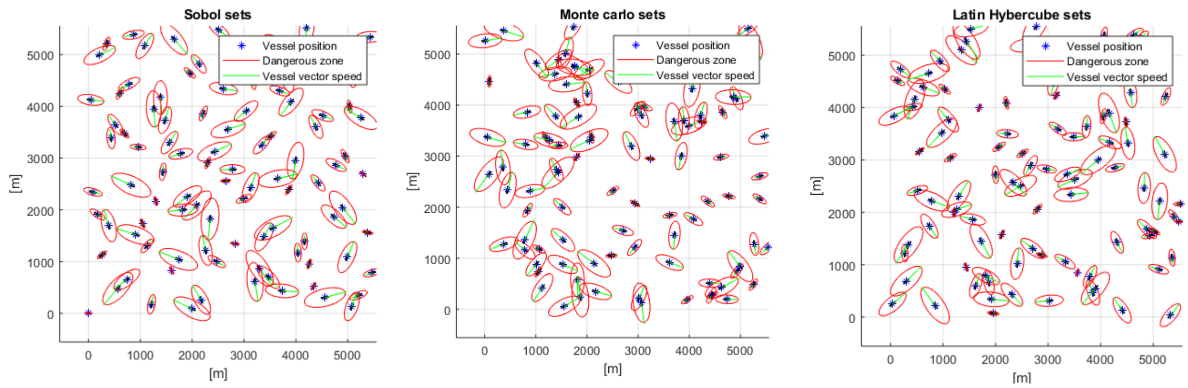


Figure 3: The filling of the test area using different Sampling techniques

Table 4: The comparison between different sampling techniques with 100 samples

	Waves height (expected mean 1)	N=100						Mean
	sampling repetition (total 10)	1	2	3	4	5	…	
LH	Collision scenarios n	6	11	7	4	3	…	5.80
LH	Waves height mean	0.73	1.26	1.14	0.65	0.62	…	1.01
LH	MSRE	0.38	0.36	0.31	0.48	0.61	…	0.35
RMC	Collision scenarios n	4	2	5	7	7	…	4.50
RMC	Waves height mean	0.84	0.91	0.45	0.98	0.96	…	0.98
RMC	MSRE	0.21	0.58	0.39	0.45	0.34	…	0.32
SB	Collision scenarios n	4						4.00
SB	Waves height mean	0.83						0.83
SB	MSRE	0.27						0.27

Table 5: The comparison between different sampling techniques with 10^2, 10^3, 10^4, 10^5 samples

	N=	10^2	10^3	10^4	10^5
LH	Collision scenarios n	5.80	42.60	446	4548
LH	Waves height mean	1.01	0.99	0.98	0.99
LH	MSRE	0.35	0.33	0.33	0.34
RMC	Collision scenarios n	4.50	42.00	443.40	4564
RMC	Waves height mean	0.98	0.98	0.99	1.00
RMC	MSRE	0.32	0.36	0.34	0.33
SB	Collision scenarios n	4.00	40.00	476.00	4496
SB	Waves height mean	0.83	0.91	1.04	1.00
SB	MSRE	0.27	0.29	0.35	0.34

5. CONCLUSIONS AND FURTHER RESEARCH

In this study an approach for automatic generation of hazardous ship traffic conditions has been presented. It was found that:

- The approach is effective for generating hazardous ship traffic conditions.
- The Sobol sampling provided accuracy comparable with other techniques.
- The Sobol sampling permitted more stable results compared to the other techniques due to quasi-random nature.

Our future work will focus on specifying more complex encountering conditions such as ship in proximity to shore. Another direction of future research is to specify coverage criteria for Sobol sampling.

Acknowledgments

The study was carried out in the framework of the AUTOSHIP project [1], which is funded by the European Union's Horizon 2020 research and innovation programme under agreement No 815012. The authors also greatly acknowledge the funding from DNV AS and RCCL for the MSRC establishment and operation. The opinions expressed herein are those of the authors and should not be construed to reflect the views of EU, DNV AS, RCCL or other involved partners in the AUTOSHIP project.

REFERENCES

[1] AUTOSHIP. Autonomous Shipping Initiative for European Waters. 2019.

[2] Yara. Yara Birkeland press kit. 2018.

[3] Smartmaritime. ASKO to build two autonomous vessels for Oslo fjord operations. 2020.

[4] MUNIN. Maritime Unmanned Navigation through Intelligence in Networks. 2016.

[5] AAWA. AAWA project introduces the project's first commercial ship operators. 2016.

[6] Daffey K. Technology Progression of Maritime Autonomous Surface Ships. 2018.

[7] Bolbot V, Theotokatos G, Boulougouris E, Vassalos D. A novel cyber-risk assessment method for ship systems. Safety Science. 2020;Under review.

[8] National Transportation Safety Board (NTSB). Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida May 7, 2016 Highway Accident Report. United States, Washington D.C.: National Transportation Safety Board; 2017.

[9] Guiochet J, Machin M, Waeselynck H. Safety-critical advanced robots: A survey. Robotics and Autonomous Systems. 2017;94:43-52.

[10] Alexander R, Hawkins HR, Rae AJ. Situation coverage–a coverage criterion for testing autonomous robots. 2015.

[11] COLREGS. International Regulations for Preventing Collisions at Sea - Articles of the Convention on the International Regulations for Preventing Collisions at Sea. 1972.

[12] Gao M, Shi G-Y. Ship collision avoidance anthropomorphic decision-making for structured learning based on AIS with Seq-CGAN. Ocean Engineering. 2020;217:107922.

[13] Mou JM, Van der Tak C, Ligteringen H. Study on collision avoidance in busy waterways by using AIS data. Ocean Engineering. 2010;37:483-90.

[14] Goerlandt F, Goite H, Valdez Banda OA, Höglund A, Ahonen-Rainio P, Lensu M. An analysis of wintertime navigational accidents in the Northern Baltic Sea. Safety Science. 2017;92:66-84.

[15] Kulkarni K, Goerlandt F, Li J, Banda OV, Kujala P. Preventing shipping accidents: Past, present, and future of waterway risk management with Baltic Sea focus. Safety Science. 2020;129:104798.

[16] IMO. Revised guidelines for the onboard operational use of shipborne Automatic Identification Systems (AIS) Resolution A.1106(29). 2015.

[17] Huang Y, van Gelder PHAJM. Collision risk measure for triggering evasive actions of maritime autonomous surface ships. Safety Science. 2020;127:104708.

[18] Namgung H, Kim J-S. Collision Risk Inference System for Maritime Autonomous Surface Ships Using COLREGs Rules Compliant Collision Avoidance. IEEE Access. 2021;9:7823-35.

[19] Brcko T, Androjna A, Srše J, Boć R. Vessel Multi-Parametric Collision Avoidance Decision Model: Fuzzy Approach. Journal of Marine Science and Engineering. 2021;9:49.

[20] Woerner K. COLREGS-compliant autonomous collision avoidance using multi-objective optimization with interval programming. MASSACHUSETTS INST OF TECH CAMBRIDGE; 2014.

[21] Pedersen TA, Glomsrud JA, Ruud E-L, Simonsen A, Sandrib J, Eriksen B-OH. Towards simulation-based verification of autonomous navigation systems. Safety Science. 2020;129:104799.

[22] Li L, Huang W, Liu Y, Zheng N, Wang F. Intelligence Testing for Autonomous Vehicles: A New Approach. IEEE Transactions on Intelligent Vehicles. 2016;1:158-66.

[23] Fremont DJ, Kim E, Pant YV, Seshia SA, Acharya A, Bruso X, et al. Formal Scenario-Based Testing of Autonomous Vehicles: From Simulation to the Real World. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)2020. p. 1-8.

[24] Lindvall M, Porter A, Magnusson G, Schulze C. Metamorphic model-based testing of autonomous systems. 2017 IEEE/ACM 2nd International Workshop on Metamorphic Testing (MET): IEEE; 2017. p. 35-41.

[25] Sobol' IyM. On the distribution of points in a cube and the approximate evaluation of integrals. Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki. 1967;7:784-802.

[26] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics. 1953;21:1087-92.

[27] McKay MD, Beckman RJ, Conover WJ. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics. 1979;21:239-45.

[28] Rødseth ØJ, Faivre J, Hjørungnes SR, Andersen P, Bolbot V, Pauwelyn A-S, et al. AUTOSHIP deliverable D3.1 Autonomous ship design standards, Revision 2.0. 2020.

[29] Burhenne S, Jacob D, Henze GP. Sampling based on Sobol'sequences for Monte Carlo techniques applied to building simulations. Proc Int Conf Build Simulat2011. p. 1816-23.

[30] Kucherenko S, Albrecht D, Saltelli A. Exploring multi-dimensional spaces: A comparison of Latin hypercube and quasi Monte Carlo sampling techniques. arXiv preprint arXiv:150502350. 2015.

[31] Qian G, Mahdi A. Sensitivity analysis methods in the biomedical sciences. Mathematical biosciences. 2020;323:108306.

CAN DRIVER IMPROVEMENT COURSES REDUCE RECIDIVISM IN BELGIUM?

Ellen Boudry, Researcher
Ricardo Nieuwkamp, Senior Researcher
Peter Silverans, Research Manager
Vias institute
Haachtsesteenweg 1405, Brussels 1130, Belgium
Ellen.Boudry@vias.be

Vanlaar Ward, Chief Operating Officer
Traffic Injury Research Foundation
171 Nepean Street, Ottawa, Canada
tirf@tirf.ca

ABSTRACT

Background. Driving under the influence of alcohol is a major threat to road safety. An important question is how recidivism can be prevented (i.e., reoffending). Sanctions can be distinguished in traditional and alternative sanctions. In this study the effectiveness of Driver Improvement (DI) courses – as alternative sanction – is determined based on recidivism data. **Method.** A quasi-experimental design (i.e., a static group comparison) is applied. Drivers in the comparison group (traditional sanction) are matched with the experimental group (DI course) on five aspects. 606 drivers convicted between 2010 and 2014 (reference conviction) were included. Recidivism was followed-up until January 2018. The effect is measured by comparing recidivism rate between groups and with a Cox regression analysis of time to recidivism. **Results.** After the follow-up period, recidivism rate of the experimental group was 41% lower than the comparison group. Based on the Cox regression analysis, the experimental group was 2.6 times less likely to reoffend. However, when the experimental group reoffends, their Blood Alcohol Concentration is higher than the comparison group. **Conclusions.** In line with international research, imposing a DI-course can reduce recidivism by 41%. Based on risk factors for recidivism, we aimed to identify driver profiles who benefit more from DI-courses. Profiles could not be identified due to ambiguous results. Replication with a larger sample can enhance the identification of driver profiles and contribute to an improved road safety in Belgium.

Keywords:
Effectiveness, Driver Improvement, Recidivism, Driving under the influence (DUI)

1. INTRODUCTION

1.1. Driving under the influence of alcohol and road safety

The consumption of alcohol makes a driver less able to correctly assess a traffic situation (Caird, Lees, & Edwards, 2005). In addition, alcohol has a visible effect on driver performance; there is an increase in lane waving and a higher variablility in speed (Irwin, Iudakhina, Desbrow, & Mccartney, 2017). In 2019, 9% of Belgian drivers involved in an accident tested positive on the use of alcohol (Slootmans, 2020). Hence, driving under the influence (DUI) of alcohol is a major treat to road safety worldwide.

Despite the higher risk of a traffic accident, DUI of alcohol occurs relatively frequent. In 2018, Vias institute (formely known as Belgian Road Safety Institute) estimated that 1.9% of the total kilometers driven in Belgium were completed by a driver who tested above the legal alcohol limit of 0.5 g/l (Brion, Meunier, Pelssers, Leblud, & Silverans, 2019). During weekend nights, the prevalence of DUI is estimated to be even higher. 12.6% of the Belgian drivers tested above the legal alcohol limit during weekend nights (Brion et al., 2019). Next, it is estimated that the crash risk of these drivers is 40% higher compared with sober drivers (Leskovšek & Goldenbeld, 2018). In addition, repeat traffic offenders are considered to be harmful for the society (Carnegie, Strawderman, & Li, 2009). More specifically, it is assumed that repeat traffic offenders have a higher chance to be involved in a traffic accident (Goldenbeld & Twisk, 2009).

Therefore, the question raises how to prevent that drivers, who drove under the influence of alcohol, will reoffend in the future.

1.2. Belgian legal framework

Belgium's legal alcohol limit is, as in most European countries, set at a blood alcohol concentration (BAC) of 0.5 g/l for all drivers. Except for professional drivers the legal alcohol limit is set at a BAC of 0.2 g/l. Next to this, the legal framework in Belgium distinguishes between a main sanction (e.g., a fine) and an additional sanction (e.g., driver license withdrawal) (Hoet, 2013). An argument for an additional sanction is that a main sanction doesn't take the context of the offences and the personal circumstances of the offender into account (Hoet, 2014). Moreover, sanctioning drivers with a traditional sanction (i.e., a fine or an imprisonment) has been shown to be ineffective in preventing recidivism (Af Wåhlberg, 2011) and stricter sanctions (i.e., higher fines) don't imply a reduction in recidivism (Elvik & Christensen, 2007).

In respons to the findings above, Vias institute developed alternative sanctions that aim to change the behavior of offenders. Alternative sanctions function as a substitute for the main sanction (Kluppels, 2018). Within the Belgian legal framework[1], the main and additional sanction can be suspended. One of the conditions to suspend the main sanction is to attend a sensibilization or driver improvent course (DI).[2] DI can be defined as an educational measure whereby the offender expierences a learning process. The main objective of this learning process is situated on the level of insight, increasing risk perception and becoming aware of one's own behavior and attitude towards traffic' (Kluppels, 2018, p. 13-14).

1.3. The effectiveness of driver improvement

Usually, recidivism data are used to evaluate the effectivesness of a DI course. A systematic literature review, in which six studies (mainly from the USA) were included with recidivism for DUI of alcohol as outcome measurement, found a decrease in recidivism in five out of the six studies. The decrease in recidivism ranged from 11% until 61% (Miller, Curtis, Sønderlund, Day, & Droste, 2015). In these studies, drivers sanctioned with a traditional sanction and drivers participated in a DI course were followed up by measuring reoffences of DUI of alcohol. Drivers that followed a DI course reoffended from 11% until 61% less for DUI of alcohol compared to traditional sanctioned drivers. Recently, a meta-analysis on the effectiveness of DI for DUI of alcohol found evidence for a decrease in the recidivism rate by 40%, which means that participants of a DI course reoffended 40% less compared to traditional sanctioned drivers (Slootmans et al., 2017).

In Belgium, to the best of our knowledge, only two studies have been published in the (grey) literature that investigated the effectiveness of DI on the base of recidivism data. The first study evaluated the initial and general DI course developed by Vias institute in 1996 by examining drivers between 1997 and 1999. The results show a decrease in the recidivism rate of 7%. However, that decrease was not significant in the first effectiveness study. Nevertheless, there was a trend that participants of the DI course reoffended less often and slower than participants of the comparison group. (Vanlaar, Kluppels, Wiseur, et al., 2003). The second, more recent study found a decrease in recidivism of 43% by following convicted drivers between 2012 and 2013. This study has some limitations, such as the small sample size, the limited observation period and the inclusion of only one juridical district (Waeyaert, 2017). Therefore, the results should be interpretated with caution.

1.4. Research design

An experimental design is considered to be the best scientific method to measure the effectiveness of an intervention. In an experimental design, participants are randomly assigned to the experimental group (i.e., drivers that participated in a DI course for DUI of alcohol) or to the comparison group (i.e., drivers that were sanctioned with a traditional sanction). In a juridical context, it is not possible to randomly impose a certain sanction to offenders. Therefore, a quasi-experimental design (i.e., a static group comparison) is usually applied in previous research on the effectiveness of DI as well as in the current study. Using existing groups implies that differences found between the experimental and comparison group can not be unambiguously explained by participation in a DI course.

A matching procedure ensures that both groups are comparable regarding important characteristics related to redivism. Hence, the problem of non-random allocation to research groups can be partially eliminated.

1.5. This study

Since the first effectiveness study in which recidivism was used as an outcome measure in Belgium by Vanlaar and colleagues (2003), the number of drivers that followed a DI course are substantially increased (Nieuwkamp & Slootmans, 2019). Even more important is that the content of the DI course is significantly evolved since the first

[1]https://www.ejustice.just.fgov.be/cgi_loi/change_lg_2.pl?language=nl&nm=1964062906&la=N

[2]https://justitie.belgium.be/sites/default/files/downloads/20180321_jp_15.pdf

courses in 1996. In the initial phase of the DI course, there was only a general DI course established for all types of traffic offenders. Whereas, nowadays there is a specific module for DUI of alcohol implemented. The current study examines the effectiveness of the module for DUI of alcohol. As in the previously mentioned studies, recidivism is the main outcome measure for effectiveness of the DI course. To determine the effectiveness, two groups of drivers who committed an offence for DUI of alcohol are compared. One group was sanctioned with a traditional sanction (e.g., a fine or an imprisonment), the other one with an alternative santion (i.e., DI course). After following the DI course or after the conviction, we observed whether the drivers reoffended. It is expected that drivers who followed the DI course will reoffend less (in time and frequence) compared to the drivers sanctioned with a traditional sanction.

1.6. Research questions and hypothesis

The main *research questions* of the present study are as follows:

1. Does following a DI course for DUI of alcohol lead to less recidivism compared to a traditional sanction?
2. Is there an effect of the type of sanction (DI course vs. traditional) on the number of offences committed after the reference conviction (recidivism frequency)?
3. Is there an effect of the type of sanction (DI course vs. traditional) on the time between the reference conviction and the next conviction (time until reoffending)?
4. Is there an effect of the type of sanction (DI course vs. traditional) on the trend of the BAC of the offender?
5. Is it possible to identify different profiles of drivers, based on the information available in the file of the offender, that can benefit more from a DI course on the one hand or a traditional sanction on the other hand?

In congruence with the previous research on this topic, we propose the following *research hypothesis*:

1. Following a DI course for DUI of alcohol reduces the risk of recidivism compared to a traditional sanction.
2. Offenders who followed a DI course have a lower recidivism frequency compared to those with a traditional sanction.
3. We expect that the time to reoffending is longer for offenders who followed a DI course compared to offenders with a traditional sanction.
4. Drivers who followed a DI course will have a lower BAC in the case of recidivism for DUI of alcohol compared to offenders traditionally sanctioned.

2. METHOD

The five matching criteria related to recidivism in the current study are the following:

1. Gender;
2. Age;
3. Blood alcohol concentration (BAC);
4. Period of conviction;
5. Type of prior convictions

2.1. Participants

The experimental group was composed using participant lists of the DI department of Vias institute. 2241 offenders followed a DI course in Belgium between 2010 and 2014. 303 offenders that completed a DI course for DUI of alcohol between 2010 and 2014 were selected from two different juridical districts. To compose the comparison group, 303 drivers in the same juridical districts with a traditional sanction for DUI of alcohol were selected using the five matching criteria described above.

The juridical information for the data collection process had to be encoded manually using a database of the Belgium court (i.e., MaCH database). The Board of Attorneys General gave their approval to use this database for the current study. Although, the database has been used by police courts in Belgium for several years, it is not very suitable for research purposes (Nieuwkamp, Slootmans, & Silverans, 2017).

2.2. Follow-up period

For all offenders, it was checked whether they had committed any new offence after the date of the last session of the DI course or the date of the conviction for those traditionally sanctioned. Recidivism was followed up until January 2018 for all the drivers in the study. By doing this, it is guaranteed that all offenders are followed for at least three years after their reference conviction.

2.3. Statistical analysis

In order to answer the research questions, a number of statistical tests are performed. Basic statictical tests, such as a chi-squared test and ANOVA are executed using the statistical software IBM SPSS Statistics 25. More advanced statical tests, such as a Cox regression analysis are performed in the statistical program R. This regression analysis, also known as a survival analysis, examines for each driver how much time there is between the reference conviction and a possible new conviction. For more information about this statistical method, Bijleveld and Commandeur (2009) can be consulted.

2.4. Characteristics offenders and comparability between the groups

At the time of the reference conviction, the average age of the offenders was 34 years (SD = 13.1) with a range from 18 until 71 years. The BAC ranges from 0.50 g/l until 4.44 g/l with an average of 1.76 g/l. The majority of the offenders were male (*n* = 534; 88.1%) and 37% of the offenders had a criminal record at the time of the reference conviction for DUI of alcohol. Table 1 and 2 show a breakdown of these results for the comparison group (traditionnaly sanctioned) and experimental group (DI course) to verify the matching procedure.

Table 1: Mean, standard deviation (SD) and statistics for the variables age and BAC disaggregated by comparison group (C) and experimental group (E)

Variable	Group	Mean	SD	Statistics
Age	C	34.90	13.30	$F(1, 602^{3}) = 1.53$, p = ns, $\eta^2 < .01$
	E	33.50	12.90	
BAC	C	1.75	0.60	$F(1, 593^{4}) = 0.42$, p = ns, $\eta^2 < .01$
	E	1.78	0.60	

Table 2: Percentage and statistics for the variables gender and criminal record disaggregated by the comparison group (C) and experimental group (E)

Variable	Label	Group	%	Statistics
Gender	Men	C	88.1%	(χ^2(1, n = 606) = 1.00, p = ns)
		E	88.1%	
	Women	C	11.9%	
		E	11.9%	
Criminal record	Yes	C	37.0%	χ^2(1, n = 606) = 1.00, p = ns
		E	37.0%	
	No	C	63.0%	
		E	63.0%	

Based on this, it can be concluded that the experimental group is similar to the comparison group with regard to the variables age, BAC, gender and criminal record (dichotomized).

3. RESULTS

3.1. Recidivism as measurement for effectiveness

At the end of the follow-up period of minimum three years, 309 of 606 drivers (51.0%) reoffended. From the drivers traditionally sanctionted, 194 of 303 (62.8%) reoffended. In contrast to 115 of 303 drivers (37.2%) that followed a DI course. This difference is statistically significant (χ^2 (1, n = 606) = 41.2, p <.001, *Cramer's V* = .26). Consequently, the recidivism rate of offenders that followed a DI course is 41% lower compared to offenders traditionally sanctioned. Most of the drivers reoffended for DUI of alcohol (*n* = 151; 49.5%), followed by reoffending for other traffic offences (*n* = 137; 44.9%). 17 drivers (5.6%) reoffended for non-traffic related offences. See Table 3 for an overview of the type of recidivism disaggregated for the experimental (i.e., DI course) and comparison group (i.e., traditional sanction).

Table 3: Overview of the type of recidivism

Type of recidivism	Experimental group (DI course)		Comparison group (traditional sanction)		Total	
	N	%	N	%	N	%
No recidivism	188	62.0	109	36.5	297	49.3
DUI of alcohol	65	21.5	86	28.8	151	25.1
Other traffic offences	43	14.2	94	31.4	137	22.8
All types of offences	7	2.3	10	3.3	17	3.0
Total	303	100	299[5]	100	602	100

The median of the recidivism frequency is 1 (SD = 1.62) with a minimum of 1 and a maximum of 12. The recidivism frequency is statistically different between both groups. Drivers that followed the DI course (M_e = 1,74; SD = 2.21) have a higher recidivism frequency compared to drivers traditionaly sanctioned (M_c = 1,20; SD = 1,10), F(1, 306) = 8.17, p = *.005*, η^2 = .03. This result is contradictory with hypothesis two.

The median of time until reoffending is 481 days with a minimum of 0 days and a maximum of 2463 days. As shown in Figure 1, fewer drivers reoffended after following a DI course compared to drivers traditionally sanctioned. More specifically, after 1000 days, approximately 25% of the drivers who followed the DI course reoffended. Whearas, approximately 50% of the drivers traditionally sanctioned reoffended.

[3]The variabele age was unknown for three drivers.
[4]The variable BAC was unknown for 12 drivers.
[5]The type of recidivism was unknown for four drivers.

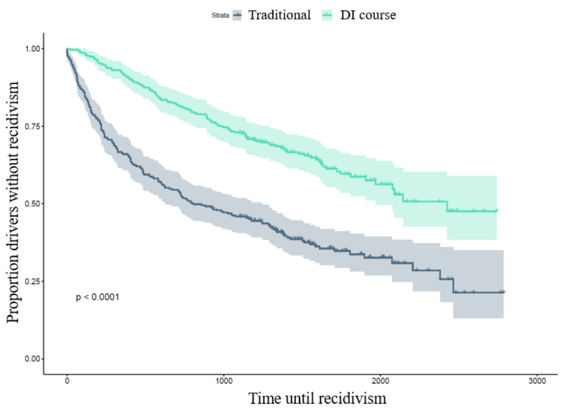


Figure 1: Time until recidivism

A Cox regression analysis measures the time until recidivism to calculate the effect of DI on recidivism. In the current study the hazard ratio (HR) represents the risk to reoffend. A HR more then one means that the risk for reoffending increases after following a DI course. Whereas, a HR less than one means that the risk to reoffend decreases. The type of sanction was included in the Cox regression analysis. The result show that following a DI course reduces the chance on reoffending with 2.6 (1/0.38; [0.28-0.52]; $p < 0.001$) compared to when the driver received a traditional sanction. Thus, a DI course reduces the risk of reoffending.

3.2. Trend in blood alcohol concentration

To answer research question four, we investigated whether there is a difference in recidivism severity (i.e., a different trend in BAC related to the type of sanction). Figure 2 shows the trend in BAC between the reference conviction and the next conviction in relation to the type of sanction (traditional vs. DI course). It should be noticed that only the reoffenders for DUI of alcohol are included in this analysis.

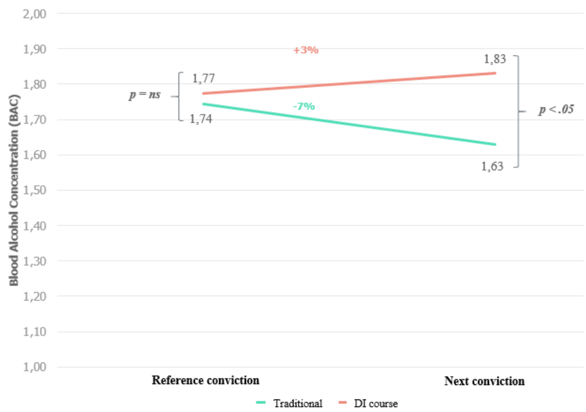


Figure 2: Trend in blood alcohol concentration of drivers alternative sanctioned (orange line) and driver traditionally sanctioned (blue line)

Based on a one-way ANOVA, the mean BAC level at the reference conviction was not significantly different between the two groups (M_e = 1.77, SD_e = 0.58; M_c = 1.74, SD_c = 0.55), $F(1, 592) = 0.63$, $p = 0.43$, $\eta^2 <.01$. Subsequently, it is important to note that there is a different trend between the two groups in the level of BAC at the reference conviction and the next conviction. More specifically, the BAC level of the traditionally sanctioned offenders declines with 7% after being traditionally sanctioned. Whereas, the BAC level of the drivers that followed the DI course slightly increases with 3% after completing the DI course (M_e = 1.83, SD_e = 0.58; M_c = 1.74, SD_c = 0.55). Thus, the BAC level of the next conviction for DUI of alcohol is significantly higher for offenders following a DI course compared to traditionally sanctioned offenders, $F(1, 181) = 4.17$, $p = .04$, $\eta^2 = .02$.

3.3. Identification of profiles of drivers

Finally, we examined whether it is possible to identify different profiles of drivers that can benefit more from imposing an alternative or traditional sanction. Herefore, the type of sanction (DI course *vs.* traditional) is combined with a level of four other variables available in the file of the offender using a Cox regression analysis. These four variables are respectively age, BAC, gender and criminal record. The reference level corresponds to offenders traditionally sanctioned in combination with another value of a variable mentioned in the file of the offender. By doing this, the added value of the DI course can be investigated. See Table 4 for an overview of the results.

For the variable age, it can be observed that older drivers have less chance to reoffend after following a DI course compared to younger drivers. In other words, the older the offender, the lower the chance to reoffend after completing a DI course. Drivers with a BAC level of more than 1.50 g/l at the reference conviction have 2.36 less chance to reoffend compared to drivers with the same BAC level traditionally sanctioned. Regarding to gender, both men and women can benefit from a DI course. However, women who followed the DI course have 4.44 less chance for recidivism. Whereas, men have a 2.67 lower chance to reoffend compared to men who were traditionally sanctioned. Drivers with no criminal record who followed the DI course have 3.44 less chance to reoffend compared to drivers with a criminal record that were imposed with a traditional sanction. This effect is also noticeable, but to a minor extent, for drivers without criminal record that were imposed with a traditional sanction. The latest group has 1.48 less chance to reoffend compared with drivers with a criminal record traditionally sanctioned.

Table 4: Overview of the type of sanction combined with age, BAC, gender and criminal record using a Cox regression analysis

	Exp(coef) [lower chance to reoffend]	p
Age category * type of sanction		
36+ * DI	4,14	***
36+ * traditional	NA	0,08
26 - 35 * DI	4,11	***
26 - 35 * traditional	NA	0,16
18 - 25 * DI	2,37	***
18 - 25 * traditional (Reference)	NA	NA
BAC * type of sanction		
0,50 - 0,99 g/l * DI	3,52	*
0,50 - 0,99 g/l * traditional	NA	0,57
1,0 - 1,50 g/l * DI	2,66	***
1,0 - 1,50 g/l * traditional	NA	0,18
1,50+ g/l * DI	2,36	***
1,50+ g/l * traditional (Reference)	NA	NA
Gender * type of sanction		
Women * DI	4,44	***
Women * traditional	NA	0,07
Men * DI	2,63	***
Men * traditional (Reference)	NA	NA
Criminal record * type of sanction		
No criminal record * DI	3,44	***
No criminal record * traditional	1,48	*
criminal record * DI	2,50	***
criminal record * traditional (Reference)	NA	NA

*Note: 'Exp(coef)' indicates the lower chance to reoffend for every interaction effect compared to the reference level. Significant differences with respect to the reference category are indicated by an asterisk (*); ***:<0,001; **: <0,01; *: <0,5.*

4. DISCUSSION

In 2003, a first study in Belgium was conducted to measure the effect of a DI course on recidivism for DUI of alcohol (Vanlaar, Kluppels, Wiseur, et al., 2003). This study served as a guideline for the methodology of the current study. More than 15 years later and after some important developments in the content of the Belgian DI courses, a new effectiveness study was required in order to measure the effect of a DI course for DUI of alcohol on recidivism.

The most important result of this effectiveness study is that following a DI course for DUI of alcohol is an effective measure to reduce recidivism for different types of traffic offences. More specifically, the recidivism rates of the group of drivers who followed the DI course was 41% lower in comparison with traditionally sanctioned. Furthermore, an analysis of the risk factors for recidivism shows that drivers who followed the DI course have 2.6 less chance to reoffend compared with traditionally sanctioned drivers. Moreover, in the case of recidivism, the group of drivers who completed the DI course reoffended later compared to reoffenders with a traditional sanction.

In order to identify which drivers would benefit more from following a DI course, the interaction between the type of sanction and other variables was investigated. It could be stated that the effect of the DI course on recidivism for all types of offences is more pronounced for women (4.44 lower chance to reoffend) than for men (2.67 lower chance to reoffend) in comparison with men with a traditional sanction. Next to this, the criminal record also plays an important role. Drivers without a criminal record, who followed a DI course, have a 3.44 lower chance to reoffend compared to drivers with a criminal record traditionally sanctioned.

Although the results are broadly in line with the findings from international review and meta-analysis (Miller et al., 2015; Slootmans et al., 2017) and follow the trend of previous Belgian studies (Vanlaar, Kluppels, Wisseur, et al., 2003; Waeyaert, 2017), the results of this study also differentiates from previous research. According to the results of Waeyaert (2017) for example, the BAC level of the reoffenders for DUI of alcohol decreased after completing the DI course. In contrast with the current study in which a longer follow-up period is used, we find that when drivers have completed the DI course and reoffend for DUI of alcohol, their level of BAC was higher compared to drivers traditionally sanctioned. It seems that for an important group of drivers, the DI training has a positive effect, but that for a minority of offenders, DUI of alcohol becomes even more severe.

4.1. Limitations

It should be noted that recidivism in this study is limited to registered behaviour. In other words, we do not know the *dark number* of drivers who have reoffended again for traffic offences. Furthermore, our recidivism results are only based on the MaCH database used by the Belgian court until January 2018.

Finally, in this study, a quasi experimental design was chosen since offenders with a alternative or traditional sanction are pre-existing groups of drivers. The validity of quasi experimental research is not guaranteed because the drivers were not randomly assigned to a group (Choi, Kho, Kim, & Park, 2019).

4.2. Future research

When databases are becoming available for research purposes, future research should focus on measuring the effect of DI courses on recidivism by using a bigger sample. Hereby, it can be possible to determine the effect of a DI course on different types of traffic offences, broader than DUI of alcohol.

REFERENCES

[1] Af Wåhlberg, A. E. (2011). Re-education of young driving offenders: Effects on recorded offences and self-reported collisions. Transportation Research Part F: Psychology and Behaviour, 14, 291–299.

[2] Bijleveld, C. C. J. H., & Commandeur, J. J. F. (2009). Multivariate analyse. Een inleiding voor criminologen en andere sociale wetenschappers. Den Haag: Boom Juridische uitgevers.

[3] Brion, M., Meunier, J.-C., Pelssers, B., Leblud, J., & Silverans, P. (2019). Alcohol achter het stuur : de situatie in België – Nationale gedragsmeting "Rijden onder invloed van alcohol" 2018. Brussel, België: Vias institute - Kenniscentrum Verkeersveiligheid.

[4] Caird, J., Lees, M., & Edwards, C. (2005). The Naturalistic Driver Model: A Review of Distraction, Impairment and Emergency Factors. IDEAS Working Paper Series from RePEc. Retrieved from http://search.proquest.com/docview/1698178518/

[5] Carnegie, J. A., Strawderman, W. E., & Li, W. (2009). Study of Recidivism Rates among Drivers Administratively Sanctioned by the New Jersey Motor Vehicle Commission. FHWA NJ-2009-019. Retrieved from: https://www.nj.gov/transportation/refdata/research/reports/FHWA-NJ-2009-019.pdf op 28/01/2019.

[6] Choi, Y.-Y., Kho, S.-Y., Kim, D.-K., & Park, B.-J. (2019). Analysis of the duration of compliance between recidivism of drunk driving and reinstatement of license after suspension or revocation. Accident Analysis & Prevention, *124*, 120–126.

[7] Goldenbeld, C., & Twisk, D. A. M. (2009). *Verkeersovertredingen, veelplegers en verkeersonveiligheid*. (R-2009-7). Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.

[8] Hoet, P. (2013). Werkbare probatie en herroeping. In P. Traest, J.-F. Leclercq, & E. Goethals (Eds.), Amicus Curiae. Liber amicorum Marc De Swaef (pp. 197–219). Antwerpen: Intersentia. Retrieved from https://www.jurisquare.be/en/book/9789400004306/werkbare-probatie-en-herroeping/

[9] Hoet, P. (2014). Vijftig jaar Probatiewet: verleden en toekomst van de individualisatie van de bestraffing. In CBR Jaarboek 2013-2014 (pp. 319–391). Mortsel: Intersentia. Retrieved from https://www.jurisquare.be/en/book/9789400004498/vijftig-jaar-probatiewet-verleden-en-toekomst-van-de-individualisatie-van-de-bestraffing/

[10] Irwin, C., Iudakhina, E., Desbrow, B., & Mccartney, D. (2017). Effects of acute alcohol consumption on measures of simulated driving: A systematic review and meta-analysis. Accident Analysis and Prevention, 102, 248–266.

[11] Kluppels, L. (2018). Educatieve maatregelen voor verkeersovertreders. Het nut van leren als straf. Brussel, België: Politea.

[12] Leskovšek, B., Goldenbeld, Ch. (2018). Influenced Driving – Alcohol, European Road Safety Decision Support System, developed by the H2020 project SafetyCube. Retrieved from www.roadsafety-dss.eu

[13] Miller, P. G., Curtis, A., Sønderlund, A., Day, A., & Droste, N. (2015). Effectiveness of interventions for convicted DUI offenders in reducing recidivism: a systematic review of the peer-reviewed scientific literature. The American Journal of Drug and Alcohol Abuse, 41(1), 16–29. Retrieved from http://www.tandfonline.com/doi/abs/10.3109/00952990.2014.966199

[14] Slootmans, F. (2020). Statistisch rapport 2020. Verkeersongevallen 2019, Brussel, België: Vias institute – Kenniscentrum Verkeersveiligheid

[15] Nieuwkamp, R., & Slootmans, F. (2019). Statistisch rapport 2019. Handhaving: controle en sancties. Brussel, België: Vias institute - Kenniscentrum Verkeersveiligheid.

[16] Nieuwkamp, R., Slootmans, F., & Silverans, P. (2017). Verkeersrecidive nader bekeken: Verkenning van een steekproef van gerechtelijke dossiers uit Leuven en Mechelen. Brussel, België: Belgisch Instituut voor de Verkeersveiligheid - Kenniscentrum Verkeersveiligheid.

[17] Slootmans, F., Martensen, H., Kluppels, L., & Meesmann, U. (2017). Rehabilitation courses as alternative measure for drink-driving offenders. European Road Safety Decision Support System, developed by the H2020 project SafetyCube. Retrieved from www.roadsafety-dss.eu on 11/01/2019.

[18] Vanlaar, W. (2018). Methode ter evaluatie van de Vias Driver Improvement cursussen voor bestuurders onder invloed van alcohol. Ottawa, Canada: Traffic Injury Research Foundation (TIRF). Unpublished internal report.

[19] Vanlaar, W., Kluppels, L., Wiseur, A., & Goossens, F. (2003). Leiden sensibilisatiecursussen voor bestuurders onder invloed van alcohol tot een lager recidivegehalte dan klassieke straffen? *Panopticon*, *4*, 365–390.

[20] Waeyaert, E. (2017). *Stageopdracht Emiel Waeyaert*. Unpublished internship report.

UNDERSTANDING THE DRIVERS AND BARRIERS IN THE PROCESS OF DEPLOYMENT AND OPERATION OF DIGITAL MOBILITY SERVICES FROM AN INCLUSION AND ACCESSIBILITY POINT OF VIEW

Hannes Delaere, PhD researcher
Samyajit Basu, Senior researcher
Imre Keserü, Professor
Cathy Macharis, Professor
Vrije Universiteit Brussel
Pleinlaan 2, Brussel, Belgium
Hannes.Delaere@vub.be, Samyajit.Basu@vub.be, Imre.Keseru@vub.be, Cathy.Macharis@vub.be

ABSTRACT

The impact of digital mobility services on the mobility landscape is growing and users are increasingly having more options to choose from every day. There is however another side to this growth, the side where vulnerable-to-exclusion groups do not have the possibility to equally benefit from these services. In order to bring changes that can extend the benefits of digital mobility services and applications equally to the vulnerable-to-exclusion groups, it is necessary to understand what the drivers and barriers are for the development and deployment of inclusive and accessible digital mobility services. In this study we have tried to explore those drivers and barriers by performing semi-structured interviews with stakeholders involved in the development and day to day operations of 10 different digital mobility services throughout Europe which served as case studies. As part of these case studies, a literature review and desktop research were performed, in combination with semi-structured interviews and a content analysis. Three main groups of stakeholders (developers, operators, policymakers and in some cases also user group representatives) were interviewed with a focus on the regulatory framework in effect and how inclusion and accessibility were taken into account. First, the semi-structured interviews were transcribed, and a qualitative thematic analysis was performed. After this, the collected experiences of the stakeholders were consolidated and discussed at an online workshop, involving the three main stakeholder groups. The main conclusions drawn from this research are that there is a multifaceted problem; clear communications, collaboration and the related co-creation are not present in most services, no mobility or development related information is shared, and although the concept and importance of inclusion were known to the interviewees, only very few services take action towards providing an inclusive digital service.

Keywords:
Inclusive digital mobility, vulnerable-to-exclusion groups, case studies, semi-structured interviews, qualitative thematic analysis

1. INTRODUCTION

The rise of the internet and - shortly after - the emergence of the mobile phone with internet access, have opened new possibilities for people to organise their travel. Since these technologies have become widely spread, personal mobility has changed dramatically: real time information has made it possible for people to adapt to the journey environment by having access to information about traffic jams, better routes, arrival times of public transport, location of shared mobility stations, etc.

In his first two laws of technology Kranzberg says: “Technology is neither good nor bad; nor is it neutral.” and “Every technical innovation seems to require additional technical advances in order to make it fully effective.”(Kranzberg, 1986). The first law simply says that the impact technological inventions have, depends on the way humans use the invention. The second law indicates that an invention always needs additional technological development so that it can be used effectively in our society. The two laws can be applied to the current technological developments resulting in the creation of the digital transport system. The digital transport system has a significant impact on our society, but it has not been used to its full potential. Kranzberg approaches the problem from a strictly technological point of view, but for the digital transport system there is need for a significant human impact factor. Technological adoption in combination with a change in behaviour takes time. An example to clarify: according to Eurostat data from 2019, on average 27% of the European inhabitants have no access to internet when on the move, 40% have never ordered goods or services using the internet and 14% have never used the internet (Eurostat, 2020a, 2020b, 2020c). So, developers of mobility services need to be aware that a lot of people do not have the skills, financial tools or understanding of the digital system to use it. This results in the current development and operation of a high impact service that cannot be used by a significant number of people in Europe (Franckx & Mayeres, 2016).

In reference with Kranzbergs' laws, there is need for understanding the process which led to the introduction of the new digital mobility services. To do that, the development process needs to be studied with a focus on the stakeholders involved in the development process of the digital mobility services. Three main groups are considered pivotal in this process: developers, operators, policy makers. Each of these groups can have a significant impact on the way digital mobility services are developed and by whom they can be used: what decisions are made, how do these stakeholders see the involvement of currently excluded groups, what initiatives have already been taken and what exactly are the barriers and drivers for the stakeholders to develop a more inclusive digital mobility service.

In this paper we try to understand the process leading to the current digital transport services, how the development and deployment took place and if that process has resulted in the exclusion of several vulnerable groups in our society. Therefore the main research question in this paper is: "What are the main barriers and drivers during the development and deployment phase for an inclusive digital transport service?"

This paper is structured in 4 sections: the methodology, the results, the discussion and conclusion. In the methodology the selection of cases, related stakeholders and the data-analysis method are explained. In the results we discuss the four main digital transport service types: car-and ridesharing, micromobility and bikesharing, smart logistics and the multimodal routeplanners and MaaS. The final part of this paper contains the discussion and conclusions.

2. METHODOLOGY

This research focused on the collection of data by interviewing stakeholders involved in the development or operation of digital transport services, using semi-structured interviews. The analysis is based on the **case study method**. For socio-economic research, case studies are one of the principal means used to collect data (Bates et al., 1998; Robinson et al., 2003).

'A Case study is an in-depth exploration from multiple perspectives of the complexity and uniqueness of a particular project, policy, institution, program or system in a "real life" context' (Simons, 2009).

The **case studies** focus on the process of the deployment of the digital mobility services from the viewpoint of the different stakeholder groups. Concretely, with these cases we will try to understand and create an overview of the drivers and barriers, requirements and needs developers, operators and policy-makers experience when attempting to develop/deploy some form of digital mobility service.

For this research **three main stakeholder groups** are identified: developers, operators, policy makers. Some user representatives were also interviewed to gain a better understanding of the different vulnerable to exclusion users.

We have identified six steps in the methodology: the first step is the selection of the 10 deployment case studies. Secondly, a desktop study is conducted; thirdly, for each case study relevant stakeholders have to be identified. In step 4, semi-structured interviews (SSI) are developed and conducted with the stakeholders. Next the interviews are transcribed and analysed using the qualitative thematic analysis method. Finally, in step 6, the results of the case studies are consolidated and discussed with the stakeholders.

2.1. Identification 10 deployment case studies

The case studies have been selected taking into account the following criteria:

- Inclusion of digital mobility services that are already being deployed or are promising in terms of their future development
- Coverrage of a great variety of digital mobility services so that we can highlight any similarities and differences in the barriers and drivers
- Ccoverage of multiple cities or regions

- Involvement of a variety of stakeholders

Therefore, we used the digital mobility services as identified by Kedmi Shahar et al. (2020), as a starting point and identified digitally driven e-scooter services, crowd logistics services, free-floating car-sharing, on-demand ridesharing and Mobility-as-a-Service (MaaS) applications as minimum categories that should be covered by the case studies.

A first selection of cases is based on input from previous projects, personal knowledge and networks and by an online research to contact services that comply with the criteria described above. The first contact with a potential case took place in July 2020, during the COVID-19 crisis, which possibly had a negative impact on the services willing to provide help for this research.

2.2. Identification of relevant stakeholders

In order to identify and contact the relevant stakeholders, a **definition for a stakeholder** must be integrated within the method. A stakeholder in the singular meaning is:

”*Any group or individual who can affect or is affected by the achievement of the organization's objectives*” (Freeman, 2010, p.46)

For the identification of stakeholders, **a stakeholder analysis** was conducted. The stakeholder analysis is not a single tool used to identify stakeholders but consists of multiple different methodologies (Crosby, 1991). Two methods were used to identify the deployment case study stakeholders: the snowball-mapping method and the stakeholder-led categorisation. For each of the stakeholder groups the snowball mapping-method for identifying new stakeholders is initially used, while afterwards, the stakeholders were asked to categorize themselves as one of the stakeholder groups. Stakeholder mapping provides an overview of the different stakeholders, but also gives an indication about the strength of the inter- and intra-stakeholder group relationships (Brugha & Varvasovszky, 2000).

For each of the cases multiple stakeholders are interviewed, their answers analysed using a thematic analysis and the results discussed during an interactive workshop. In total, 22 interviews were conducted, 5 with developers, 8 with operators, 9 with policy makers and one with a user group representative.

2.3. Development of the interviews

The data collected from the interviews includes several aspects of the digital mobility services:

- How weas inclusion and accessibility aspects taken into account?
- How were the regulatory framework taken into effect?
- How was the service or application developed/deployed?

During the development of the questions we decided that the use of semi-structured interviews (SSI) is preferable for this research. The language barrier, in combination with unknown stakeholders and topics that are relatively new, which might require more elaborate explanations, make the SSI the more suited method.

We have developed the interviews based on literature (Goodwin, 2013; Pope, 2020; Schmeer, 1999; Wick, 2012). The questions were developed based on topics related to user involvement, the current regulatory framework in effect, how a planning was realised and followed, how decisions were made, the barriers and drivers that were experienced and other relevant topics. Depending on the type of stakeholder, the topics touched during the interview were slightly different.

2.4. Analysis of the interviews

The analysis of the interviews consists of different steps. The overall method used is the thematic analysis method, as explained by Braun and Clarke (2006, p. 79). The authors describe the thematic method as an independent qualitative method that can be defined as “a method for identifying, analysing and reporting patterns (themes) within data. It minimally organises and describes your data set in (rich) detail.”

For this research, NVivo analytic software is used, which made it possible to analyse the interviews in a structured and thorough manner. For a good presentation of the results, verbatims are also used to strengthen certain claims made based on the input from the interviews that were validated during the workshop.

2.5. Development of the thematic analysis codes

During the initial stage of the analysis, the desktop research was used to create a first summary of the context in each of the cases. This provided some of the initial information for the development of the semi-structured interviews and for the codes used within the thematic analysis. During the actual thematic analysis some additional codes were added to complement the predefined ones. In Table 1, all codes are split up in two categories: “primary coding topics” and “secondary coding topics” which are respectively the initial topics selected for the thematic analysis and the topics developed during the thematic analysis itself, based on new information collected during the analysis of the interview.

Table 1: Thematic analysis codes

Primary coding topics	Secondary coding topics
Regulatory framework	Future ideas and concepts
Inclusion aspects	Financial aspects
Data-protection/-collection and privacy	Business as usual (BAU) development
User involvement/ co-creation	General drivers for development
	General barriers for development

Source: own data

2.6. Co-creation workshop

After the initial results from the interviews, an online workshop was organised to consolidate and discuss the first results from the stakeholder interviews with external experts in mobility and logistics services consisting of transport and logistics service operators, software developers, policy makers, mobility researchers. In total, 36 experts participated in the interactive workshop. The objective of the workshop was to present and discuss the insights gathered in the case studies. The discussion was mainly used to validate the information that was collected from the interviews with the stakeholders.

3. RESULTS

3.1. Selection of cases and stakeholders

The ten services that were selected to serve as case studies are presented in Table 2. The results are presented in four different sections, based on the major categories in the digital transport system: first car- and ridesharing is discussed, followed by bike sharing and micromobility, smart logistics systems and finaly also multimodal routeplanners and MaaS.

Over the period of half a year, 22 stakeholder organisations, linked to a mobility or logistics service, were contacted to participate in this research. Finally, 10 suited cases were selected with 22 interviewees willing to participate.

3.2. Car-and ridesharing

The topic of car-and ridesharing is the first of four topics and is covered by two case studies, Cambio and Mobitwin, which respectively resulted in two and three interviews. Cambio is a provider of station-based shared cars in Belgium. The service was first introduced in Germany, but is also active in Belgium since 2002. Mobitwin is a rideservice for elderly people and people who experience difficulties with their mobility. For Cambio, a developer/operator and a policy maker from Brussels Mobility, the regional administration responsible for mobility were interviewed. For the Mobitwin case, three interviews took place with a developer, an operator and a user group representative. For the Brussels Capital region, a framework was developed related to shared mobility. There are no specific demands in the framework related to inclusive digital mobility, but some general accessibility and inclusion-related remarks are included e.g. a clear and accessible subscription system should be in place, the operator can also choose to make the cars available without the need for a subscription (Belgisch staatsblad, 2016).

3.2.1. Interviews

As mentioned above, car- and ridesharing services are the oldest types of shared mobility studied in this project. This results in a **regulatory framework** that is older and has been further developed and adapted compared to other services. During the interviews it was noticeable that the developers/operators have been in the business for a longer time and that the market is more stable. Because of that stability and the fact that carsharing is already quite popular, it has been present in all major Belgian cities for quite some time.

Several aspects related to **inclusion** have already been introduced within carsharing-services, mostly focusing on providing a service available in multiple neighbourhoods within the city. Much fewer adaptations have been made based on digital inclusivity, financial inclusivity, etc. For Cambio, this is different, their main ‘inclusivity’-aspect is embedded in the core of the service: the call service. On their website they also explicitly mention they want to provide a service that is affordable for all (Cambio Flanders, 2020).

The interviewee from the Cambio operator also said that not only the digital exclusion of lower educated citizens is an issue to use their service, but general knowledge about carsharing is lacking. Especially among the lower educated people, the knowledge about shared or digital mobility was almost non-existing. This shows that even though there is a lot of digital marketing for similar services, the vulnerable-to-exclusion groups are not reached.

User involvement and the strongly related bottom-up method for development of services could lead to a more inclusive service. One of the operators of the Mobitwin app, whose primary job was actually that of a social worker, said: “knowledge of vulnerable to exclusion groups is low among developers” which was confirmed by the developer of the application. The issues of app development and co-creation are not that easily solved according to the stakeholders because of several reasons: first, it is very hard for developers to reach vulnerable to exclusion groups, secondly, there is much difficulty to find enough people to have a sufficient large group and thirdly, activating and persuading ‘vulnerable’ groups to participate in e.g. testing phases has proven to be a major issue.

For the **data collection, protection & privacy** of different partners in a digital environment there are several issues that can be addressed, according to the stakeholders. First, it is important that the implications of data sharing are made available and understandable for all users. Elderly people do not have the knowledge about the amount of private data they share. Secondly, they are rather keen on personal connection which led in some cases in sharing too much personal information. The operators from Mobitwin complained in multiple cases about the users having access to personal phone-numbers of the voluntary drivers (which is sometimes allowed in case of emergency):

The Mobitwin app was tested among elderly people in 2019 and, after being introduced to the service, many of them downloaded the app and used it a few times, but afterwards actual use of the app was negligible. Many considered it too difficult with too many options. A completely different reaction came from the social and healthcare workers in elderly homes, who considered the app a very useful tool which made the process for making reservations a lot easier and faster. The operators from both services said that there was need for better guidance and understanding of those specific groups.

3.3. Bike sharing and micro mobility

For the micro-mobility related services, three interviews were performed: two with stakeholders of the HIVE e-scooters in Lisbon and one with a policy maker for micro-mobility services in Brussels. HIVE (part of Free Now) is a free-floating e-scooter service, started in Lisbon in 2018 and is part of the Daimler group. They introduced 600 shared e-scooters in the city and are even expanding, combined with the introduction of extra e-scooters, they also put a lot of effort in close collaboration with other local organisations and especially with local policy makers (Hinchliffe, 2018). The app can be used in 11 languages, with the Catalan language included. Lisbon is currently using a soft regulatory framework in order to attract providers of shared mobility.

For the Brussels' case, no provider of micro mobility was willing to take part in the research. During the interview with the policy maker, some questions related to the operation of a micro mobility or shared bike service were answered, since the Brussels Region also has its own public shared bike scheme 'Villo'. Villo is a station-based bike sharing service provided by the Brussels municipality. The service currently offers about 5000 bikes stalled in 360 stations all over the Brussels Region. Blue bike is also present at all train stations in Brussels providing station-based shared bikes. Villo and Blue bike are the two best known providers of shared bikes, but several other providers have been present as well over the years.

Next to bike sharing schemes, there are also plenty of free-floating scooters present in the city such as Dott, Lime and Bird. Bird was already present in the city from 2018 untill the winter of 2019, after which they retracted their scooters until November 2020. The company restarted their service in Brussels with 200 e-scooters (Belga, 2020). Even though several other providers have left the city, the scooters are still widely used by more than 100 000 users, but there is also a lot of criticism because of their impact on the public space and especially related to safety issues when used on sidewalks, as they often are (Wallemacq, 2019).

3.3.1. Interviews

Although no actual bike provider was interviewed, the free-floating bikes have similar issues as the free-floating e-scooters. The experts in the workshop considered their barriers and strategies to overcome these barriers applicable to both e-scooters and shared bikes.

The first aspect discussed is the way in which the **regulatory framework** is being developed in the two cases and how this might impact the digital inclusivity of the service. Many European countries have developed some sort of regulatory framework for shared mobility. The regulatory framework is often developed at a national level, but in many cases, also at a city-level (Eltis, 2020). Both for Brussels and Lisbon, a regional policy has been developed. This has not yet been fine-tuned so there are still a lot of gaps, also related to the limited knowledge (about use, impact, (dis)advantages) policy makers often have about the types of services. Both have some form of regulations to increase the accessibility and inclusiveness of the services provided in their city, but they are based on physical accessibility rather than digital accessibility and inclusiveness. In Brussels and Lisbon, the regulatory framework was primarily developed for car- and ridesharing services; only later were micro mobility and bike sharing added. Although a lot of efforts have been made for the continuous development of the framework, the Brussels policy maker stated that the framework is still not extensive enough and a bit outdated.

Lisbon was one of the early adopters within the shared and digital mobility market. The city chose a soft regulation approach, which making it more attractive for operators, which resulted in a large number of them wanting to operate in Lisbon. This was also confirmed by the interviewed operator: "I think the approach of Lisbon it just gives the benefit of the doubt for the companies to come" These types of regulations were combined with hard regulations as the city's parking regulations prohibiting vehicles to park on the sidewalks. When the local government received complaints from elderly people and people who are visually impaired regarding all the scooters on the sidewalk, the city used parking regulations to prohibit scooters to park on the sidewalk.

Due to other hard regulations, such as the parking regulations, which have effect on the way in which e.g. e-scooters are used in public space, the city can still impose their will on the e-scooter providers.

The Lisbon policy maker suggested a second reason why soft regulations were introduced. There was a lack of knowledge about micro-mobility among the policy makers. Hard regulations simply require more knowledge about the sector and its effects on society and public space in order for these regulations to be effective and useful. In these cases the policy maker said it was better to use soft regulations. In Brussels a stricter, license-based approach was adopted, but the Brussels' policy maker stated that, many cities (incl. Brussels) do not have the funding or personnel to monitor if all the digital transport services operating in the city comply to those regulations.

Strongly related to the regulatory framework is the introduction of measures for a more **inclusive service**. The main issue the interviewees mentioned was the instability of the market and the fact that, although it evolves quickly, it is still a niche-product, which leads to the providers/developers wanting to develop the service in its totality, with only a (very) limited focus on the digital inclusivity of the service. For operators, developers and policy makers the development of a good 'general' service, within a reasonable regulatory framework that is financially stable is currently the main goal. Policy makers did acknowledge that research into more inclusive services is needed, but as stated before, it is currently not possible due to lack of knowledge and financial resources. It was considered as a failure of policy by both policy makers. Although in general no massive attempts or regulations are made to change this on a large scale, some financial-inclusivity aspects are considered, such as 30 min free e-scooter time for inhabitants of Lisbon with limited financial means, hoping this would show the opportunity micro-mobility can provide for these people. Compared to the shared car market, it shows that financial and market stability leaves room for more inclusive development. It was mentioned by stakeholders for both topics that if financial or market stability is not present, not many developers and/or operators are willing to spend a lot of money on inclusiveness.

User involvement and communication will result in better relations between the developers, operators and stakeholders, which has a positive impact on the service. In both cities, information events are organized by the local authority for people to learn about the new services and their digital aspects, and when asked, all stakeholder groups interviewed consider the input from end-users as being important and having a positive effect on the match of what is being developed and what is needed, but also admitted that this type of cooperation is still lacking in many cases.

There are two main issues concerning the co-creation or info-events: first there is the composition of the group knowing about and attending these events and secondly, many of these events, as described by the Brussels policy maker, are rather info-events, where information that has already been decided is shared with social organisations (e.g. unions, local social groups, neighbourhood groups, etc.). Certain groups in society (e.g. older people, migrants, people with limited education) are not as easily reached and as a result, their opinion and issues are not heard. So, these workshops will have a positive effect on the service, but many vulnerable groups are still overlooked. In Brussels some of these issues are being tackled by involving social organizations in the info-events (e.g. private organizations, workers unions etc.), policy makers hope that by introducing these organizations to the development of the services they will get wider support, and the service developed will be better tailored to the needs of different groups within society.

According to the Brussels policy maker, the concept of **data collection, protection & privacy** will also become an important matter for the creation of better digital mobility services. The stakeholder stated: "Well, it's (the exchange of data and protection) definitely a work in progress, this trust issue between private sector and public authority, and public transport operator as well, actually, it's definitely something that we have to address." One of the major arguments against intensive data collection given by the Lisbon operator was: "… we have a less pressure to collect massive data. And, I think it's good because it just takes a very big risk in terms of data management." The same issue comes up in each of the interviews, data-collection is hard, the advantages are not always very clear, lack of trust, lack of knowledge, not enough resources for specific data collection. When asked for a vision on future data collection and especially management, the operator did not think intensive collection of data (e.g. data about vulnerable groups, specific people, areas etc.) will become the standard. This vision was not shared by the policy makers, they see a lot of potential in data collection and management.

3.4. Smart logistics services

Two types of logistics are studied. First there are smart lockers, in Valencia and the 'Mobile Lockers' in Flanders, and the delivery of goods using (cargo)bikes in Madrid. Logistics services struggle with several aspects related to digital inclusion (e.g. how do blind people access lockers, how does someone without a smartphone order food and how can last mile package delivery become lucrative, but also available for all?).

For the Citypack lockers in Valencia a city representative, who was involved in the deployment of the lockers was interviewed. For Mobile Locker, their CEO responded the questions from both a developer and an operator point of view. For the Coopcycle case in Madrid, three people were interviewed, a developer, a policy maker and an operator. The Coopcycle case is a cooperative service, focusing on food delivery using (cargo)bikes. Some of the issues related to the delivery of food in cities, which has grown exponentially over the years, were dominating the news for quite some time, especially related to the Coopcycle case in Madrid. The Coopcycle service was developed in 2017 as a response to the initial problems that surfaced related to the 'self-employment' status of the drivers for other good-delivery services. La Pajara started working on their first collaboration projects and helped customers to fight against the working conditions of mainstreaming platform models (security, abuse, holidays, sick pay) inside national and international networks and associations. For this case a developer, operator and local policy maker were interviewed.

The second case study is also based in Spain and was developed as part of the European Horizon2020 project SPROUT in the city of Valencia. The development and operations are performed with 'Ferrocarrils de la Generalitat Valencian' (local public transport agency) and 'Fundación Valenciaport' (the local seaport organisation). The use of lockers, provided by Citypack, for the last-mile delivery give operators more flexibility as it decreases the operational costs and reduce the failed home deliveries. To the city level, it reduces the urban freight vehicles occupying the public space and traffic jumps. In this project, public infrastructures such as metro stations opened up the adoption of this service to any commuter using the metro.

The third and last logistics service is a Flemish private locker service named Mobile Locker, providing very large variety of smart lockers. The idea was developed in 2012 and has grown to become a high-end provider of lockers for all types of events and locations. Contrary to many other providers, they can provide tailored lockers developed in collaboration with the customer, which can be a (local) government, an organization or a company in need of solving issues with their package flow. These lockers are permanent and provide, next to the logistic aspect, also some extra services, such as WiFi, cellphone charger on solar energy, etc. For the use of a locker, not even an application is needed. Just by scanning the QR (quick response) code, someone can use the service.

3.4.1. Interviews

For logistics services, the development of a **legal framework** is less clear compared to the mobility cases. There have been problems with the status of riders, their payment and the social and job protection they enjoyed while working (Gómez, 2020). Although some form of legal framework is present in some of the cities where these services are active, none of the interviewees could give a clear answer on the content of that framework. The Madrid policy maker claimed that there was a plan for a decent regulatory framework to protect the riders but could not go into detail about the content or when this framework might be ready. Except for a potential regulatory framework to protect the drivers, there is no such framework to protect the customers, no rules were mentioned in order to be able to access vulnerable to exclusion groups.

Ini the madrid case, the applications and service itself were developed by 'activists', within Coopcycle with a major focus on developing fair services for which they are active on two levels: they are lobbying on the European level, but also with local government. Because of its foundation within the political activist-scene, Coopcycle have developed their own framework to protect their riders and customers.

The Belgian policy makers did not mention the existence of a legal framework related to good delivery by lockers, they provide a rather supportive role without too much interference in the way the market is run. The developer of the Mobile Locker system claimed that, with the current status of the market, "The government should be involved as less as possible." but also stated that the lack of profits in the last mile is caused by the prices that are too low, which could be adapted with a regulatory framework developed on a nation-level. Inclusion wise there are some general rules, related to access of screens in public places for people in a wheelchair, but no recommendations of rules have been drafted from a digital exclusion point of view.

Generally speaking, no clear framework has been developed, although several of the interviewees stated there were plans (e.g. policy makers) for a regulatory framework in which only very few aspects related to digital inclusion were mentioned. That lack can be explained by the urge to create a decent service, which is in the first place accessible to the broader public. Secondly, and especially related to lockers, the profit in last and first mile delivery is negligible, so before a profit can be made on these deliveries, not many investments are going to be made for a more (digital) inclusive system.

Often closely related to the lack of a regulatory framework is the limited factors promoting (digital) inclusiveness. Efforts in order to develop a **digital**

inclusive service are very limited from a policy point of view, in both the Spanish and Belgian cases (and this is applicable in Europe) there were some efforts towards inclusion, mostly related to the accessibility of buildings and services as regulated by the government (e.g. screen displays should be accessible for people in a wheelchair).

Initiatives for a more digital inclusive system are not mandatory and not really promoted as well. Coopcycle, being a quite inclusive organisation, has some aspects they are working of in order to create a more inclusive service. Reservations are possible using the phone, cash can always be used for payment and when neighbourhoods are too far away to make delivery profitable, they try and work out a solution. This was corroborated by the Madrid policy maker, who acknowledged the problem, but except for acknowledging the issue, nothing much is happening to solve it. Based on the inputs from the intrviewees it can be concluded that there is a structural problem, rather than only the lack of funds and initiatives.

The lockers in Valencia are developed by the public railway organisation and are therefore located near train stations. The link between these stations and the lockers creates a more inclusive story. People with more limited mobility due to not having a car have more difficulty to combine trips and to carry goods and food home after work for example. The combination with lockers, where both food and goods could be delivered, not only provides the users with benefits, but also reduces stress on the public transport. If fewer (connected) trips related to shopping need to be made, there are less people using the public transport. Also, public transport does not favour one group over other, so the lockers are available to a broad spectrum of society. Similar to the one in Madrid, the Valencia policy makers the acknowledged that issues about excluded groups are known, but not much action is undertaken at this time.

For the smart logistics services, there is still a lack of rules, efforts and methods to make the services more (digitally) inclusive. Most of the efforts made towards this goal are own initiatives from the companies, rather than regulated by policy makers.

User involvement such as co-creations could be a potential solution for some of the earlier mentioned issues, keeping in mind that for a more elaborate and better regulated framework, user involvement solely with potential users will not suffice. A bottom-up approach, combined with close collaboration with developers, operators and especially the policy makers (both at local level and beyond) are necessary to achieve an inclusive regulated service. The issue here, is that during the development of the services, certainly the lockers, a very limited amount of co-creation is part of the development process.

During the interviews, co-creation was only mentioned a few times even though it is embedded in the development of the Coopcycle service, which is co-owned by everyone participating within the business and therefore is submitted to a lot of different viewpoints. The user involvement part is mainly based on the input from the riders, who were underrepresented before this service was developed. Before and during the development of the lockers in the city of Valencia, potential users were asked to provide input, but no further comments were made about this during the interview.

La Pajara have a very close relationship with the developers of their service (Coopcycle) which results in good communications and faster response to errors. “It is a good thing that the more users you have, the more they validate it (the app), the more it is tested and the more errors are seen.” As was said above, collaboration with other organisations is pivotal to the business idea of Coopcycle. For the development of the service, they worked together with an organisation specialised in promoting cooperatives. Their vision on this form of cooperation was very positive and gave them the time to learn, create a network and discover alternative ways of thinking and working. Related to the lockers, co-creation is very limited or even absent. The lockers from Mobile lockers are developed with input during the prototyping, if the project is big enough.

The fourth main aspect is **data collection, protection & privacy**, which is very different for the bike delivery and the lockers. In the case of the lockers, the collected personal data in both cases is rather limited and mostly focused on the use of the lockers. The main goal is to advance the way in which the service works, which is done in two ways. First, general data is collected from the locker (time, use, which types of lockers, duration etc.), which is used by both providers, and is especially useful when the developer is in charge of the actual locker (e.g. when it is not bought by a government for example, but is installed on their own initiative). Secondly, and this was mainly the case for the lockers in Valencia, a lot of studies were performed to find out what the reach of the lockers was, meaning the area that was serviced by these lockers, to evaluate how profitable they were. The policy maker for Valencia said: “Since it was a data intensive project, but it was only me who used and had access to that data and it was for their benefit, it was not going to be published or anything.” Which led to the users providing a lot of information about the acceptance level of the lockers. Coopcycle was not that busy with data collection, their focus was primarily on the creation of a fair system for the drivers and customers.

3.5. Multimodal routeplanners and MaaS

For this part, focusing on multimodal routeplanners and Mobility as a Service or MaaS systems four cases are studied: the BKK FUTAR app Budapest, the HVV Switch app, used in Hamburg, the HSL MaaS app from Helsinki and Jeasy, a new MaaS developer active in Belgium.

The first case is a multimodal application for the city of Hamburg. It is a mobile app, named Switch, operated by Hamburger Hochbahn AG and the 'Hamburger Verkehrsverbund' (HVV), the local Transport Association that makes moving around Hamburg easier and faster. With this app it is possible to buy HVV tickets and book your ride with MOIA (local car sharing service) with just one registration. HVV is a company that coordinates the public transport system in and around Hamburg, Germany. The full application was only available from 2020 onwards, in German and English. The app does not have multiple digital payment options, the only one is PayPal.

The second case study took place in Budapest (Hungary) where the BKK FUTÁR service was studied. The application is a multimodal route-planning service mainly focusing on tram- and bus-use, but it can be used also for subway. For each of the modal options, real-time information has been integrated in the application and webservice, a main difference with the 'Switch'-app is the lack of an integrated payment option. The app is available in 2 languages, English and Hungarian, so for tourists this might cause some issues. A web-based version of the service is available as well, leading to more people having access to the service. The application provides real-time information about the mobility services and their stops, also information about ticket vending machines and kiosks is available.

The third service is HSL (Helsingin Seudun Liikenne or Helsinki regional transport) Public transport application used in the city of Helsinki and the surrounding areas. Helsinki is a leading example for multimodal and sustainable urban mobility and the HSL app is a good example of how apps can be developed and how other services are integrated: the app contains fully integrated real time information and payment systems (debit- or credit card or by using your phone bill). The application is not a full-option MaaS app because of its focus on public transport (e.g. in comparison with Whim from MaaS Global). The application is available in English, Finnish and Swedish.

The last service is the multimodal routeplanner Jeasy which is currently only available in the Belgian B2B market, focusing on multimodal commuting. Jeasy has developed an application that, based on preferences (e.g. shortest route, cheapest, environmentally friendly) from the user, presents a multimodal route for commuters. While they are currently only working with companies or organisations, there is already a beta-testing version for users available on the website. An application is in development but is not yet available. One of the strong suits of the Jeasy service is the integration of MaaS in the Belgian mobility budget.

3.5.1. Interviews

There are more barriers/drivers compared to the other topics because of the recent and fast development of MaaS-like systems, the large amount of data and information that is needed and because of strong debates about the development of new tools. Another point of discussion is the relationship between public and private initiatives. The cases also show that both private and public organisations have introduced multimodal routeplanners or MaaS or are at least on their way to do so.

The **regulatory framework** for routeplanners or MaaS varied strongly in different cities and between the organisations. None of the operators mentioned a strict regulatory framework which they had to take into account. For the three (semi-) public organisations (BKK FUAR, Switch and HSL) there were however some general rules, mostly focusing on providing a service for 'everyone', 'the general or broad public', although this was not always supported by other claims about their target groups: "…with 6 type of tickets you get 95% of all people satisfied. Their approach is in line with the comments about the lacking regulations about the expectations from the city, regions they are operating in, the lack of (inter)national regulations about data sharing and the lack of standards. HSL tries to solve part of this issue by releasing licenses for limited time with some aspects of control (similar to KPI's).

Inclusion seems to be quite the challenge for multimodal routeplanners and MaaS-systems. None of the interviewees could indicate that they were considering specific groups of people vulnerable to exclusion. However, the developer from Jeasy approached inclusive design in another way, as part of a stepwise evolution of their service, which is, from a financial point of view a logical choice: "…it is difficult to work with those groups (vulnerable to exclusion), as you said, we should include them. But, to make a service successful, you start with the biggest group.". Also, the stakeholder considered it very important to collect information about excluded groups and to broaden their knowledge about the excluded groups and how they can be approached. It's also clear that different stakeholders were focusing on different aspects of inclusion, Jeasy, working in the B2B market, wants all employees to benefit from the advantages of the mobility budget, in combination with MaaS, which could replace company-cars, so they were interested to see how generational differences would impact the preferences people had.

User involvement is present in a very limited manner for the development of routeplanners and MaaS-systems. According to the Helsinki policy maker it is a very competitive market where both private and public organisations are active. The first difference between both is the development. For Jeasy it is an ‘in-house’ development, for the other services the development and coding itself is carried out by specialised ‘third-parties’, which in some cases results in the developers having almost no input contentwise, next to only developing an app as ordered by the operator, and in other cases results in the lack of collaboration in the app development. It is clarified by the HSL policy maker why their involvement is kept to a minimum: “Because they can be different people this year, different people next year.”. Although this might change in time. In all the cases it was also clear that co-creation with potential end users was not really considered, rather they did interviews and surveys to see how many people were using a service, how often and how satisfied people were using the service: “We concentrated essentially on a few co-workers and then on the enterprises which were involved anyway in the development, thus not necessarily were regular users involved directly in the app development.”. Similar answers and remarks were made by the other stakeholders, Jeasy however, did claim they could provide API’s (Application Program Interface) to other organisations (e.g. representing vulnerable user groups) in order for them to use the MaaS application in combination with the app for the vulnerable users.

It also becomes clear that the speed of development of services is very different for the public service compared to the private sector. In the first group there are more stakeholders, more boards and development moves at lower pace compared to the private initiatives which was considered a barrier by the developer from Jeasy. Nevertheless, the developer considered the conversations with policy makers and city representatives as very useful. by the developer: “… every time we knocked on a door, we reached a discussion that was constructive, not always providing what we expected, but at least it was constructive, positive and so forth.” This gave the impression that the slower pace at which policies change related to these services had as a result that close co-creation was not a real possibility at the moment. For the public organisations in Helsinki, Hamburg and Budapest, there are many more conversations, discussions at hand before changes can be made.

All of the stakeholders mentioned similar aspects about user involvement, if it was present, it was mostly in a very limited capacity, and could rather be defined as collaboration between companies, mostly in the form of tenders or assignments. Multiple organisations also mentioned their interest to provide access to those people who were currently excluded. As confirmed by Jeasy: because of the newness, competition and volatility of the market the main focus in on the ‘masses’ in order to create a viable service, which can later on, stepwise, become more digitally inclusive.

Data collection, protection & privacy are very relevant in the case of multimodal routeplanners and MaaS-systems, similar to data collection, protection & privacy when providing door to door logistics services. In order for these services to work at their best, a lot of personal information is needed: preferences, skills, needs, expectations etc. This was also acknowledged by the stakeholders as described above. There is quite some data collection from the users in order to have a better tailored service, and, compared to the other services seems to be analysed more in-depth.

Another thing pointed out by the HSL policy maker is the need for shared data and information about routeplanners and MaaS-systems. This was also strongly confirmed by the Jeasy developer. When asked about European guidelines, almost all interviewees confirmed that if the EU wants decent MaaS systems, there is a need for standardised data sharing. Currently, data sharing is already present on a minimal level and is performed according to the EU-regulations, but in many cases this framework needs to be elaborated. In the Hamburg case the importance of data protection was also confirmed. However, at the moment this is still too limited and services are being developed in all countries and even cities: “What I do see is that countries are working alone, so they don't work with other countries and that within those countries, cities are working alone. Antwerp does not have the same expectations from the MaaS players than in Brussels.”.

Table 2: Synthesis results driver and barrers for an inclusive mobility serice

Type of service	Service name	Regulatory framework	Co-creation	Inclusivity	Data collection/ protection and privacy	Other relevant aspects
Car- & ridesharing	Cambio Brussels	A regulatory framework is present in most cases, usually on a national level. The framework is present due to stability market and is starting to push towards more inclusive services, but this is not mandatory in case of digital inclusivity.	Services are unknown among vulnerable to exclusion groups and not enough accessible communication. Rather info-events and introductions. Homogeneity of 'test-groups': highly educated and white. Post-development co-creation events are important.	Knowledge about vulnerable groups too limited Depending on service different attempts: call center, cheap, personal approach, test-events. Digitally contacting elderly is hard	Only mobility related data, with limited use, especially about vulnerable to exclusion users	Internal or third party app development influences flexibility. Users are suspicious towards new technologies. Lack of resources for research towards inclusion vulnerable people.
	Mobitwin				Elderly people have trouble understanding privacy related issues and don't understand importance of GDPR.	
Bike sharing and micro mobility	HIVE Lisbon	Regulatory framework is regional and based on carsharing framework, so not always suited to micro mobility. Knowledge about service among policy makers is too limited. Both hard or soft regulations are adopted to control micro mobility.	Lack of funding and instability of market make providers focus on 'easy to reach' users, in order to be financially feasible. More frequent and intensive communication between stakeholders and with users. In most cases there is a lack of communication with local authority.	Small steps are taken towards a more inclusive service, but no regulations are present to push towards more inclusivity. Need for smartphone and credit card in almost all cases. Link with public transport and use of one payment method	Data collection is very limited and is not really used for analysis. Contradicting vision on use of data: in detail or rather a general approach. Data is needed to address impact on public space	Impact on public space is major, affecting elderly people, blind people etc. Young and volatile market was severely impacted during COVID-19 crisis. Main focus on financial revenue.
	Brussels Mobility					
Smart Logistics services	La Pajara Madrid	No national regulatory framework for the protection of couriers, initiative by organisations themselves.	Co-creation during development with riders created more fair value sharing. All neighbourhoods are provided for without extra cost	Acknowledgement of need for more inclusive services, but lack of funding an knowledge. Similar to the regulations, there are no rules to make the service more inclusive. If inclusivity is not introduced by the developer/operator, it is not present in the service.	Only information relevant to delivery is stored, but not used for any other purpose.	Locker providers showed no intention for the development of inclusive services, rather focus on the financial aspects, this was not the case for the public provider.
	Mobile Locker Citypack Valencia Lockers	No consensus on need and purpose of regulatory framework among different services	Except for the location, there are not that many measures towards an inclusive service. No collaboration with users or other organisations present.		Information about efficiency/use of the lockers is collected and used to find most profitable location.	
Multimodal routeplanners & MaaS	HSL HVV Switch BKK FUTAR Jeasy	No relevant regulatory framework was present	Isolated development simultaneously in different cities Different development speed in private vs. public organisations. Closer cooperation with users necessary	No measures were mentioned by operators or developers. Public organisations develop services for all inhabitants, but this goal is not often reached. Service still unknown	Fear that sharing data will lead to advantage for competitors Lack of trust in public services handling data.	Public service, open data sharing with other cities is becoming more important. Unfair competition between private and public development

Source: Interviews and results from co-creation workshop

4. DISCUSSION

Several topics were discussed, based on which a number of common requirements, needs and challenges of policy makers, operators and developers that are relevant for all types of services were investigated. These barriers and drivers are related to the **market position of the services, the regulatory framework in place, the integration of the service into the public transport service network, the diversity of the vulnerable-to-exclusion groups, knowledge about the vulnerable groups, the level of user involvement (co-creation), the fast evolution of digital mobility services and communication and collaboration between stakeholders.**

The **market position** of a service is defined by how established an operator is in a local market, i.e. how long it has been operating, how many users it reaches, if it makes profits and what its position is compared to the competitors.In a more stable and predictable market, there is a long-term growth potential for service operators. This has been observed by the interviewed car-sharing operators. In contrast, if the market is volatile with many players appearing and disappearing in a short time period, and with uncertain profitability, the focus of the market players is short-term profit maximalization that may exclude any accessibility and inclusivity features.

The **regulatory framework** can set minimum service requirements in terms of accessibility and inclusiveness. It can also contain requirements for sharing service data, define key performance indicators and require collaboration between complementary services. Current frameworks in the case studies focus primarily on the regulations related to operators keeping to certain rules e.g. parking regulations, regulations related to use of public space, speed regulations etc., the topics of accessibility and inclusivity are not usually covered. The regulatory frameworks are often adapted to just one type of digital mobility service, usually the one that was introduced as the first one (usually car sharing). The frameworks are then extended to other types of services, but often not addressing the specificities of these new services (e.g. e-scooters).

Related to the regulatory framework, **the integration of the digital service into the public transport service network** may act as a major driver for accessibility and inclusion. Once a service is considered a 'public service', it needs to be available for all citizens, it needs to comply with minimum accessibility standards and eventually also with a regulated pricing system.

Knowledge about the needs and requirements of vulnerable to exclusion groups is limited among operators and developers. This aspect was observed among all developers and operators of the different services studied and it was also confirmed by the user group representatives. If knowledge about user groups with disabilities is available, it is about the needs of people with physical disablilities. These are also often the groups that are not reached for information events. Another issue linked with this is the lack of **diversity among groups participating in info- or co-creation events.** There is currently no or limited interest from the operators to explore why certain groups of society do not use their services. They are mainly interested in market expansion, focusing on target groups and areas that can provide sufficient profit margins.

Combining data from multiple operators and other sources (e.g. statistical databases of local authorities) could expand the possibilities for the analysis of users' needs and for the identification of non-users. Sharing of mobility data is a very sensitive topic. The fear for sharing data is twofold. Firstly, many private providers of digital mobility do not trust private and public organisations in terms of sharing their data. Secondly, many providers are afraid they will lose their market share to competitors when sharing information, so they all rather keep their data. This showed in the **lack of communication and especially collaboration between stakeholders.**

The level of user involvement (co-creation) in the development of new services remains a key issue. The importance of involving (potential future) users in the development of applications and services has been mentioned as a key requirement by most stakeholders. However, actual co-creation, i.e. the involvement of the users in the design of multiple aspects of a service or application is limited. User involvement can be very diverse. From the case studies about Coopcycle, HSL, Cambio and HIVE we identified several approaches which can act as **drivers** for creating more inclusive and accessible services:There are, however, also many barriers. A **barrier** to co-creation is that many potential users have very limited knowledge about the service, showing there is still a lack of decent introduction, explanation and guidance for groups who have no access to the digital transport system. It is possible that for **some groups** it is very hard to make the transition from physical to digital services, as was proven in the Mobitwin case.

5. CONCLUSION

We can conclude that there are a multitude of digital mobility services available in the current transport system. Although all four services are very different they struggle mostly with the same issues. The development of an inclusive service seems to be a difficult challenge depending on and influenced by many different factors. The carsharing services were the most inclusive, both from a digital and a physical point of view. This was mostly related to their stability and the approach they take in providing a fair service. Both the micro mobility and logistics services struggle with the concept of inclusivity, which they claimed had to do with their financial instability and with a lacking regulatory framework related to these topics. Similar to the carsharing services, most inclusivity measures were developed as part of their business plan. The routeplanners and MaaS were not inclusive as they were still rather at the start of their deployment process and were still looking for their market share. The main aspect that needs attention regarding these services is a regulatory framework with specific focus on shared data and the management of that data. The digital transport system has a lot of potential for the future, but as was stated in the introduction, futher development of the services will be pivotal to provide its service to all groups in society.

REFERENCES

[1] Belga. (2020). Deelsteps van Bird strijken opnieuw neer in Brussel. De Tijd. https://www.tijd.be/ondernemen/transport/deelsteps-van-bird-strijken-opnieuw-neer-in-brussel/10263011.html

[2] Brugha, R., & Varvasovszky, Z. (2000). Stakeholder analysis: A review. Health Policy and Planning, 15(3), 239–246. https://doi.org/10.1093/heapol/15.3.239

[3] Bates, R. H., Greif, A., Levi, M., Rosenthal, J.-L., & Weingast, B. R. (Red.). (1998). Analytic narratives. Princeton University Press.

[4] Belgisch staatsblad. (2016). Wet van 28 april 2016 betreffende de houdende voorwaarden voor het gebruik van voorbehouden parkeerplaats aan operatoren van gedeelde voertuigen. http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=nl&la=N&cn=2016042817&table_name=wet

[5] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Brussel Mobiliteit. (2017, December 14). Alternatieve mobiliteit. https://www.brussel.be/alternatieve-mobiliteit

[7] Cambio Flanders. (2020). Voordelen | Cambio autodelen | Vlaanderen. https://www.cambio.be/nl-vla/voordelen

[8] Crosby, B. L. (1991). Stakeholder Analysis: A Vital Tool for Strategic Managers. USAID Implementing Policy Change Project, 2, 6.

[9] Kedmi-Shahar, E., Delaere, H., Vanobberghen, W., & Ciommo, F. D. (2020). D1.1 – Analysis Framework of User Needs, Capabilities, Limitations & Constraints of Digital Mobility Services. 105.

[10] Eurostat. (n.d.). Internet access and use statistics—Households and individuals. Eurostat Statistics Explained. Retrieved 15 January 2021, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Internet_access_and_use_statistics_-_households_and_individuals&oldid=379591

[11] Eurostat. (2020a). Digital economy and society statistics—Households and individuals. Eurostat Statistics Explained. https://ec.europa.eu/eurostat/ statisticsexplained/index.php/Digital_economy_and_society_statistics_households_and_individuals

[12] Eurostat. (2020b). Individuals using mobile devices to access the internet on the move (Dataset No. tin00083).https://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tin0

[13] Eltis. (2020). Overview of policy relating to e-scooters in European countries. https://www.eltis.org/resources/casestudies/overview-policy-relating-e-scooters-european-countries

[14] Franckx, L., & Mayeres, I. (2016). Future trends in mobility: The rise of the sharing economy and automated transport-Annex A (Deliverable D3.3; Mind-Sets, p. 113). European Commision.

[15] Freeman, R. E. (2010). Strategic Management: A Stakeholder Approach. Cambridge University Press.https://doi.org/10.1017/CBO9781139192675

[16] Gómez, M. V. (2020, September 24). Spanish Supreme Court rules food-delivery riders are employees, not self-employed. EL PAÍS. https://english.elpais.com/economy_and_business/2020-09-24/spanish-supreme-court-rules-food-delivery-riders-are-employees.html

[17] Goodwin, K. (2013, January). A Stakeholder Interview Checklist. Boxes and Arrows. https://boxesandarrows.com/a-stakeholder-interview-checklist/

[18] Hinchliffe, T. (2018, November 28). mytaxi Launches 'hive' E-scooter Pilot Program in Lisbon—Portugal Startups. https://portugalstartups.com/2018/11/mytaxi-hive-e-scooters-lisbon/

[19] Kedmi-Shahar, E., Delaere, H., Vanobberghen, W., & Ciommo, F. D. (2020). D1.1 – Analysis Framework of User Needs, Capabilities, Limitations & Constraints of Digital Mobility Services. 105.

[20] Kedmi-Shahar, E., Delaere, H., Vanobberghen, W., & Di Ciommo, F. (2020). D1.1 – AnalysisFramework of User Needs, Capabilities, Limitations & Constraints of Digital Mobility Services (Deliverable D1.1; INDIMO, p. 103).

[21] Kranzberg, M. (1986). Technology and History: 'Kranzberg's Laws. Technology and Culture, 27(3), 16.

[22] Pope, L. (2020). Stakeholder interviews: What to ask and how to ask it. https://gathercontent.com/ blog/stakeholderinterviews-what-to-ask-and-how-to-ask-it

[23] Robinson, J., Acemoglu, D., & Johnson, S. (2003). In Search of Prosperity: Analytic Narratives on Economic Growth. Princeton University Press.

[24] Simons, H. (2009). Case Study Research in Practice. SAGE Publications, Ltd. https://doi.org/10.4135/9781446268322

[25] Schmeer, K. (1999). Stakeholder Analysis Guidelines. 48.

[26] STARS. (2018). The influence of socioeconomic factors in the diffusion of car sharing Deliverable D4.1 (Deliverable 4.1; Shared Mobility OppoTunities And Challenges For European CitieS, p. 243). European Commision.

[27] Wallemacq, T. (2019, May 7). De verborgen kant van elektrische deelsteps | Metro. De Verborgen Kant van Elektrische Deelsteps. https://nl.metrotime.be/2019/05/07/must-read/de-verborgen-kant-van-elektrische-deelsteps/

[28] Wick, A. (2012). Six Effective Elicitation Questions to Ask Your Stakeholders. BA Times. https://www.batimes.com/articles/six-effective-elicitation-questions-to-ask-your-stakeholders.html

IMPACT OF AUTOMATED VEHICLES ON ROAD SAFETY

Rins de Zwart
SWOV – Institute for Road Safety Research
Bezuidenhoutseweg 61, The Hague, The Netherlands
rins.de.zwart@swov.nl

ABSTRACT

The introduction of connected and automated vehicles (CAVs) has far-reaching impacts on the traffic system and society. The Levitate project (EU, Horizon2020, levitate-project.eu) aims to forecast those impacts. Road safety is one of the relevant areas of impact that is analysed in the Levitate project. CAVs influence road safety directly by replacing the human driver, and indirectly due to secondary impacts such as modal split and rate of travel. In order to forecast the road safety impacts of an increasing penetration rate of automatic vehicles several methods are combined.

Microsimulation is used to model traffic behaviour of automated and human driven vehicles in different road networks. Vulnerable road users (pedestrians and cyclists) are not included in the microsimulation model but are affected by the introduction of connected and automated vehicles. Additional safety effects due to a change in modal split are determined by estimating the changes in travel mode and it's effects on total accident numbers.

By combining the different approaches a more complete overview of the impacts can be made. The goal of this paper is to describe the road safety impacts of increasing CAV penetration in the traffic system as they are identified in the Levitate project.

Keywords:
Automated transport, Automated vehicles, Vulnerable road users, Safety

1. INTRODUCTION

The development and introduction of connected and automated transport systems (CATS) is expected to have far-reaching impacts on the traffic system and society as a whole. This poses policy makers with new opportunities and challenges for the integration of these technologies in a controlled and safe manner. The Horizon2020 project LEVITATE (levitate-project.eu) aims to aid policy makers by forecasting the societal, environmental and economic impacts caused by the rise and eventual widespread use of CATS. One of the main impacts taken into account within the LEVITATE project is road safety. This paper describes the approach taken within the project to determine the different road safety impacts and the results of this method.

Within the project the impacts on road safety are looked at from three different use cases: public transport, personal transport, and freight transport. These use cases offer different looks into the impact of automated transport on road safety. Further specification within the use cases is done by way of different sub-use cases that explore the impacts of specific implementations or interventions.

In order to forecast the impact of AV introduction it is important to take into account the moment and speed of the introduction. As it is exceptionally difficult to predict when AVs will become commonplace the project uses the market penetration rate of automated vehicles instead of different years. This way it is possible to fit the results to new predictions on when certain penetration rates are met.

The introduction of automated transport is expected to influence road safety in a number of different ways. Within the LEVITATE project a distinction is made between two types of impact: primary and secondary impacts. Primary impacts is an impact that directly results from the automation of a vehicle. Secondary impacts, also called rebound impacts, are generated by the primary impacts and feed back to the source of that primary impact.

Automated vehicles (AVs) affect road safety directly (primary impact) as they will have a lower risk of being involved in a crash compared to human driven vehicles. This effect becomes even greater when vehicles are able to communicate with each other, i.e. they become interconnected. In addition to these primary impacts some secondary rebound effects are to be expected. For example, it is likely that the total distance traveled and the modal split of travel change as a response to the availability of automated transport. Changes in modal split and distance traveled have the potential to impact the number of crashes.

The current paper aims to give a concise overview of the different methods used within the levitate project and describes the identified road safety impacts of an increasing penetration of CAVs.

2. METHODOLOGY

Within the project the safety impacts are looked at from three different use cases: public transport, personal transport, and freight transport. This distinction between use cases is made due to the different expected effects on road safety of the three different types of transport. A further specification within the use cases into so-called sub-use cases is done afterwards by looking at specific implementations or interventions. Safety impacts are first determined for the overall introduction of automated transport, with specification of impacts on a sub-use case level taking place in a later stage.

The project combines multiple different methodologies in order to quantify the impacts of AV introduction. For the road safety impacts a major part is played by traffic microsimulation. Traffic microsimulation entails the simulation of traffic in different situations, allowing for the manipulation of vehicle and environment characteristics. This enables the modelling of traffic behavior of automated and human driven vehicles in different settings. Results of the microsimulation allow for the determination of effects relating to surrogate safety measures such as time-headway and speed differences. As the microsimulation is limited to vehicles such as cars, trucks and busses, the vulnerable road users are not included. However, these are of course also part of the traffic system and are impacted by the introduction of automated vehicles. In order to include the safety impacts on vulnerable road users a separate method is used. Additional safety effects from modal split changes are not captured in the above mentioned methods and are thus determined separately by estimating changes in travel mode and the effects on total accident numbers. These three methods will be shortly discussed below.

In order to forecast the effects of increasing penetration rates of AVs a plausible implementation scenario is chosen. This scenario consists of different penetration steps that are chosen to represent the most likely steps of implementation over time. A distinction is made between two different types of automated vehicle, cautious vehicles and ambitious vehicles. These are chosen in order to represent technical developments over time resulting in more advanced vehicles at a later stage. Cautious vehicles represent automated vehicles that do have full self driving capabilities but do not drive closer together than human vehicles, instead choosing a similar time-headway. Effects of a shorter reaction time and never waning attention are present in these vehicles. Ambitious vehicles represent the later stages of implementation where penetration rates are high. These vehicles follow much closer together than human driven vehicles, choosing time-headways that would be considered unsafe for human drivers. Table 1 shows the different deployment scenarios used within the project.

Scenario A represents the current composition. There are no fully self-driving vehicles available so penetration rates are 0%. Scenarios B and C represent the first introduction of cautious AVs, with a quicker introduction within freight transport. From scenario D onwards the share of cautious AVs does not further increase. Instead the share of ambitious AVs increases until all human driven vehicles are replaced with automated vehicles. Scenarios G and H represent the final stages of implementation where older more cautious AVs are replaced with later more ambitious versions.

Table 1: AV deployment scenarios

Type of Vehicle	A	B	C	D	E	F	G	H
Human-Driven Vehicle - passenger vehicle	100%	80%	60%	40%	20%	0%	0%	0%
1st Generation (Cautious) CAV - passenger vehicle	0%	20%	40%	40%	40%	40%	20%	0%
2nd Generation (Aggressive) CAV -- passenger vehicle	0%	0%	0%	20%	40%	60%	80%	100%
Human- driven - Freight vehicle	100%	80%	40%	0%	0%	0%	0%	0%
Freight CAV	0%	20%	60%	100%	100%	100%	100%	100%

2.1. Microscopic simulation

As human behavior modelling has improved it became possible to also use microscopic traffic simulation to determine impacts related to road safety (Guido, Vitale, Astarita, & Giofrè, 2019). By using well calibrated models for human behavior and utilizing model inputs that are known to represent real world safety performance at given locations a prediction on impacts can be made. This allows for the estimation of different impacts such as speed, traffic volume and congestion from an increasing penetration rate of automatic vehicles. By comparing different sets of situations with each other the impact of different parameters can be determined. It should be noted that microsimulation is not a perfect representation of the real word situation, assumptions have to be made. However, by comparing different situations with each other it is possible to work around the disadvantage of assumptions as they will influence all situations the same.

In order to evaluate the microsimulation results with regards to road safety, a surrogate safety analysis is performed using SSAM. This produces a number of conflicts based on vehicle trajectories. The number of conflicts is then converted to a number of crashes using a crash-conflict ratio based approach. The end result allows for a comparison of the number of conflicts and crashes in a before situation with numbers in alternate situations. Information about distance travelled and where the traffic is located are also output from the microscopic simulation. This enables the comparison of different implementation scenarios and interventions by comparing crash data per kilometer travel. It is important to note that while driving behavior of AVs is different, travel behaviors such as amount of travel of all vehicles or time of travel do not change just by the introduction of automated vehicles in the microsimulation model. Effect on travel behavior related to behavioral changes are excluded from the microsimulation model. In layman terms the microscopic simulation replaces human driven vehicles one-to-one with automated ones, reducing factors such as reaction time, following distance, speed variation. It does however not change the number of vehicles that want to travel or the distance those vehicles would travel. These changes are collected in the modal split paragraph.

Within the Levitate project the microsimulation environment AIMSUN (www.aimsun.com) is used with a model based on the city of Manchester.

2.2. Vulnerable road users

The interactions of vulnerable road users with automated vehicles are complex and require a carefully chosen approach in order to ensure safety for these users for whom a collision would be even more dangerous than a car to car collision (Constant & Lagarde, 2010). Due to the inability of capturing safety impacts on vulnerable road users with the microsimulation environment the project relies on a different two pronged approach to estimate the relevant safety impacts. Using accident statistics based on data available from Austria an analysis is performed to determine current accident causes. This allows for the separation of accident causes between VRUs and cars that could be prevented by AVs and those accidents that can not.

For the accidents that can not be prevented by AVs the assumption is made that there is the potential for a reduction in accident severity based on the lower reaction times of automated vehicles compared to human driver reaction times. Using this reduction in reaction time it follows that the braking action of an automated vehicle starts earlier than that of a human driven vehicle, resulting in a lower speed at the time of impact. By comparing fatality rates at the different impact speeds an estimate can be made about the safety effects of improved reaction times for those accidents that can not be fully prevented by replacing the human driver with an automated one.

2.3. Modal split changes

The above mentioned methods are used for determining safety effects for replacement of human driven vehicles with automated ones and the effects on vulnerable road users. The effects relating to changes in mobility and driving behavior are estimated using a third approach concerning the modal split changes.

Determining the impact of modal split change on road safety is done by combining information about the exposure of different travel modes with information about the risks of different travel modes. The introduction of automated vehicles is expected to change the balance between private, public and active transport. In order to determine the changes in amount of travel for the different travel modes a combination of a mesoscopic model and a system dynamics approach is used. This model utilizes agent-based modelling to simulate people completing normal daily tasks while different options for transport are available. By modifying the settings and parameters it is possible to determine what changes occur in the travel mode distribution. An increase in travel with a certain mode translates to an increase in exposure for that mode. This change in exposure can then be combined with the risk of that type of transport.

The introduction of automated transport also impacts the risks associated with different kinds of travel. Reduction of human driver error will also reduce the risks for all road users that interact with human drivers. In order to capture this change in risk a comparison is made between the current risks of different transport modes and information about the new risks coming from the microscopic simulation and vulnerable road user methods.

Multiplying the change in risk rate of a certain mode of transport by its change in share of the modal split gives an estimate of the change in road safety impact. When this is done for all modes of transport a new total road safety impact can be determined. A distinction is made between four different modes of transport within this approach: human driven vehicles, cautious CAVs, ambitious CAVs, vulnerable road users, and other transport which consists of busses and trucks. A two-step approach is taken which consists of a separate VRU determination and motorized transport determination, which are then combined for a total impact on safety. The end result allows for a comparison of the different modal splits that occur at varying market penetration rates of CAVs.

3. RESULTS

Within the Levitate project a distinction is made between two different kinds of impacts with relation to road safety, direct and indirect impacts. A direct impact is a direct result for the automation technology being implemented, while an indirect impact occurs as a rebound effect or behavioral adaptation. More information about all impacts considered within the Levitate project can be found in Elvik et al. (2019).

Automated vehicles have a direct effect on road safety due to a lower crash risk than the traditional human driven vehicles (Fagnant & Kockelman, 2015; Logan, Young, Allen, & Horberry, 2017). This is especially the case when vehicle to vehicle communication is added and the vehicles become connected (Malone et al., 2014). In addition to these direct effects there are indirect effects expected. It is likely that the modal split of travel and the amount of travel are effected by an increasing penetration rate of automated vehicles, and it is known that modal split and the distance traveled have an impact on the number of crashes.

3.1. Direct impacts

At the core of automated transport is the goal of replacing the human driver with a automated system that allows the vehicle to drive itself. Most crashes involve the human driver making an error at some point, even if this does not always mean that the driver is at fault or to blame. By moving the driving task away from the human it is expected that automated vehicles will reduce the effect of these driver errors in crashes. While it is still unclear exactly what percentage of human driver errors can be prevented by introducing CAVs, it is clear that sufficiently advanced vehicles should be able to prevent most (Fagnant & Kockelman, 2015; Logan et al., 2017).

However, CAVs also have the potential to introduce new risks. Equipment or software might fail due to the many different parts involved, software malfunctioning or cyber security problems. Other issues might occur with the AV not being able to detect or recognize other road users and infrastructure such as road signs and markings due to poor visibility in the case of bad weather or temporary lane markings confusing the system. In more general terms, human drivers are likely better at dealing with unexpected situations. When the automation technology is not yet fully capable of handling every situation it might be necessary to transfer control back to the human driver in certain situations. This brings an increase in risk due to the transition of control that has to happen, often suddenly and with minor warning (Zhang, de Winter, Varotto, Happee, & Martens, 2019).

The most important direct impacts are discussed below in more detail.

3.1.1. Human driver error

Even the most experienced drivers make mistakes. Human drivers sometimes disobey the traffic rules and make unintentional mistakes. These mistakes and violations increase the risks of a crash occurring. When fully automated vehicles take over the driving task from the human occupant the risk of a driver error is eliminated.

Automated vehicles are assumed to always follow the rules of the road and will therefore reduce or eliminate the crashes that are a result of disobeying the traffic rules. Additionally, human issues such as tiredness, distraction and lack of situational awareness will likely be things of the past when vehicles can drive themselves. As a final point it is expected that automated vehicles will be able to respond much quicker than human drivers and will be better equipped for detecting other road users due to a multitude of cameras and sensors. These improvements are even bigger when vehicles are able to communicate between each other and become connected.

3.1.2. Transitional effects

While fully automated vehicles are able to perform the driving task in all situations there will likely be a period of transition where less advanced vehicles are being used. These less advanced vehicles will be able to perform the task most of the time but require the human driver to either pay attention or take over control when the system can no longer keep up. This transitional period has the potential for decreased road safety as unfamiliarity with a imperfect system can result in an increase in human error. These errors are most notable when the human driver assumes the vehicle can drive itself when it can not (mode confusion) or when the vehicle indicates the driver needs to take control back (take-over requests).

When a take-over request occurs the human driver has to stop what they are doing and resume manual control of the vehicle. This is a demanding task. Literature reports different amounts of time needed to perform a take-over safely, varying between less than a second and more than 20 seconds (Zhang et al., 2019). This time is influenced by what the driver was doing at the time of the request, prior experience, type of take-over request and capabilities of the vehicle. Studies show that reaction times increase when automation can no longer perform the driving task (Strand, Nilsson, Karlsson, & Nilsson, 2014), and an increase in collisions when requests for take-over are presented (Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014; Rudin-Brown & Parker, 2004). These results indicate that take-over requests pose a risk for road safety. However, due to the complexity of take-over requests and the incomplete knowledge about real-life situations the actual impact is as of yet unclear.

3.1.3. System security and failure

As with any vehicular system, failure of the system can occur. In addition to the normal systems present in vehicles, AVs introduce a multitude of sensors and software for object recognition. This has the potential to increase the risk of a system failure resulting in the vehicle being unable to detect an object or other road user.

Even when the vehicle systems perform well there is the potential for outside interference. As AVs are computer controlled they introduce so-called cyber risks. These risks become even more prevalent when vehicles have the ability to communicate with each other and have connection capabilities. Some of the current vehicles on the road already have the ability to communicate with outside devices and are therefore at risk of a hacking attempt. The industry has not yet been willing to significantly invest in cyber security in order to prevent outside attacks on vehicle software.

3.2. Indirect impacts

Next to the direct impacts of AV introduction, indirect impacts are also present. These impacts occur due to changes in factors that in their turn impact road safety. People could for example change their travel behavior due to the availability of automated vehicles.

3.2.1. Driver behavior adaptation

As the share of automated vehicles on the road increases other road users such as cyclists, pedestrians and human driver might adapt their driving behavior to the automated vehicles. Human drivers might adopt the smaller time-headways that the AVs are using when surrounded by these vehicles. Research shows that this can already occur when driving next to a platoon of freight vehicles driving close together (Gouy, Wiedemann, Stevens, Brunett, & Reed, 2014). As automated vehicles strictly adhere to the traffic rules a discrepancy might occur between AVs and human driven drivers. This could result in unwanted behavior when a human driver has to adjust their speed or overtake the vehicle due to the AV’s strict adherence to the speed limit. A large scale study performed in the Netherlands shows that 2 out of 3 test drivers with a mandatory speed assist experienced negative interactions with other drivers (AVV, 2001). This is likely due to the inability of the test drivers to exceed the speed limit, frustrating the drivers behind.

In addition to human drivers, cyclists and pedestrians might also change their behavior when confronted with an AV. Research shows that pedestrians are currently not confident in the capabilities of self-driving vehicles to detect and stop when approaching a pedestrian (Rodríguez Palmeiro et al., 2018). However, no difference in pedestrian crossing behavior was found between vehicles that were explicitly identified as self-driving compared to those that were not (Rodríguez Palmeiro et al., 2018). A similar experiment designed for cyclists shows no differences in crossing behavior either (Vlakveld, van der Kint, & Hagenzieker, 2020). An overview of the current knowledge gaps related to pedestrian and cyclist interactions with automated vehicles can be found in (Schagen, Kint, & Hagenzieker, 2017).

3.2.2. Travel changes

The introduction of automated transport is likely to influence the decisions people make concerning travel. These changes might be due to higher availability of motorized transport, the ability to travel longer distances without the need for the driver to rest, or a lower valuation of travel time because the driver is able to engage with other tasks than driving. There is also the potential of vehicles making empty kilometers, in order to return home and be available for others or in order to avoid parking fees. Interventions have the potential to influence the amount of travel generated directly. For example by limiting access for AVs or implementing pricing on empty travel.

Generally speaking an increase in travel results in an increase of exposure. This increased exposure results in an increase in the number of crashes. It is also relevant to know where the travel takes place, as the risks of different road types can differ greatly. Changes in route choice might occur due to the connected nature of automated transport enabling the sharing of information about congestion. There is evidence from current navigation systems influencing the route choices and impacting road safety due to increased travel on roads not meant for through traffic (Cabannes, Fighiera, Ugirumurera, Sundt, & Bayen, 2018).

3.2.3. Modal split changes

As automated transport becomes more readily available it has the potential to greatly effect the split between different modes of transport. The ability to use an AV without a driver license, do work while traveling and automated options for urban transport and goods delivery all influence the modal split. As cars become a more attractive choice the usage of public transport or active modes of transport such as walking and cycling declines. As crash rates differ between the modes of transport it is important to know what the changes to the modal split are and how these impact the overall road safety.

4. DISCUSSION

As the introduction of CAVs into the traffic system is still years away from reality there are a number of assumptions that have to be made within the project. These assumptions are made based on currently available literature and expert opinion but are still assumptions that might prove to be incorrect at a later date. It is assumed that CAVs will be able to adapt their behavior to their specific conditions in order to minimize risks. CAVs are expected to always respect traffic rules and have a positive impact on road safety. The scope of this paper is limited to effects of CAVs on road safety, additional effects of CAV introduction related to societal and environmental issues is not discussed but does have the potential to effect the implementation scenarios greatly.

The project uses different methods to create a complete overview of the road safety impacts of CAVs. However, the different methods utilize sources of information from different locations. For future projects is might be valuable to focus on a single location in order to allow for easier comparison. The approach used in the current paper does allow for wider generalization.

5. CONCLUSIONS

The Levitate project aims to determine the impacts of automated vehicle introduction. Road safety is one of the important aspects to take into account when looking at automated vehicles. The impacts on road safety are determined using multiple different methods that allow for the inclusion of city wide traffic simulation, vulnerable road user impacts and travel choice analysis. This paper describes the methods used within the Levitate project to determine the different road safety impacts and provides an overview of identified impacts.

Impacts related to road safety are split into direct impacts and indirect impacts. Direct impacts represent the impacts that are direct effects of the usage of automated vehicles as opposed to human driven vehicles. Reductions in reaction time, deviations from traffic rules, distraction and mental impairment are expected when humans are no longer performing the driving task. However, in the transition toward fully automated vehicles there might be a period where the human driver serves as a fallback for the automation system. Sudden need to take over the driving task from the automated system poses a risk to road safety. Additional risks could occur due to the possibly of cyber attacks on the automated vehicles.

Indirect impacts relate to effects that occur in response to the direct impacts. Changes in driving and travel behavior due to automated transport availability are the most important indirect impacts. While human drivers share the road with an increasing number of automated vehicles it is expected that human driver behavior will adapt some of the automated driving behavior. Automated vehicles are likely to drive closer together in order to maximize road capacity, something that is possible only due to improved reaction times. When this behavior is copied by human drivers it has a negative effect on road safety. Next to expected changes in driving behavior there are also changes expected in the choice around travel. The amount of travel is likely to change with the introduction of automated vehicles. This could for example be due to the easier access to travel when a driving license is no longer needed. In addition to changes in the amount of travel there are also likely changes in the split between ways of travel. The modal split is likely to move away from active modes of transports and move towards a bigger proportion of single vehicle travel such as cars.

The general approach to road safety effects as described in this paper enables policy makers and safety experts to get a clear overview of expected safety improvements and areas that could be of concern, allowing for adjustments to be made on time in order to guarantee safe implementation of automated vehicles.

REFERENCES

[1] AVV. (2001). Evaluatie Intelligente SnelheidsAanpassing (ISA): Het effect op het rijgedrag in Tilburg.

[2] Cabannes, T., Fighiera, V., Ugirumurera, J., Sundt, A., & Bayen, A. (2018). The impact of GPS-enabled shortest path routing on mobility: a game theoretic approach.

[3] Constant, A., & Lagarde, E. (2010). Protecting Vulnerable Road Users from Injury. PLOS Medicine, 7(3), https://doi.org/10.1371/journal.pmed.1000228

[4] Elvik, R., Quddus, M., Papadoulis, A., Cleij, D., Weijermars, W., Millonig, A., Vorwagner, A., Hu, B., & Nitsche, P. (2019). A taxonomy of potential impacts of connected and automated vehicles at different levels of implementation Deliverable D3.1 of the H2020 project LEVITATE.

[5] Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice, 77, 167-181. https://doi.org/https://doi.org/10.1016/j.tra.2015.04.003

[6] Gouy, M., Wiedemann, K., Stevens, A., Brunett, G., & Reed, N. (2014). Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers' longitudinal control? Transportation Research Part F: Traffic Psychology and Behaviour, 27, 264-273. https://doi.org/10.1016/j.trf.2014.03.003

[7] Guido, G., Vitale, A., Astarita, V., & Giofrè, V. P. (2019). Comparison Analysis between Real Accident Locations and Simulated Risk Areas in An Urban Road Network. Safety, 5(3). https://doi.org/10.3390/safety5030060

[8] Logan, D., Young, K., Allen, T., & Horberry, T. (2017). Safety Benefits of Cooperative ITS and Automated Driving in Australia and New Zealand.

[9] Malone, K., Rech, J., Hogema, J., Innamaa, S., Stefan, H., Van Noort, M., & Gusrafsson, D. (2014). DRIVE C2X Deliverable 11.4 Impact assessment and user perception of cooperative systems. http://personal.ee.surrey.ac.uk/Personal/K.Katsaros/papers/DRIVE_C2X_D11.4_Impact_Assessment_v1.0_full_version.pdf

[10] Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How Traffic Situations and Non-Driving Related Tasks Affect the Take-Over Quality in Highly Automated Driving. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58, 2063-2067. https://doi.org/10.1177/1541931214581434

[11] Rodríguez Palmeiro, A., Kint, S. v. d., Vissers, L., Farah, H., Winter, J. C. F. d., & Hagenzieker, M. (2018). Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. Transportation Research Part F, 58, 1005-1020. https://doi.org/https://doi.org/10.1016/j.trf.2018.07.020

[12] Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): Implications for preventive strategies. Transportation Research Part F: Traffic Psychology and Behaviour, 7, 59-76. https://doi.org/10.1016/j.trf.2004.02.001

[13] Schagen, I. v., Kint, S. v. d., & Hagenzieker, M. (2017). Zelfrijdende voertuigen: wat betekent dat voor fietsers en voetgangers?

[14] Strand, N., Nilsson, J., Karlsson, M., & Nilsson, L. (2014). Semi-automated versus highly automated driving in critical situations caused by automation failures. Transportation Research Part F Traffic Psychology and Behaviour, 27, 218-228. https://doi.org/10.1016/j.trf.2014.04.005

[15] Vlakveld, W., van der Kint, S., & Hagenzieker, M. P. (2020). Cyclists' intentions to yield for automated cars at intersections when they have right of way: Results of an experiment using high-quality video animations. Transportation Research Part F: Traffic Psychology and Behaviour, 71, 288-307. https://doi.org/https://doi.org/10.1016/j.trf.2020.04.012

[16] Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. Transportation Research Part F: Traffic Psychology and Behaviour, 64, 285-307. https://doi.org/https://doi.org/10.1016/j.trf.2019.04.020

A NOVEL GLOBAL ARCHITECTURE FOR MANAGING PEDESTRIAN FLOWS IN TRAIN STATIONS

Capucine-Marin Dubroca-Voisin, PhD student
LVMT, Ecole des Ponts ParisTech
Boulevard Newton, Champs-sur-Marne, France
capucine-marin.dubroca-voisin@enpc.fr

ABSTRACT

This paper addresses the problem of managing increasing pedestrian flows in metropolitan rail stations, in the context of decreasing financial resources. Based on research carried out both in academic and corporate contexts, the paper proposes a novel architecture including data collection, management levers activation, collaborative policy-making, efficiency evaluation and agent formation. The functional architecture is adapted into two potential systems, a simple one designed to test and validate the principles, and a more mature one. The application of these systems to an actual station is then explained.

In urban rail mass transit systems such as Paris' RER, passenger demand has been heavily growing in the past years, while the transportation offer struggled to develop. This resulted in an increased concentration of passengers in the major stations. Reached flows and densities became difficult to manage for the train and stations operators, who lack theoretical and organisational tools.

This paper aims to contribute to improving the organisational side, by proposing a novel digital architecture. In such a solution, data could be collected, fused and analysed in order to activate pedestrian management levers. Once they have been activated, collected data would be used once more to assess the efficiency of the used strategy. Finally, a core part of the architecture would be to use this process as a pedagogic framework, enabling an efficient formation for involved operator agents.

Once functionally described, the system is adapted to the case of Bibliothèque-François-Mitterrand station in Paris.

Keywords:
pedestrian management, stations, architecture, pedestrian flows, railway traffic

1. INTRODUCTION

In a pre-covid world, travellers flows in the metropolitan cities were mostly increasing. These flows concentrate in high-frequented rail lines and stations, leading in pedestrian accumulation, with the potential risk of crowd movements. In urban rail mass transit systems, it can be difficult to manage the pedestrian flows for the railway operators, especially if they are much more familiar with the historical railway procedures than with crowd control.

In the meantime, it is complicated to develop the transportation offer or to extend the stations, as these developments are extremely expensive while the public finances are largely constrained.

This paper proposes an architecture aiming to help railway and station operators to deal with these pedestrians flows, in a context where operators try to develop data-collection and real-time management, but they lack theoretical and organisational tools.

This paper starts by a brief presentation of our our use case station, Bibliothèque François Mitterrand, located in Paris. We then present the difference elements which constraint both the flow management itself and the development of a flow management system.
In a third part, we provide a brief overview of systems already proposed by research teams.
The fourth part focuses on the data collection techniques, those available in most stations and in our use case station.
Inspired by these examples, we propose our own functional architecture. We apply it to the case of Bibliothèque François Mitterrand station, and propose a technical architecture fitted to the local needs.

That paper was conceived during a PhD contract both in the LVMT research laboratory and in the French national railway operator, SNCF. It then benefits from an inner perspective, but may be biased due to the same reason.

2. THE BIBLIOTHEQUE FRANÇOIS MITTERRAND STATION

Bibliothèque François Mitterrand is a station located in the southwestern Paris, in a neighbourhood with recent urban developments, important density and mixed functional usage (national library, cinemas and shops, offices, university, housing).

In its three underground levels can be be found two important public transport routes:

- At the level -3 can be found the metro 14 line, opened in 1999. This automatic line is the fastest of the Parisian metro lines, and passengers are able to reach the center of Paris in less than 10 minutes. It operates at a very-high frequency (with 90 seconds intervals) and offers an impressive capacity of more than 30,000 passengers per hour and per direction (pphpd).
- At the level -1 can be found the RER C, a suburban line running via Paris and opened in 1979 (while the Bibliothèque François Mitterrand itself was opened in 2000). RER C capacity is over 40,000 pphpd, but the intervals are much higher between trains, with an average around 3 minutes between trains on peak hours, and 5-15 minutes for the same stations served. Bibliothèque François Mitterrand is the last station within Paris and constitues a pivot-station between a slow, omnibus Parisian section of the line, and a section served by both express and local trains, with potentially high speed, up to 140 kmph.
- The intermediate -2 level is a vast concourse, with accesses to both metro 14, RER C, and most of the station exits. While vast, it is not used by travellers for waiting, as they prefer to reach the platforms immediately. In the mean time, only two of three available platforms are used by the RER C (one in each direction). These platforms, despite their recent construction, are relatively narrow.
- This situation leads to different pedestrian flow issues in the station:
- The difference of intervals between Metro 14 and RER C leads to an accumulation of passengers coming from the metro on RER C southbound platform, particularly on the evening peak hours.
- This problem is reinforced by the unequal distribution along the 200 m-long platform, most people staying at the end of the platform to minimise their walking at destination.
- There is also an unequal distribution between stairs and the different escalators, with a saturation of the middle escalator, which leads itself to the most crowded part of the platform.

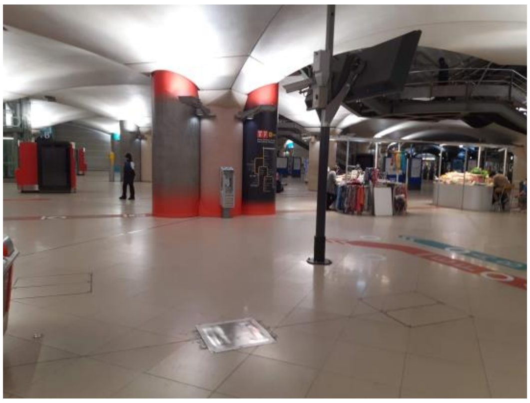


Figure 1: Concourse of the Bibliothèque François Mitterrand in 2020.

3. ISSUES AND CONSTRAINTS FOR PEDESTRIAN FLOW MANAGEMENT WITHIN AN URBAN RAIL MASS TRANSIT STATION

We’ve identified different key issues for pedestrian flow management, basing on an analysis of the station system, ground observations and interviews with stakeholders.

Flow management is deeply constrained by different elements :

- Stations are part of a complex railway system. Flows of trains must not be disrupted by flows of passengers, but safety must be guaranteed at all times.
- Moreover, railway systems offers a very low flexibility : it is difficult to change the train times and even more the intervals, as it is a timetable based system.
- Few human resources are present in the stations, and they are not specifically trained to flow management.
- It is very difficult for station managers to have the global picture (line or network state), and it is very difficult for line managers to know the current state of the stations.
- Data collection is very limited by legal constraints, and the station itself is a very challenging environment.
- Data predictions are also challenging, due to unpredictable events having huge impacts.

Flow in station are a mix of continuous processes and discrete events.

Building a system that helps train station operators to manage flow is also complex:

- It needs to be reliable, while data collection quality can hugely vary.
- It needs to be understood by the different actors , while flow phenomena are themselves challenging to understand.
- It also needs to be quick, as the decisions have to been taken in real-time.
- While the pedestrian levers themselves can be relatively cheap, the system itself must not be too expensive, as investment capacities in transportation are limited.

4. PROPOSED SYSTEMS IN RESEARCH

Several research teams have already proposed different architectures for pedestrian flow management systems. We've found three of them particularly relevant to the needs we identified.

In (Wagoum et al. 2012), a simple architecture is identified for crowd management in real-time. Using automated counting and a safety/security management system, it uses a simulation to display key indicators and a visualisation of the crowd. The originality of this work is its ability to show than a real-time simulation is usable in an actual case.

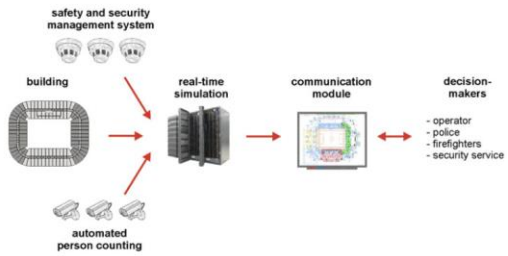


Figure 2: Simple architecture using real-time simulation (Wagoum et al. 2012)

(Martella et al. 2017) have identified two loops of crowd management by interviewing 10 Dutch professionals of crowd and flow management, including stations. This paper underlines the importance of planning, with an output constituted by scenarios and plans who are used in decision making. Data collected is then collected between events to improve the planning itself. The second loop takes places within the event itself, with the execution of action plans and the measure of the crowd.

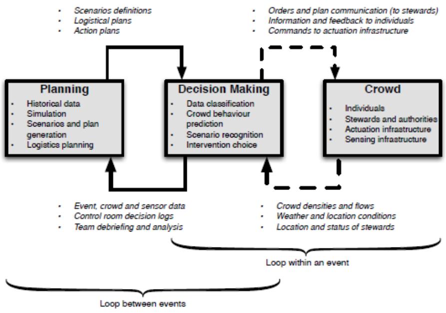


Figure 3: Double loop of crowd management during events, proposed in (Martella et al. 2017)

The most advanced system proposed in research seems the one described in (Molyneaux et al. 2020). Adapted from road traffic management systems, it proposes an estimation of the current state of the station based on several datasources. This estimation is used for a prediction of the future state of the system. Both are used to compute KPI, on which rely the decision making, with is determined by a control policy.

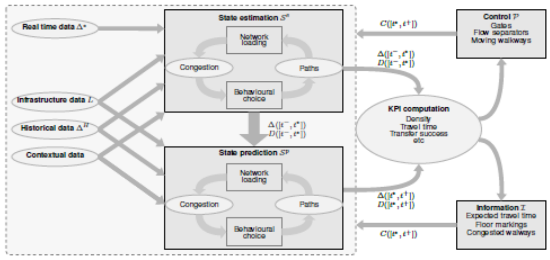


Figure 4: Pedestrian traffic management system from (Molyneaux et al. 2020)

While the double loop system highlights the importance of humans and a learning process, the two others are more focused on a technical approach, showing the complexity of flow management.

There could be a need for integrating human learning in technical solutions, that's the reason why we propose a different architecture in this paper.

5. AVAILABLE DATA COLLECTION TECHNIQUES

Several techniques can be used to collect data about pedestrian movements in stations. For instance, the system introduced previously and described in (Wagoum et al. 2012) used camera automated person counting (APC), which is a very common technique.

We want to present here a very brief survey of such techniques, as it's useful to know them for proposing a global architecture. Most of these techniques are rapidly evolving, so this survey has not any pretension to furnish a complete state-of-the-art. We will simply propose a rapid overview, then see which see which of these techniques are available in the case of Bibliothèque François Mitterrand station.

5.1. Brief overview of the techniques

An important part of data collection about pedestrian movement in stations is automated pedestrian country (APC). In this category, we make an important distinction between country the number of pedestrians in a space (the stock of passengers) and the number of pedestrians passing a line (the flow of passengers). Measuring density can be included in the first category.

Despite the huge importance of pedestrian counting, other measurements can be made, such as travel times, speed, trajectories or even qualitative observations about the crowd behaviour.

We consider that the sources used for pedestrian data collection collection can use three kinds of footprints: physical ones, hybrid (both physical and digital) ones and purely digital ones.

5.1.1. Physical techniques

Use of cameras is the best known and probably the most used technique. Automated image processing has known great process in the past years, notably by using artificial intelligence tools such as neural networks in important networks, and thanks to the improvement in calculation capacity. It enables both stock and flow pedestrian counting, but it’s hard to deduce one from the other (Duives et al. 2018). One interesting things with this technique is that stations often already have an expansive video collection system, even if the use of this system can be restricted for legal reasons. However, the position of cameras in stations is generally suboptimal for automated counting, notably with problems of occlusion and luminosity variation. Associated technologies can also be expensive. An improvement of the technique is the use of 3D cameras, or thermic cameras.

A common simple technique is the use of infrared sensors, but in the context of stations it can only count the pedestrian flow in very narrow spaces, due to the importance of occlusion in that case.

Some other physical techniques include counting floor carpets (measuring the pressure from walking pedestrians), CO_2 concentration measurement (Han et al. 2013). or ultra wide band counting (Yang et al. 2019). The latter two techniques do not seem sufficiently mature for a use out of the field of research.

Manual data counting and observation also belong to physical techniques. If humans are rapidly overwhelmed when the number of pedestrians grows, they are difficultly replaceable for qualitative observations. Notably, presence of agents in station may be used to collect such observations, but that may need some training and normalisation of the crowds behaviours to observe.

5.1.2. Hybrid techniques

These techniques involve both the physical presence of the pedestrian in the station, and for they to use or possess a device. These are mostly dependant of the possession rate of the device.

GSM footprints are not very precise and cannot help to measure pedestrian flows at the scale of the station, but they can be useful to determine the flow at the scale of the network, and then at least know the origin-destination pairs in station (Mun et al. 2008).

WiFi (Kurkcu et Ozbay 2017) and Bluetooth (Liebig et Wagoum 2012) can be more precise and help to determine the number of pedestrians in a space, and even the trajectory using several hotspots or beacons. They’re heavily dependant on the possession and activation rates, quite high for Wifi but lower for Bluetooth, and potential interference.

Another very commun hybrid source is the automated fare collection (AFC). Ticket gates are very common in stations, and very precise as the only errors are linked to fraud or the dysfunction of the equipment, generally short due to the crucial role of AFC in operators revenue. They provide interesting data about pedestrian flow but only at some points of the station, and are not always designed for real-time data transmission.

5.1.3. Digital techniques

By using massively digital devices, travellers can produce, voluntarily or not, digital footprints. Some of these footprints can be used to predict flow in station, for example using journey requests on a calculator and deducing the demand at the station. GPS tracking could also be used, but can be as imprecise as GSM footprints, needs to be shared by the user, and are frequently not available in underground spaces, generally the most crowded spaces in stations (so the ones needing pedestrian management and data).

Another technique we characterise as digital is crowdsourcing : lots of passenger themselves will report their own observations about the flow. Therefore, this is actually a physical technique, but with the point of view of the data collector, this is a purely digital technique. It has already been largely implemented by tech companies and train operations, especially for on-board load.

A last possibility is to use social networks to detect abnormal situations. Watching the networks could not be very useful in normal times, as users are not very keen to share information about a banal situation in a situation, but on the contrary it could be very useful to be alerted when an abnormal situation is reported on these networks.

5.2. Available techniques in the case of Bibliothèque François Mitterrand station

Some on the previously cited techniques are available on our use case station, Bibliothèque François Mitterrand. The stations as has an extensive cameras network, who can be used for counting the number of passengers along the different platforms, in the concourse. This camera-based counting system has already been deployed previously in the station (Dubroca-Voisin & Bertaux 2020). This system could also be used to count the flow of pedestrian within the stairs and the escalators, and we will use this possibility in our proposed architecture.

Wifi hotspots are available in the station and can be used to have the global of passengers within the station. As it is less precise than the camera system, redundant, and for the sake of simplicity, we will not consider this data source for this paper.

The station is fully equipped with ticket gates at its three entrances. In 2020, the existing gates started to be replaced with new ones, able to transmit in real-time the number of pedestrians at each gate (even if for some reason the gate is open). For the sake of simplicity again, we will consider that the entire station has been equipped with these new gates.

Therefore, we will use two main real-time data sources for our proposed architecture at Bibliothèque François Mitterrand: camera counting for both flows and stocks of passengers within the station, and ticket validation for entrance and exit flows.

6. AN ARCHITECTURE INCLUDING DATA COLLECTION, DECISION MAKING AND HUMAN LEARNING

In this paper, we want to propose an architecture which includes technical elements and human learning as well. This architecture is schematised in the Figure 5.

Four categories of humans are involved in this proposal. First, the data manager is responsible of maintaining the quality of the data collected. Both automatic and manual data collection (including crowdsourcing) can suffer from different biases, errors and problems. An human intervention can be very useful to check and potentially correct this data when processing it.

Levers managers have a key role in managing the flows: they decide of not to activate the different flow management levers available. To do, they

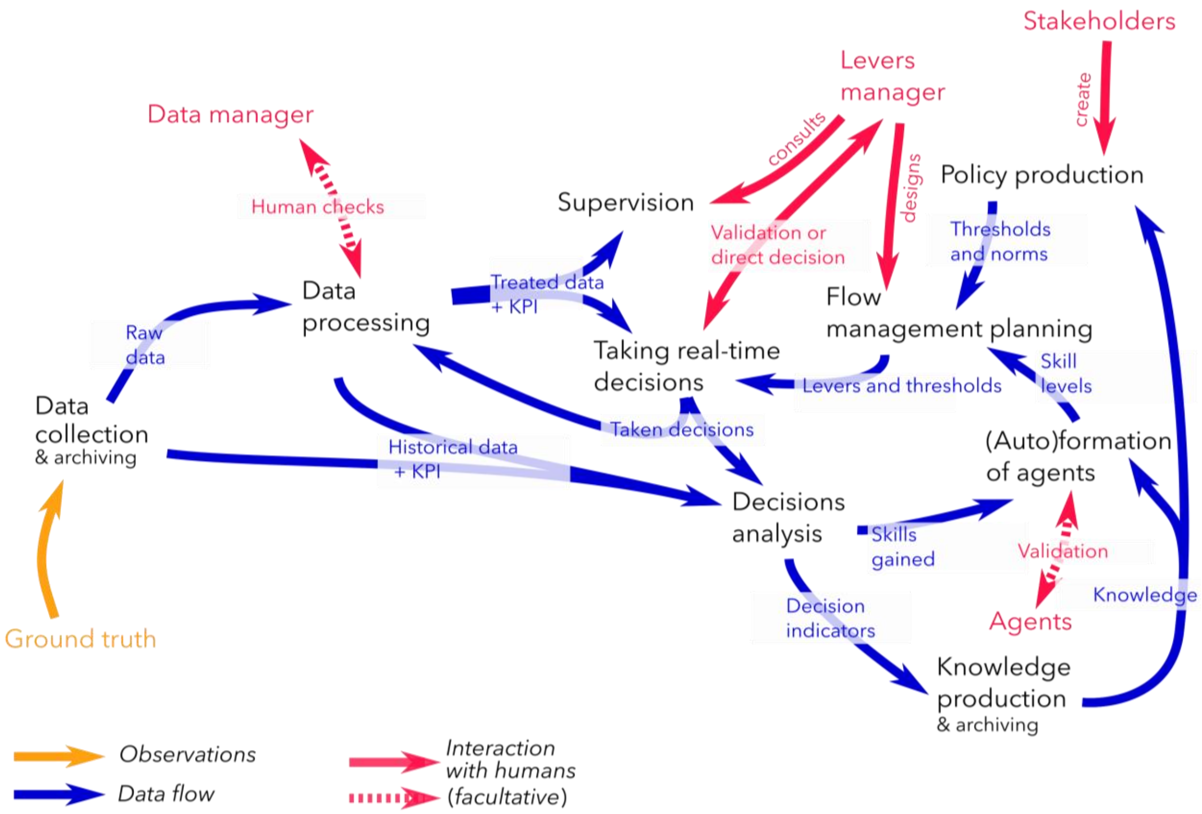


Figure 5: Global proposed architecture of the system, including human-data interactions

consult the flow supervision systems, which provide them useful KPI and flow visualisation. They're able to approve the decisions proposed by the system or to take their own decisions. They help to apply a flow management plan, but they also take a key role in designing this plan, using their experience. This plan is dependant on the policy, which can be forged taking into account the desires from the different stakeholders (train operating companies, public regulators, passengers, etc.).

Finally, all agents, including levers managers but also ground flow managers, need training. The skills gained can be tracked by an adequate system.

These humans interact with data in different states.

Data is first collected, using different techniques such as image analysis, weight counting, numerical footprints detection.

Data is then processed to ensure its consistency and make it usable for KPI and visualisations. This step could involve real-time simulation, as proposed in (Wagoum et al. 2012), but simple fusion techniques could be sufficient to ensure a sufficient quality.

The processed data is used for both supervision systems and decision making. These decisions are tracked, which is important for future data processing.

The flow management plan is a key data provider for decision making, which appeals to very structured plans. It it itself base on the policy and an available skills.

Once the flow has been managed, the analysis provides new knowledge, which needs to be encapsulated in a knowledge management system. It therefore influences policy production itself.

As this system is quite conceptual, we've decided to apply it the actual case of Bibliothèque François Mitterrand station.

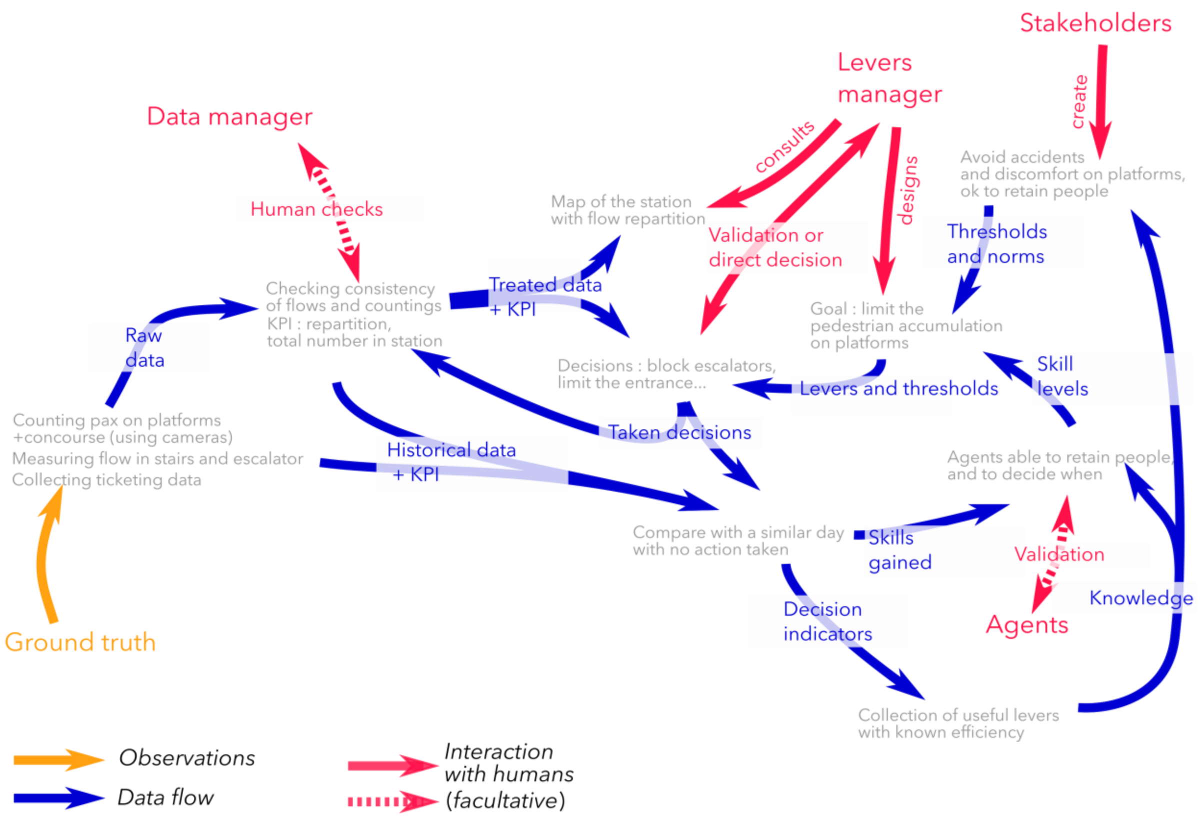


Figure 6: Functional architecture applied to the case of Bibliothèque François Mitterrand station

7. APPLICATION ON BIBLIOTHÈQUE FRANÇOIS MITTERRAND STATION USE CASE

Bibliothèque François Mitterrand station offers an interesting use car, due its current pedestrian issue described above, and its relative simplicity. To apply our proposed architecture, we first want to identify the local stakeholders then to propose relevant KPIs for this case. Finally, we propose both functional and technical architectures for that case.

7.1. Local stakeholders

The station is managed by a local team of the train operating company: station staff, train drivers or ticket inspectors work for the same company. However, this company also outsources pedestrian flow management on the southbound platform at peak hours, creating a second team with different working conditions.

As of 2020, there is no specific staff known as 'levers managers' or even 'flow manager'. The function is however taken by the station staff manager who's present at the station. He's a key stakeholder as he will have the mission, in our system, to monitor the situation, participate in the flow management planning, and decide in real-time which levers to apply.

He will be in direct link with the agents who will enforce his decisions, give him direct feedback, etc. and also with control centres (mainly the railway operation center of RER C, the RATP line 14 control center, and potentially the signalling cabin) to exchange information.

Other stakeholders will mainly be involved in previous steps, notably elaborating the flow management policy. The stakeholders notably include the passengers, who could have a key role in helping to determine which kind of levers and regulations are acceptable or not.

7.2. Relevant KPIs

As we explained above, there are three main problems in the station : accumulation of passengers in the southbound platform, unequal repartition on the platform and unequal use of stairs and escalator. We can define three immediate KPI:

- K1: number of passengers on the southbound platform.
- K2: number of passengers who would need to move along the platform to have an equal repartition between every section of the platform. Unequal repartition could also be represented by indicators such as Gini coefficient, but using number of passengers makes the indicator both more directly understandable and sensitive on the number of pedestrians on the platform.
- K3: proportion of passengers using the stairs, the middle escalator and the last escalator.
- However, some other KPI could be useful for the levers manager:
- K1.1: number of minutes before reaching the saturation threshold at the current pedestrian flow rate. This indicator could be useful to alert the manager than an action is needed.
- K1.2: maximal density on the platform. This could indicate a local dangerous situation even if the number of pedestrians on the platform is still normal.
- In the proposed system, the precise KPIs to use, their ideal values and the alerts threshold should be defined by the different stakeholders during the policy production process. In our example, avoiding accidents and ensuring the best confort possible for customer are global goals, while retaining pedestrians for doing so is an acceptable lever. Some global KPIs could also be used; they're not KPIs dedicated to the real-time operation of the system, but instead to ensure the global efficiency of the system, such as:
- K4: number of accidents linked to pedestrian crowding (the target value for this value should of course be 0).
- K5: average level of comfort experienced by each traveller, for instance by using Fruin level of service scale.
- K6: total duration when pedestrians had to be retained (as this is an acceptable but not desirable solution).

These KPIs are important for the decision analysis, knowledge production, who could then be used for policy production and formation.

7.3. Functional architecture

We so applied the global architecture to the case of Bibliothèque François Mitterrand station. This application is summarised in Figure 6, using the same structure than Figure 5. Each one of the processes is detailed.

Data collection is made using cameras counting passenger stock on each section of the platforms and in the concourse, and passenger flow at stairs and escalators, while ticket gates are counting entrance and exit flows.

Data processing uses two modes of verification : ticket gates data is used to check consistency of the cameras data, while human checks are also performed by the station staff. This enables to compute KPI such as those described in the previous section.

The supervision process could here take the form of a map, showing the flows and the stocks of passengers in the station, along with the relevant KPI. This map would be updated frequently and consulted by the station staff manager to help him take decisions.

These decisions would need to be communicated and saved, ideally using the same tool. They would be made to ensure the goals defined by flow management planning, and particularly the main one, which is to limit the pedestrian accumulation on platforms.

As the levers are not always activated, a good way to perform an analysis is to compare data between two days, one where the levers were activated and one where they weren't. This comparison will help to determine which levers are the most efficient and thus gaining knowledge.

Producing and obtaining this knowledge is an important part of the formation and training of the agents, including the station staff manager who's responsible for flow management. This can either be relatively informal or supported by a tool.

All this functional architecture must be completed by a technical one, in order to be applied properly.

7.4. Technical architecture

In this last section, we propose a very simple technical architecture for the case of Bibliothèque François Mitterrand station. This solution is voluntarily simple if not simplest in order to propose a first cheap and feasible solution for the stakeholders. However, this solution still needs to be complete and provide all the needed tools.

Our technical solution is schematised in Figure 7. The schema is slightly modified to take into account the fact that the data and levers managers are 'technically' the same person.

The two data collection systems are complicated themselves and could be described in depth, but the use of proprietary solutions reduce the complexity o the global implementation.

Data is then collected via APIs and computed trough a small custom solution. The implementation itself is not complicated but the algorithms can be, their ability to ensure the data consistency is key in the success of the system.

This data comes to a global tool, enabling to monitor (via map and KPIs) but also to interact by indicating if data looks consistent with human observations or not, but also the decisions taken.

Decisions and data are stored with simple CSV databases, enabling analyses made on notebooks. This feeds a knowledge database listing the different levers used and their effects.

Policy and planning are made simply using unstructured documents, describing the global policy wanted by stakeholders and its local application to the station. This is technically transformed into a config file, necessary for processing data and computing KPI.

A simple technical solution would git to the functional needs at the station. Its most complicated elements are the map and data processing, due to the need of precise algorithms and very powerful visualisation.

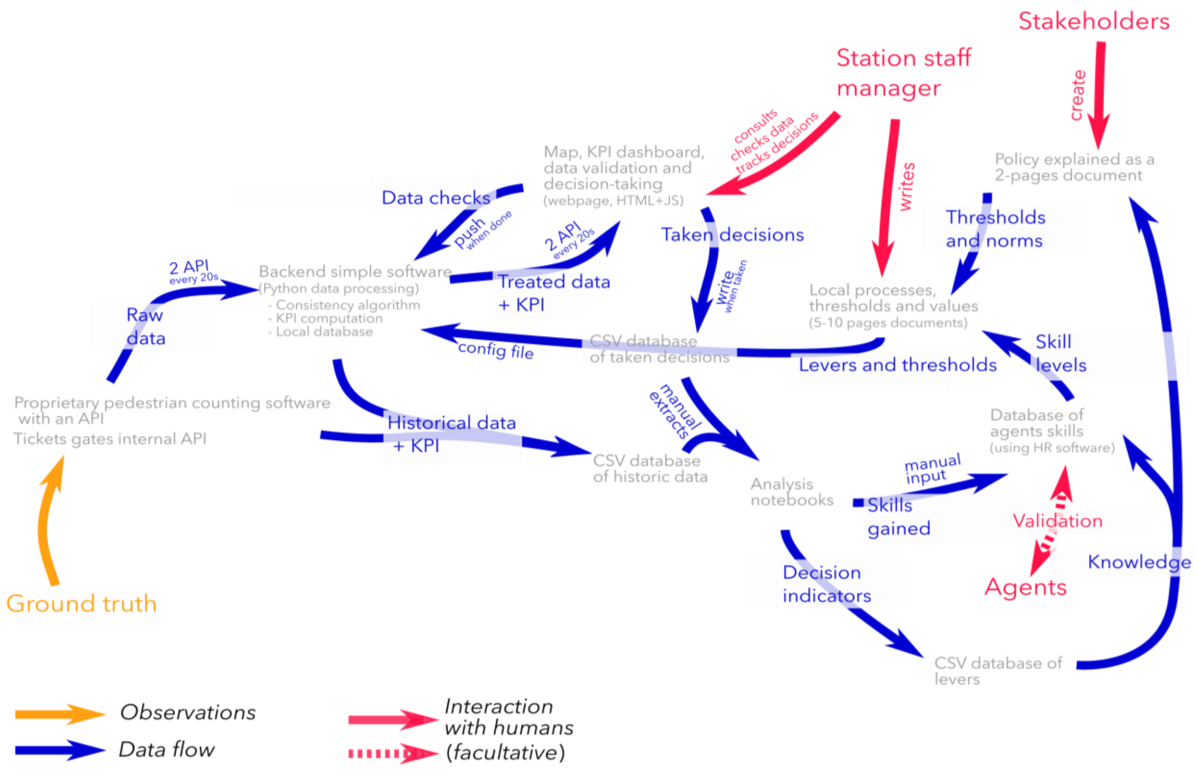


Figure 7: Architecture applied to the case of Bibliothèque François Mitterrand station

8. CONCLUSION

Constraints are high when designing a pedestrian flow management system. In this paper, we've tried to propose a novel architecture taking into account both technical constraints and necessity of training.

This architecture relies on efficient data collection, for instance with automatic pedestrian counting using cameras or automated fare collection. It is then necessary to propose a monitoring system, which favours taking adequate decisions. Those needs to be traced to enable analysis and so the production of knowledge concerning the levers at the station.

We've applied this architecture to a simple case, Bibliothèque François Mitterrand station in Paris. This station has recurrent issued, and we proposed a simple data collection system to help tackle these problems.

REFERENCES

[1] Dubroca-Voisin, C.-M., & Bertaux, P. (2020). TCO Voyageurs : une visualisation des voyageurs au services des voyageurs (Revue générale des chemins de fer, pp. 62–73).

[2] Duives, D. C., Daamen, W., & Hoogendoorn, S. P. (2018). Monitoring the Number of Pedestrians in an Area: The Applicability of Counting Systems for Density State Estimation. Journal of Advanced Transportation, 1–14. https://doi.org/10.1155/2018/7328074

[3] Han, H., Jang, K.-J., Han, C. & Lee J. (2013). Occupancy estimation based on CO2 concentration using dynamic neural network model. Proceedings of the 34th AIVC - 3rd TightVent - 2nd Cool Roofs' - 1st venticool Conference , 25-26 September, Athens 2013. https://www.aivc.org/resource/occupancy-estimation-based-co2-concentration-using-dynamic-neural-network-model

[4] Kurkcu, A., & Ozbay, K. (2017). Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors. Transportation Research Record: Journal of the Transportation Research Board, 1, 72–82. https://doi.org/10.3141/2644-09

[5] Liebig T. and Kemloh Wagoum A. (2012). Modelling microscopic pedestrian mobility using Bluetooth. Proceedings of the 4th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART, ISBN 978-989-8425-96-6, pages 270-275. DOI: 10.5220/0003833802700275

[6] Martella, C., Li, J., Conrado, C., & Vermeeren, A. (2017). On current crowd management practices and the need for increased situation awareness, prediction, and intervention. Safety Science, 381–393. https://doi.org/10.1016/j.ssci.2016.09.006

[7] Molyneaux, N., Scarinci, R., & Bierlaire, M. (2020). Design and analysis of control strategies for pedestrian flows. Transportation. https://doi.org/10.1007/s11116-020-10111-1

[8] Wagoum, A. U. K., Chraibi, M., Mehlich, J., Seyfried, A., & Schadschneider, A. (2012). Efficient and validated simulation of crowds for an evacuation assistant. Computer Animation and Virtual Worlds, 1, 3–15. https://doi.org/10.1002/cav.1420

[9] Yang, X., Yin, W., Li, L., & Zhang, L. (2019). Dense People Counting Using IR-UWB Radar With a Hybrid Feature Extraction Method. IEEE Geoscience and Remote Sensing Letters, 1, 30–34. https://doi.org/10.1109/lgrs.2018.2869287

ACCEPTANCE AND USABILITY OF AUTONOMOUS VEHICLES OVER THE LAST YEARS

Olga Maria Giannakari, Civil Engineer MEng
Hellenic Institute of Transport
6th km Charilaou-Thermi Rd., Thessaloniki, Greece
giannakari@certh.gr

ABSTRACT

Autonomous Vehicles (AVs) constitute an emerging area of research and development in the automotive industry. Automation in vehicles refers both to automated applications and features, such as the autopilot, Adaptive Cruise Control (ACC) and automated parking, and to fully autonomous systems where a physical driver is not necessary (SAE L4-L5 automation). So far, fully automated shuttles have been deployed in some countries, whereas several automated features are deployed in passenger vehicles already available on the market. In parallel, user acceptance of such systems is being studied, aiming to capture the end users' needs, wants and priorities concerning automation in transport. Undeniably, these features and technologies are being developed having as ultimate aim to enhance traffic safety and decrease consecutive accidents and deaths. Yet, this does not appear as an adequate reason for convincing people to accept and use autonomous vehicles. Users seem to be reluctant, mainly due to the lack of trust towards a vehicle operating without a supervisor/driver.

In this paper, a review of several surveys on user acceptance is being conducted, aiming at investigating users' opinions and identify – where relevant - the source of their lack of acceptance towards AVs. Moreover, the usability of autonomous vehicles is being addressed, tackling issues that are commonly understated, such as the possible increase in mobility, as a consequence of the wide accessibility of AVs, as well as the expected increase in repair and replace costs of these automated, advanced safety systems.

Keywords:
autonomous vehicles, automated features, user acceptance, usability

1. ACCEPTANCE OF AUTONOMOUS VEHICLES FOR PRIVATE USE

Autonomous vehicles (AVs) are emerging in our lives. Specifically, driving assistance functions are already being integrated in new vehicles, such as automated pilot, park assist, cruise control and more. Vehicles are also increasingly connected to the internet and are constantly engaging digital technologies, which are moving away from the pure "electro-mechanical" world. All of these new features and functionalities aim at providing safer, more accessible, cost and fuel efficient travels, while contributing to a more sustainable environment. Although the general public is adapting in these changes slowly, they seem to be particularly hesitant when it comes to fully automated vehicles or even vehicles operating with a minor assistance from the driver. State of the art technology, artificial intelligence and machine learning are used to equip autonomous vehicles, while communication infrastructure, such as 5G, G5 and IoT, is developing in order to reinforce connected vehicles. Automated and connected vehicles shape the future of transport, which is expected to be safer, based on the elimination of human error. Furthermore, AVs will be accessible to a wider audience, including elderly, children, people with disabilities that are incapable of driving a vehicle, and other vulnerable road users (VRUs). It seems that it will take some time to ensure passengers and drivers about automated driving, but the benefits will always outstand their worries.

1.1. Review of acceptance of private autonomous vehicles by end users

End users, such as passenger vehicles' drivers, passengers and VRUs, are seeking for improved, affordable and green means of transport nowadays. Electrified and automated mobility comes as a solution to a lot of everyday problems commuters face. Yet, people's main concern is road safety. Autonomous vehicles are designed to offer an unambiguous way of interaction between the vehicle and the driver/passenger. But does the technology and the infrastructure meet users' needs and wants? The prioritized needs and priorities from the part of the involved end users include the sense of perceived safety while using automated driving solutions. Automated vehicles are expected to improve road safety since they are designed to communicate with the environment, including pedestrians, other vehicles – equipped or not, VRUs, and to act quicker than a driver, whose attention may be destructed or his reaction time may be slower. Yet, autonomous vehicles as of today operate only with the assistance of a driver, at levels 2-4 as the Society of Automotive Engineers (SAE) has indicated. This fact raises new concerns about road safety such as risks of driver confusion/distraction, misuse of the systems and liability issues, which have to be addressed in order to ensure introduction and use of automated vehicles in our transportation system (Kulmala, Jääskeläinen, & Pakarinen, 2019). Furthermore, in order to encourage the acceptability of automated solutions and potential replacement of the current status of provided services, all of the end users, passengers, drivers, need to be properly informed about of this new era of automated and connected mobility. According to a survey conducted between the 28 Member States of the EU in 2019 and published in the Special Eurobarometer 496, "Expectations and concerns of connected and automated driving Europe" (Joint Research Centre, 2020), 58% of the people questioned, have heard, read or seen something concerning automated vehicles during the past twelve months. This percentage varies at national level, with the Netherlands and Sweden concentrating above 90% positive replies, while in Slovakia, Malta and Czech Republic positive answers are just above 50%. Another challenging aspect to look for is people's willingness to pay for autonomous vehicles (AVs) or services they provide. Surveys show that opinions differ depending on age and gender: younger people are more eager to use or pay for AVs or different services offered in AVs, such as communicating and productivity (Alonso Raposo et al., 2019). Also, male users are more open to use AVs than female ones (Alonso Raposo et al., 2019). Age is also a factor affecting the perception of accessibility in both the phase of design and implementation/operation. Internet illiteracy addresses most elderly people, let alone IoT and C-ITS services. Therefore, the deployment of AVs needs to be specially adapted to vulnerable road users, disabled, elderly etc. To begin with, Human-Machine Interface (HMI) is vital for the deployment of autonomous vehicles operating with the assistance of a driver. HMI has been developed in autonomous vehicles operating at SAE L2-L4, where a driver operating or supervising the operation of the vehicle is needed. This interface also assists the communication between the driver/operator of the vehicle, the other road users and the authorities (e.g. the police) (Kulmala, Jääskeläinen, & Pakarinen, 2019). The use of adapted HMI aims to mitigate risks that can occur in driverless or assisted driving modes. In the context of the EU project Adas&Me[1], an HMI application stated is detecting when the driver is not capacited to drive, due to fatigue, stress and distraction. Particularly, a survey was conducted to assure that the developed systems are able to accurately detect when the driver/rider is not capacitated to drive, and consequently mitigate these states and avoid dangerous situations through the use of adapted HMI and automation mode (Cocron, et al., 2018). The evaluation focused on:

- verifying the effectiveness of the systems to recognize the driver's state,
- attesting the capacity of the HMI to display clear and unambiguous information,
- evaluating the driver behaviour following a system warning/suggestion,
- collecting the driver's opinion on the system's usability,
- knowing the driver's trust and acceptance levels regarding the ADAS&ME functions.

In an automation-ready survey conducted for the project CoExist[2], which aims at preparing the introduction of the automated vehicles, the preliminary results showed that respondents believe that the most relevant actions to take are the development of regulations for CAVs (96%) and the definition of responsibilities regarding data management (85%). Surprisingly, the adaptation of local mobility plans (56%), the involvement of citizens (56%) and capacity building (40%) are the measures considered the least relevant to prepare for the introduction of CAVs on our roads (CoEXist, 2019). According to Alonso Raposo et al., four key game changers are shaping the future of road transport: automation, connectivity, decarbonisation and sharing. Preliminary studies on user willingness to use (or pay for) AVs seem to reflect an overall positive acceptance of these new systems. Gender and age differences appear: male users seem to be more willing to use AVs than female counterparts and young people tend to show greater willingness to use or pay for AVs compared to elderly people (Alonso Raposo et al., 2019). Users have also expressed willingness to pay for different services offered in AVs, with those relating to communication (e.g. social networks) and productivity ranking highest in their ratings compared to, for example, entertainment-related services (Alonso Raposo et al., 2019). However, a

[1] https://www.adasandme.com/

[2] https://www.h2020-coexist.eu/

significant portion of the population still has a negative attitude towards driverless vehicles. On top of that, another JRC's report, "The r-evolution of driving: from Connected Vehicles to Coordinated Automated Road Transport (C-ART)" , states that overall users' acceptance towards AVs is slowly improving as people are starting to experience more and more with the technologies through the existing demonstration and marketing activities of the various manufacturers and technology companies (Alonso Raposo, Ciuffo, M., & Thiel, 2017). Furthermore, in the context of the British project UK Autodrive[3], the researchers at Cambridge University's Engineering Department and the Department of Psychology carried out a national survey of public attitudes towards self-driving vehicles, gathering 2850 responses. Results show that more than three quarters (76%) of those surveyed had heard of driverless vehicles, comparing favourably to driver assistance technologies that are already commercially available, such as adaptive cruise control (familiar to 40% of those surveyed), automated emergency braking (38%) and lane-keeping systems (34%) (UK Autodrive, 2017). Another survey conducted for the Eurobarometer report aims at assessing EU citizens' awareness of automated vehicles and their experience with automated or semi-automated driving functions; measuring attitudes towards driving in or interacting with automated vehicles on the road; evaluating citizens' willingness to purchase and use automated vehicles; determining what citizens expect with regard to automated vehicles. The survey was carried out in the 28 Member States of the EU, with a total of 27656 respondents from different social and demographic groups being interviewed face to face. A majority of respondents say they would not feel comfortable in a fully automated vehicle without the supervision of a human operator in it, but 7/10 would feel comfortable travelling in a fully automated vehicle with the supervision of a human operator in it. In the question "To what extend would you feel comfortable or not travelling in a fully automated vehicle with the remote supervision of a human operator?", 30% answered that they would feel not at all comfortable, while 31% would feel not very comfortable (Joint Research Centre, 2020). However, it is interesting that in a 2017 Eurobarometer survey, between 52 and 63% of users would feel uncomfortable "Being driven in a driverless car in traffic" (Alonso Raposo et al., 2019). This shows an encouraging aspect as the more aware people become towards vehicle automation, the more prone they are to accept it. Still, a relative majority of respondents are not in favour of the deployment of fully automated vehicles on roads, but the majority is in favour of the deployment of connected vehicles. Overall the results suggest respondents are not yet ready to fully adopt connected and automated vehicles. Another interesting survey was conducted for the EU project AUTOPILOT[4]. In the project's deliverable "User requirements analysis" one of the scenarios developed, addressed carsharing of an AV as a touristic experience. The survey was online and it was conducted in eight countries: Finland, France, Germany, Greece, Italy, Netherlands, the UK and Spain. Participants willingness to use the service was generally positive, with two thirds of them agreeing that they would use the service if it was available (Aittoniemi, et al., 2018). Furthermore, in the deliverable named "User acceptance assessment", another user acceptance evaluation was carried out for five European pilot sites, where four use cases were tested, using autonomous vehicles: Automated valet parking, Highway pilot, Platooning and Urban driving. In total, 199 public participants took part in user tests across the AUTOPILOT project. Some interesting results detected in all pilot sites were that a) the majority of the users (> 88%) found it (very) important to have the option to take back control over the vehicle at any time, b) information that makes the trip safer and comfortable, such as information about possible hazards or waiting time was evaluated as an important factor and c) requirements on customization options, especially that the information is provided in the own language, were also an important feature from the user's point of view. Overall, participants' concerns were about:

- the functionality of the system (worried about system failure e.g. in detection of objects, by hitting pedestrians),
- the unexpected or harsh brakings/accelerations,
- the otherwise uncomfortable driving style (e.g. cuts-in, lane change, jerks, swaying, slow speed, driving close to objects),
- the technical failure of the system (e.g. navigation, take-over, manual braking) and
- the properties of vehicle or automation/service (e.g. uncomfortable seat belts, turning of steering wheel, HMI, automatic gear shift).

Generally, although there were a lot of differences between the services tested and the way in which they were tested, the outcomes were rather positive and, in many cases, similar. Levels of acceptance varied, but there was no real rejection of the services (Aittoniemi, et al., 2019). Zoellick, J. et al measured : 1) acceptance of AV, 2) perceived safety, 3) trust, 4) intention to use and emotions associated with AVs surprise, fear, boredom and amusement upon a

[3] http://www.ukautodrive.com/

[4] https://autopilot-project.eu/

pilot test with 125 participants who have experienced a ride in an electric AV on a large clinic area in Berlin, Germany. Respondents had previously taken a ride with an AV on a private terrain along 2 campuses in mixed traffic. Results state that respondents were amused, surprised and not afraid after the experience. The results stand in contrast to critical comments addressing uncomfortable interior, slow driving and abrupt braking (Zoellick, Kuhlmey, Schenk, Schindel, & & Blüher, 2019). Another report on autonomous vehicles was conducted from a customer perspective from the research institute CAP Gemini. The survey, addressing over 5500 consumers from around the world and executives at 280 companies, from automotive OEMs to technology players states that automotive consumers seem to be ready for the future and for self-driving cars operating at SAE level 4. Consumers seem very clear about their expectations, concerns, and desires, and they are willing to pay a premium to receive them. In response, automotive and tech organizations are making significant investments in hardware and software and racing to test and pilot vehicles. The recommendations for accelerating the journey towards a self-driving future include: informing the consumers, understanding and reassuring them, building an ecosystem of services, developing of software skills. As results indicate, 52% share of consumers would prefer self-driving cars five years from now, while 56% would be willing to pay a premium of up to 20% over their current budget for a self-driving car (Winkler, et al., 2019).

1.2. Gaps, needs, wants and priorities of automation in private vehicles

As it is suggested by the surveys and reports reviewed, people are not ready to accept fully automated vehicles in their lives, either as drivers/passengers or as pedestrians in environments where autonomous vehicles operate. This hesitance is justified in the lack of trust towards a vehicle operating without a supervisor, as many issues accompany this situation. A critical issue raised is the assignment of the responsibility for the vehicle operation; whether this should be aligned to the vehicle manufacturer, the providers of the system or an insurance company that may be in charge of the vehicle. The same issue also applies to lower level of automation vehicles, that do equip a driver/supervisor to perform specific tasks while driving (SAE L4). It seems that a common and operational legislative framework needs to be developed prior to introducing semi and fully automated vehicles in order to promote both safety and harmonization of transportation and to reassure users of automated vehicles. The matrix in Table 1 summarises the findings relating to the needs, wants and priorities of end users, as well as the existing gaps in establishing autonomous vehicles for private use.

Table 1: Needs, wants and priorities of end users and existing gaps for automation in private vehicles

	Gaps	Needs	Wants	Priorities
1	Familiarise with automation	User friendly-HMI, Accessibl e,	Driver to be able to take control of the vehicle	Safety, Willingness to pay
2	Costumer training based on age groups (with VR/AR)			Environme ntally sustainable

It seems that one of users' main concern is a possible system failure in fully automated vehicles.

2. ACCEPTANCE OF AUTONOMOUS VEHICLES FOR PUBLIC TRANSPORT

Automation in public transport has long been along of the curve. Metro lines in London and Paris are operating autonomously for the past decades already. Automated Road Transport Systems (ARTS) though, are still developing. Automated vehicles in road transport are clearly more challenging to be developed and operate efficiently than track-based transport modes. Automation and connectivity are rapidly increasing in road transport and mobility, making every imagination of seamless, public transport feasible. Cities suffer from traffic congestion due to the excessive number of private vehicles and consequently travel time is high. Although cars offer a more pleasant and safe environment and a personalised commute, door-to-door, the vast number of them has negative impacts on the environment, causing air pollution, and people's social and psychological status, since they consume a lot of time travelling instead of having this time for personal pleasure. Public transport always aimed at alleviating this lifestyle, but failed to succeed many times. Autonomous vehicles are designed to be convenient, functional, accessible to all users, including VRUs, and fuel efficient since AVs used for public transport will be electric. Therefore, the deployment of autonomous vehicles for public transport is expected to lead to a reduction of vehicle ownership, with public transport becoming an efficient and convenient mode of transport for all, while providing a safe, automated, time saving option for commuting. Road transport and mobility of the future are characterized by increased automation and connectivity. User safety, energy consumption and efficiency, traffic congestion and drivers' and passengers' comfort and convenience are expected to be assessed by automation and connectivity. In public transport, autonomous vehicles can result in a cost reduction of approximately 50%. Instead of a driver in each vehicle, one person in a control centre may or shall monitor and, if needed, manoeuvre, several vehicles (Kulmala, Jääskeläinen, & Pakarinen, 2019). At the same time, new technologies and features, such as sensors, wireless communications (G5), 5G networks, IoT and big data will be largely developed and will provide new business models for the automotive sector and their cost is expected to decrease. Therefore, plenty stakeholders are affected by this new era in transport. Public transport operators need to provide efficient, safe and connected transport with low cost. Cities promote automation as it constitutes a potential alternative solution to current and future problems, such as increased traffic congestion, while also providing for environmental sustainability, by increasing the use of public transport. Besides, a sustainable, smart city, comprising automated and connected public transport, not only buses, but metro and tram lines as well, is more appealing to residents and business. The combination of high-capacity public transport with automated car sharing services are expected to result in a considerable reduction of the average travel time comparing to using a private vehicle (Kulmala, Jääskeläinen, & Pakarinen, 2019). Also, by investing in the enhancement of public transport networks, systems and services, policy makers and authorities may accomplish the elimination of vehicle ownership. That is the reason why research institutes and academia focus their research interests in the integration of automated vehicles, in all aspects of road transport.

2.1. Review of acceptance of autonomous vehicles for public transportation by end users

Quite a few pilot projects have taken place across Europe. A questionnaire survey conducted in the city of Trikala, where an autonomous mini bus was operating at SAE level 4, for a period of six months, in mixed traffic in the city centre. The pilot was part of the CityMobil2 EU project. The questionnaire was addressed towards two different audiences; the first part of the survey was intended to passengers at the moment when they descended from their trip and the second one was a citizens' offline survey, during demonstrations were taking place. The first part of the survey, which gathered 200 responses from passengers, indicated that the mini buses were widely accepted and integrated in citizens' everyday life without highlighting any significant concerns regarding the safety and security arose. The second part of the survey among citizens, concentrating 519 responses, revealed that age is a significant factor when it comes to using the autonomous mini buses, as younger people were identified as regular users. Experienced with driving automation, younger people would prefer fully autonomous over manual driving and would use an autonomous vehicle (Portouli, et al., 2017). Similar to the shuttle bus in Trikala, another self-driving bus called Digibus, operating at SAE L2-L3, was tested on a public road in Austria. The shuttle from the French company Navya Tech has been tested on a 1.4km long track in the village of Koppl. The trial in Koppl was one of the first trials of autonomous public transport vehicles worldwide on public roads with mixed traffic in a rural area. The focus of this trial was on the real-world evaluation of a self-driving shuttle for bridging the first/last mile in public transport. A questionnaire survey was conducted to collect data on passenger experience. The majority of passengers were between the ages of 21 and 50. 15.1% were between 51 and 60, 9.6% were over 60. Just over 5% of respondents were between 13 and 20. Concerning employment status, 65.4% of passengers were employed, 11.0% were in training or were self-employed and 8.2% were retired. Almost 80% of passengers own a private car, while 20.5% of them do not. Almost 40% of those questioned, claimed they could imagine that a self-driving shuttle bus in their municipality could remove the need for a (second) car. Around 59% do not think a self-driving shuttle bus is a realistic option and may not replace a private (second car). Regarding prior knowledge on the topic of self-driving vehicles, 13% of passengers indicated they had none. Almost 43% had heard of self-driving vehicles, 44% were more familiar with the topic. This high percentage is due to the fact that many of the test passengers were part of company delegations who had relevant knowledge in this area. For the majority of passengers (84.7%), their Digibus test drive represented their first ride in a self-driving shuttle. 9.2% indicated that they had been on a self-driving shuttle before and 5% indicated that they had ridden on the Digibus in the past. The reasons for taking part in a test drive varied and were equally distributed: 29.4% took part out of professional or scientific interest, 25.5% out of interest in an innovative means of transport, 23.0% were motivated by curiosity, 20.4% were interested in the technology. Around 92% reported enjoying or very much enjoying their ride on the Digibus. According to passenger testimonies, they especially appreciated the vehicle's "comfortable driving style", "feeling of safety", "automated nature of the driving" or its "reliable detection of other road users or obstacles". A little more than 6% of passengers indicated that they had not particularly enjoyed their Digibus ride and 1% said they did not enjoy their ride at all. The reasons for this included, for example "the sharp braking or frequent braking", "the prototypical nature of the technology" or "the lack of smooth driving behaviour“. The passenger survey also revealed very positive results regarding perceptions of safety on board. Almost 90% of passengers felt safe or very safe on board the Digibus. It must be assumed, however, that passengers' feelings of security would probably be lower if the shuttle were completely driverless. As per the reasons passengers gave for not feeling safe on board, they were "abrupt or jerky braking", "lack of confidence in the new technology", "lack of experience", " poor sensor system" or that "the shuttle cannot differentiate between people and vehicles". When asked what passengers could imagine using the Digibus for in their neighborhood, 28.3% answered they would use it to commute to work or school or as a shuttle to the next public transport stop. 20.7% of them indicated they would use a shuttlebus like this for daily errands, such as going shopping, attending doctor's appointments, visits to local government offices etc. Around 16% of passengers could imagine the Digibus being used for either leisure activities (trips to football training or music lessons, as a shuttle to ski-lifts or hiking trail start points etc.), as a delivery service for parcels, shopping etc. or as work-site transportation on enclosed company grounds. 1.4% of the respondents could not imagine an application for a shuttle bus of this kind (Rehrl & Zankl, 2018). Another questionnaire was conducted in the context of the HORIZON 2020 project AVENUE[5], focusing on people with disabilities. Yet, the methodology, questionnaires and replies collected are generic, targeting all types of population. People with some kind of impairment (hard of hearing, reduced mobility or visually impaired) were 57% of the total interviewees. None of the participants with hearing impairment would use an autonomous bus, while 57% of the visual impaired would use it. However, there might not to be a correlation between hearing impaired and acceptance of autonomous vehicles, because all of them are over 60 years of age. In the age group of over 60 only 44% (15 Interviewees) would use an autonomous bus, while 10 would not do it. In the group of users without any impairment only 32 % would not use autonomous vehicles. It has also been noticed that almost 60% of the participants talk with the driver, half of them for conversation, regardless the age. Generally, this survey shows that, although results suggest a high acceptance and expectation from the interviewees, a hesitant attitude towards autonomous vehicles is manifested. This hesitance is due to empathy for drivers who will lose their jobs and, at the same time, due to the fear of the absence of a “supervisor”. Interview partners state that they would use autonomous busses “only if they were in service for a long time without accidents”. So, even the interviewees with the most welcome towards autonomous buses attitude, would prefer a security supervisor in autonomous public transit. Even if passengers’ needs differ according to their disability, all of the user groups expect a significant added value in the use of autonomous buses in public transport. In general, user acceptance increases to the same extent as the autonomous vehicles are optimised and adapted to fit the different user requirements (Dubielzig, Reisch, & Panou, 2018). Also, as part of the CityMobil2 project, a study focused on the users’ expectancies which might influence behavioural intentions regarding the use of Automated Road Transport Systems (ARTS). A total of 349 valid responses were collected. Results show that performance expectancy is the strongest predictor, suggesting that the most important factor that people will consider in deciding whether or not to use an ARTS is how well they believe it will perform in comparison to other public transport systems. Social Influence and Effort Expectancy also had an impact on behavioural intentions, indicating that the influence of other people, and perceptions of how difficult the system is to use will also both influence the decision to use an ARTS. Apart from that, other factors such as perceived safety or on-board comfort should be considered in future work in this area (Madigan, et al., 2016). A similar survey was conducted in order to investigate the factors that influence users’ acceptance of automated road transport systems (ARTS). The audience was the users of the CityMobil2 pilot including a shuttle bus in the city of Trikala. In total, 315 participants completed the questionnaire. All of them had used the shuttled bus at least once, with 14 participants having used it more than 5 times. The main aim of this study was to use an adapted version of the UTAUT framework to investigate the social-psychological factors that influence users’ acceptance of an automated road transport system (ARTS). In particular, users’ enjoyment of the system plays a big part in their desire to use it again, while the performance of the system, the resources provided to support its use and the social popularity of the system all appear to be important factors (Madigan, Louw, Wilbrink, Schieben, & Merat, 2017).

2.2. Gaps, needs, wants and priorities of automation in public transportation

Results occur out of the many pilots that have taken place around Europe, (e.g. Finland, Greece, Austria), in larger and smaller cities. Mini autonomous buses (shuttles) were occupied in city centers, with users welcoming the concept and using the service for their commute. The autonomous vehicles demonstrated required a supervisor on board, since they are not yet qualified for fully autonomous driving. Along with the pilots, many surveys on users’ acceptance and perception of the automated public transport were conducted. Pertaining to those surveys, it appears that the autonomous mini buses were widely accepted and used for small commutes. A significant factor

[5] https://h2020-avenue.eu/

affecting users' acceptance is age, since younger users were identified as more frequent users. Safety was mainly inspired from the presence of the supervisor inside the vehicle. Passengers declare that they would be uncomfortable riding on a bus with no physical supervisor to be in charge of the vehicle and take control when/if needed. In the future plans for automated public transport services, the supervisor of the vehicle will operate the vehicle remotely from a control centre. Although the feedback was positive, as to if users would use again an autonomous bus for their commute, a fear for the driver's absence emerged. Therefore, passengers' psychological status when riding on an autonomous vehicle needs to be estimated so as to detect which circumstances make them feel comfortable and which not. Public transport users' needs and priorities have to do with the safety during their trip, the comfort in the vehicle, the frequency of the scheduled routes and the price of the corresponding services. Overall, automation in public transport will have an impact on travelers' behavior, their social and psychological attitude towards automation, but also on urban development, environment, entertainment and commerce, growth and jobs. The findings from the literature review indicate the most important needs, wants and priorities of end users, along with the existing gaps in automated public transportation. A table collecting these is presented below.

Table 2: Needs, wants and priorities of end users and existing gaps for automation in public transportation

	Gaps	Needs	Wants	Priorities
1	Safety assurance	Frequent service	Accessibility to all (including VECs*), Affordable tickets	Safety-need of a security personnel, Clean interior
2		User-friendly app, Smooth driving style-no harsh brakes or impatient driving	Access remote areas	
3	On demand service: Door to door	Clear announcements on board, Comfortable interior of vehicles	Door to door service	Environmentally-friendly commute

** Vulnerable to Exclusion, e.g. elderly, disabled, tourists, internet/language illiterate, unemployed, unprivileged, refugees*

3. USABILITY OF AUTONOMOUS VEHICLES

3.1. AVs impact in mobility

As a matter of fact, autonomous vehicles will broaden the landscape of potential drivers/operators, since underserved population, such as non-drivers, teens, elderly and people with disabilities, will be able to use AVs for their commuting. Therefore, vehicle automation is subject to increase mobility, a side effect that is not desirable. The development of autonomous vehicles aims to mitigate traffic accidents, enhance traffic safety and reduce emissions. An increase of the number of vehicles will further deteriorate traffic congestion. In order to avoid such an unwanted consequence, autonomous vehicles can be used by more than one family member or among strangers, through a car sharing scheme. Also, autonomous vehicles are equipped with technology infrastructure that allows them to be connected as well. Connected and autonomous vehicles (CAVs) can communicate between them, improving the efficiency of transportation systems (Elliott, Keen, & Miao, 2019). As a result, developing a mobility web instead of promoting the private use of AVs can efficiently reduce the number of cars in the streets (Duarte & Ratti, 2018). A mobility web will also impact parking spaces. Since AVs will be user more frequently as a matter of car sharing, fewer cars and therefore parking spaces will be necessary (Duarte & Ratti, 2018). According to Hannon et al., the boundaries among private, shared, and public transport will be blurred, given that connected and autonomous vehicles can operate upon request, as well as plan trips upon request, while offering travelers flexible ways to commute and keeping a low cost (Hannon, Knupfer, Stern, & Nijssen, 2019). Replacing private car ownership is expected under the utilization of Shared Autonomous Vehicles (SAVs), by offering an efficient, flexible and affordable on-demand mobility (Mohammadzadeh, 2021). Eventually, researchers acknowledge the possible increase in private vehicle's ownership and have already come up with strategies, technologies and plans in order to tackle the threat and actually promote a sustainable transportation solution.

3.2. Advanced driver assistance systems (ADAS)

Autonomous vehicles acquire a lot of advanced systems in order to operate all of the automated functions. Already, vehicles of SAE level 2 and 3 are equipped with plenty of advanced systems, such as the Adaptive Cruise Control (ACC), Lane Keeping Assistance, Blind Spot Warning, Parking Assistance and more. Obviously, vehicles of higher automation are equipped with even more advanced systems so as to offer safe transportation. According to AAA

(American Automobile Association), "at least one ADAS feature is available on 92.7% of new vehicles available in the U.S. as of May 2018" (American Automobile Association, 2019). Some of the most important advanced safety systems integrated in vehicles are presented in Table 3, along with the corresponding technologies required for their operation (Edmonds, 2018).

Table 3: Technologies for advanced safety systems

Technology	ADAS feature
Front Camera Sensors	Automatic Emergency Braking
	Adaptive Cruise Control
	Lane Departure/Keeping
Front Radar Sensors	Automatic Emergency Braking
	Adaptive Cruise Control
Side Mirror Sensors	Around View Monitoring
Rear Radar Sensors	Blind Spot Monitoring
	Rear Cross Traffic Alert
Ultrasonic Sensors	Parking Assistance

Source: (Edmonds, 2018)

All of these systems are developed to provide a safe driving experience, reduce traffic accidents and save lives. However, their invasion in the automotive market did not allow for drivers to get familiar with using these systems, and as a result many buy new cars, not knowing how to operate the new, advanced safety systems. This situation may cause even more significant accidents, if drivers trust that these systems are capable of avoiding any possible threat without their own involvement in the driving procedure. To conclude, specialized training should be offered to past, new and future drivers, focusing on advanced safety systems operation and capabilities.

3.3. Repair cost

While technology is rapidly evolving and safety systems are designed with a view to exclude the driver from actually operating the vehicle, a field commonly neglected is the repair cost of these systems. According to an AAA research from 2018, "vehicles equipped with advanced driver assistance systems (ADAS) such as automatic emergency braking, blind spot monitoring, lane departure warning and others, can cost twice as much to repair following a collision due to expensive sensors and their calibration requirements." (Edmonds, 2018). The cost of repairing or replacing these new features is unfortunately unbearable for most users. Repairing these systems can also occur from minor incidents, related to crashing a door or even a mirror. Furthermore, repairing or replacing of some typical car features, such as the windshield, will also require the replacement of the camera that is behind the glass in an equipped vehicle, multiplying the repair cost to even three times the cost without the advanced system (Edmonds, 2018). AAA estimated repair costs of sensors, taking into considerations the different types and models. The results are presented in Table 4.

Table 4: Estimated repair costs per part

Technology	Repair cost
Front Camera Sensors	$850-$1900
Front Radar Sensors	$900-$1300
Side Mirror Sensors	$500-$1100
Rear Radar Sensors	$850-$2050
Ultrasonic Sensors	$500-$1300

Source: (Edmonds, 2018)

4. CONCLUSION

Focusing on the passenger acceptance of AVs, safety is the driving force in establishing AVs, with the "fear of the unknown" or "the lack of confidence" pointed out as the main reason of concern, inconvenience and doubt. This endogenous mistrust is empowered by the fact that until now the legislative framework is a little obscure pertaining to the share of responsibility in case of accident and who pays except for the insurance companies. In addition, the willingness to pay is also a key issue for everyone, according to who is going to be burdened with the initial investment and how affordable is the new automated passenger transportation system is going to be for the average user with or without public subsidy and / or other support or participation from the private domain or through public private partnership schemes. In fact, the potential costs versus the expected revenues constitutes one of the most important parameters taken into consideration in the frame of spatial planning at strategic and governmental decision making level, but it is equally significant at tactical and operational level as well. Last but not least, there is also the accessibility matter which is twofold: of course there is the issue of accessibility in the phase of design concerning the VRUs, but on the other hand there is the need for easy and equal access from all the users no matter the level of their familiarization with the new technologies, the IoT, the C-ITS and the smart mobility concepts and services, also meeting the expectations for the minimization of the required equipment and respective services. This means that any service or application developed towards the integration or upgrading of the provided services is desired to be equally accessible by all, irrespective of the socioeconomic status of the user, the available equipment - hardware or software (e.g. smart phone or tab running a simple easy to download and operate app) and the level of knowledge, know how, technology illiteracy and familiarization. What is more, another issue is the difficulty from the part of the users towards the sharing and compilation of their personal data. Furthermore, factors that affect users' preference towards automated road transport systems (ARTS) over other public transport systems is each one's performance, as well as the social influence from other people, depending on their recommendations, preferences and indications. A user-friendly service is appealing to a larger audience, including the elderly, people with disabilities and other VRUs. A balance between private autonomous vehicles and AVs for public transportation is expected to occur, taking into consideration the tendency to shared mobility researchers focus on, along with the need for accessible, convenient and low cost transport. The cost accompanying owning and maintaining an autonomous vehicle could be a suspending factor for increased vehicle ownership as well. Sustainable cities need sustainable solutions to combat traffic problems. More lanes will not fix the problem of traffic congestion. Neither will more vehicles.

REFERENCES

[1] Aittoniemi, E., Barnard, Y., Chen, H., Ertl, D., Harrison, G., Kolarova, V., . . . Touliou, K. (2018). D4.7: User requirements analysis. AUTOPILOT EU.

[2] Aittoniemi, E., Barnard, Y., Ertl, D., Harrison, G., Kolarova, V., Malin, F., & Touliou, K. (2019). D4.8: User requirements analysis. AUTOPILOT EU.

[3] Alonso Raposo, M., Ciuffo, B., M., & Thiel, C. (2017). The r-evolution of driving: from Connected Vehicles to Coordinated Automated Road Transport (C-ART) Part I: Framework for a safe &. doi:10.2760/225671: EUR 28575 EN.

[4] Alonso Raposo, M., Ciuffo, B., Alves Dias, P., Ardente, F., Aurambout, J., Baldini, G., Baranzelli, C., Blagoeva, D., Bobba, S., Braun, R., Cassio, L., Chawdhry, P., Christidis, P., Christodoulou, A., Corrado, S., Duboz, A., Duch Brown, N., Felici, S., Fernandez Macias, E., Ferragut Martinez Vara De Rey, J., Fulli, G., Galassi, M., Georgakaki, A., Gkoumas, K., Grosso, M., Gomez Vilchez, J., Hajdu, M., Iglesias Portela, M., Julea, A., Krause, J., Kriston, A., Lavalle, C., Lonza, L., Rocha Pinto Lucas, A., Makridis, M., Marinopoulos, A., Marmier, A., Marques Dos Santos, F., Martens, B., Mattas, K., Mathieux, F., Menzel, G., Minarini, F., Mondello, S., Moretto, P., Mortara, B., Navajas Cawood, E., Paffumi, E., Pasimeni, F., Pavel, C., Pekar, F., Pisoni, E., Raileanu, I., Sala, S., Saveyn, B., Scholz, H., Serra, N., Tamba, M., Thiel, C., Trentadue, G., Tecchio, P., Tsakalidis, A., Uihlein, A., Van Balen, M. and Vandecasteele, I., The future of road transport, EUR 29748 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-14318-5 (online),978-92-76-14319-2 (print), doi:10.2760/668964 (online),10.2760/524662 (print), JRC116644.

[5] American Automobile Association. (2019, January). ADVANCED DRIVER ASSISTANCE TECHNOLOGY NAMES - AAA's recommendation for common naming of advanced safety systems. Retrieved from NewsRoom.AAA.com: https://www.aaa.com/AAA/common/AAR/files/ADAS-Technology-Names-Research-Report.pdf

[6] Cocron, M. P., Vallejo, A., Delgado, M. B., Wilbrink, M., Anund, A., Krupenia, S., & Zanovello, L. (2018). D7.1: Evaluation Framework. ADAS&ME.

[7] CoEXist. (2019, December 4). CoEXist Automation-Ready survey – preliminary results available. Retrieved from CoEXist: https://www.h2020-coexist.eu/coexist-automation-ready-survey-first-results-available/

[8] Duarte, F., & Ratti, C. (2018). The Impact of Autonomous Vehicles on Cities: A Review. Journal of Urban Technology, 3-18. doi:10.1080/10630732.2018.1493883

[9] Dubielzig, M., Reisch, M., & Panou, M. (2018). D2.4: First Passenger needs analysis and specifications. AVENUE project.

[10] Edmonds, E. (2018, October 25). New Vehicle Technologies Double Repair Bills for Minor Collisions. Retrieved from NewsRoom.AAA.com: https://newsroom.aaa.com/2018/10/new-vehicle-technologies-double-repair-bills-minor-collisions/

[11] Elliott, D., Keen, W., & Miao, L. (2019). Recent advances in connected and automated vehicles. Journal of Traffic and Transportation Engineering (English Edition), 109-131. doi:https://doi.org/10.1016/j.jtte.2018.09.005

[12] Hannon, E., Knupfer, S., Stern, S., & Nijssen, J. T. (2019, January 16). The road to seamless urban mobility. Retrieved from McKinsey&Company: https://www.mckinsey.com/business-functions/sustainability/our-insights/the-road-to-seamless-urban-mobility

[13] Joint Research Centre. (2020). Special Eurobarometer 496 - Expectations and concerns of connected and automated driving Europe. ISBN: 978-92-76-17323-6.

[14] Kulmala, R., Jääskeläinen, J., & Pakarinen, S. (2019, June). The impact of automated transport on the role, operations and costs of road operators and authorities in Finland. ISSN (online) 2342-0294, ISBN (online) 978-952-311-306-0: Finnish Transport and Communications Agency Traficom. Retrieved from TRAFICOM: https://www.traficom.fi/sites/default/files/media/publication/EU_EIP_Impact_of_Automated_Transport_Finland_Traficom_6_2019.pdf

[15] Madigan, R., Louw, T., Dziennus, M., Graindorge, T., Ortega, E., Graindorge, M., & Merat, N. (2016). Acceptance of Automated Road Transport Systems (ARTS): An Adaptation of the UTAUT Model. Transportation Research Procedia Vol. 14, 2217–2226. doi:https://doi.org/10.1016/j.trpro.2016.05.237

[16] Madigan, R., Louw, T., Wilbrink, M., Schieben, A., & Merat, N. (2017). What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. Transportation Research Part F: Traffic Psychology and Behaviour, 50, 55–64. doi:https://doi.org/10.1016/j.trf.2017.07.007

[17] Mohammadzadeh, M. (2021). Sharing or owning autonomous vehicles? Comprehending the role of ideology in the adoption of autonomous vehicles in the society of automobility. Transportation Research Interdisciplinary Perspectives, Volume 9. doi:https://doi.org/10.1016/j.trip.2020.100294

[18] Portouli, E., Karaseitanidis, G., Lytrivis, P., Amditis, A., Raptis, O., & Karaberi, C. (2017). Public attitudes towards autonomous mini buses operating in real conditions in a Hellenic city. 2017 IEEE Intelligent Vehicles Symposium (IV) (pp. 571-576). Los Angeles, CA: doi: 10.1109/IVS.2017.7995779.

[19] Rehrl, K., & Zankl, C. (2018). Digibus©: results from the first self-driving shuttle trial on a public road in Austria. Eur. Transp. Res. Rev. 10, 51, https://doi.org/10.1186/s12544-018-0326-4.

[20] UK Autodrive. (2017, August 8). Survey finds UK public still “open minded” about self-driving vehicles. Retrieved from UK Autodrive: http://www.ukautodrive.com/survey-finds-uk-public-still-open-minded-about-self-driving-vehicles/

[21] Winkler, M., Mehl, R., Erander, H., Sule, S., Buvat, J., KVJ, S., . . . Khemka, Y. (2019). The autonomous car A consumer perspective. Capgemini Research Institute.

[22] Zoellick, J. C., Kuhlmey, A., Schenk, L., Schindel, D., & & Blüher, S. (2019). Assessing acceptance of electric automated vehicles after exposure in a realistic traffic environment. https://doi.org/10.1371/journal.pone.0215969: PLoS ONE, 14(5).

REDUCTION OF NOISE ON ROAD SURFACES (WEARING COURSES) IN THE CZECH REPUBLIC

Blanka Hablovičová, M.Sc.
Vítězslav Křivánek, Ph.D.
Petra Marková, M.Sc.
Transport Research Centre (CDV)
Líšeňská 33a, 636 00 Brno, Czech Republic
blanka.hablovicova@cdv.cz

ABSTRACT

Changes in acoustic characteristics of the road's wearing courses during years of using have an immediate influence on the total noise level from the road traffic. According to the CNOSSOS-EU for the calculation of strategic noise maps, protection of the environment against traffic noise is very important. Long-term noise measurements by the CPX method are performed by only one company in the Czech Republic (CZE): Transport Research Centre (CDV). Low-noise wearing courses began to be used a few years ago (from mixture PA, through BBTM, to SMA NH (NH is the Czech abbreviation and marking for low-noise)). CDV cooperated on the elaboration of the technical specification TP 259 defining the rules for designing asphalt mixtures for noise reducing wearing courses and issued the certificated method for assessing their long-term noise performance, both of them were issued in 2017. Noise reduction increased slowly until documents were announced but rapidly increased after this because everyone had to start following the rules that had not been there before. Due to constantly changing low-noise mixtures, it is necessary to observe and evaluate acoustic characteristics of road surfaces during the lifetime of wearing courses and the whole roadway. Tyre/road noise very much depends on the asphalt mixture and condition of the wearing course. Low-noise and common asphalt mixtures are laid in some locations in the CZE right next to each other so that their behavior over time can be studied. Practical results measured in the CZE are described in this study.

Keywords:
Low-noise wearing course, asphalt mixture, road noise reduction, CPX.

1. INTRODUCTION

Noise is one of the most important (and still rising) pollutant affecting human health (World Health Organization, 2018; Klæboe, 2011) with impact such as sleep disturbance (Pirrera et al., 2014; Ohrstrom et al., 2006), cardiovascular and physiological effects (Babisch, 2008; Bluhm et al., 2007), mental health effects (Klompmaker et al., 2019; Stansfeld et al., 2009), annoyance (Okokon et al., 2018) or cognitive impairment (Jafari et al., 2018; Jafari et al., 2020). Traffic noise has also impact on wildlife (Francis et al., 2009; Castaneda et al., 2020; Finch et al., 2020), on economy and social life (Fyhri et al., 2006; Kim et al., 2019) (detail study about noise costs is contained in the chapter 6 of handbook (Van Essen et al., 2019)).

The growing trend of noise pollution has led to strategic and operational legal measures through EU, namely with the adoption of the Environmental Noise Directive (Directive 2002/49/EC, 2002). The directive implies requirements for Strategic Noise Mapping (SNM) with regular updates in five-year cycles in all EU countries. The methodology CNOSSOS-EU (Kephalopoulis et al., 2012) for calculating SNM is mandatory for EU members since 31 December 2018 (in legislation recorded as Commission Directive 2015/996 (2015)). The Environmental Noise Directive also requires to implement action plans in which noise reduction solutions are needed. The most used solutions for traffic noise reduction are noise barriers which can achieve high noise reduction rate. Their construction and maintenance costs are unfortunately high and they cannot be implemented into the urban areas in many cases as applicability in a street canyon is problematic, e.g. possible non-compliance with the conditions for daylighting of residential units, noise barrier would not be effective due to interruptions by side streets etc. Replacement of the road surface for low-noise wearing course is one of the approaches to reduce traffic noise, especially in urban areas, with lower cost than noise barrier or façade insulation (Larsen et al., 2001; Krivanek et al., 2020).

The aim of this paper is to present the equipment for monitoring of wearing courses in the Czech Republic, to introduce the most important documents about low-noise wearing courses at the national level, and to show some partial results from tire/road contact noise measuring in the Czech Republic.

2. LOW-NOISE WEARING COURSES AND NOISE MONITORING

The properties of the road wearing course are highly influencing the total road traffic noise emission. The importance of surface and condition of road will continue to increase as a consequence of the development of electric vehicles and electromobility. Noise generated by tire/road contact, whose mechanism is described in (Sandberg & Ejsmont, 2002), is prevailing from speed approx. 40 km/h for current passenger cars with combustion engines and for heavy vehicles it is around 70 km/h (Sandberg & Ejsmont, 2002), whereas for electric vehicles, tire/road contact noise dominates from speeds as low as 20 km/h due to very quiet engines (Czuka et al., 2016). Basic structural properties of low-noise pavements for appropriate tire/road noise reduction are: high voids content, minimal megatexture, and not a very smooth macrotexture (Sandberg, 1999).

Sandberg (1999) described the history in short about the low-noise pavements before year 2000. Descornet (2000) introduced low-noise techniques for many materials such as cement concrete, resin-bound surface, porous asphalt, thin layers and some specific structures. Aksnes et al. (2009) presented results from development and testing of low-noise pavements adapted to Norwegian conditions, Anfosso-Lédée and Brosseaud (2009) monitoried porous asphalt concrete and very thin asphalt concrete in France. Vaitkus et al. (2014) performed laboratory testing (acoustic properties included) of stone mastix asphalt and porous asphalt used in Lithuania according to Lithuanian standards. Acoustic ageing of rubberized pavements was modelled by Licitra et al. (2019). PIARC published a summary (PIARC, 2019) of global knowledge of traffic noise at the end of 2019 – noise limits, impact of noise on human health, environmental aspects, economic evaluations, prediction, measurements, evaluation, anti-noise measures and maintenance of road surfaces.

Implementation of reliable and comparable measurements of wearing courses over a long period is a necessary condition for the correct evaluation of the influence of road surfaces on traffic noise. Recommended method for long-term monitoring acoustic behavior of wearing courses (to compare the results from different locations) is the close-proximity (CPX) method in accordance with the results of the European project ROSANNE (Haider et al., 2016) since 2015. CPX measurements are carried out by various types of measuring systems across the world, such as enclosed CPX trailers (Kleiziene et al., 2019; Miljkovic et al., 2012; Vazquez et al., 2019), open CPX trailers (Tonin et al., 2014; Van Leewen et al., 2007; Kragh et al., 2013). Test tire with microphones can be also assembled directly on a test vehicle (Cesbron & Klein, 2017; Licitra et al., 2014; Campillo-Davo et al., 2019). Round robin test between nine CPX systems from four countries performed by Peeters et al. (2018) concluded that all systems met requirements, which are differences from average reference values, for random errors (range 0.11–0.27 dB) and repeatability (range 0.07–0.22 dB), two didn't fulfil the requirements for systematic error, differences were above 1 dB. Vieira and Sandberg (2019) presented similar conclusions from round robin test in Sweden (two open trailers and two enclosed trailers were tested) with some recommendations for the future of CPX method. So the results obtained with different CPX trailers which meets ISO 11819-2 (2017) are comparable.

3. SITUATION IN THE CZECH REPUBLIC

Long-term road noise measurements are performed by Transport Research Centre (CDV) in the Czech Republic (CZE). SPB method was used for road noise measuring until 2011 (Cholava et al., 2009). CPX method was tested and compared with SPB during the project CG712-102-120 (Transport Research Centre, 2011). Important output of the project was the approval of the CPX trailer for road noise measurements in the Czech Republic. The comparative noise measurement were proceeded between CPX systems of CDV and Eurovia (Krivanek et al., 2014) in the years 2013–2018 within project CESTI (2018). So, the regular monitoring of road noise surfaces by CPX method has been carried out by CDV since 2012 (Krivanek, Markova & Spicka., 2016) and measurements along distances over 5,500 km were performed until 2020. Methodology from 2014 (Transport Research Centre, 2014) is an effective CZE tool for independent verification, evaluation and assessment of noise properties of roads since 2014. Based on the obtained data since 2012, two important documents were published in 2017: the first was certified methodology (Transport Research Centre, 2017) which presents values of acoustic emission of road surfaces and their change over time, and the second was Technical Specification No. 259 (TP 259) (Valentin et al., 2017) which e.g. defines the requirements for materials of low-noise pavements, specifies the method of marking, determines the technical requirements as void content, and construction conditions as compaction. Both documents are fully integrated into the quality system in the field of roads.

According to the National Reference Laboratory, CNOSSOS-EU is intended as a decisive method of traffic noise determination for the Czech Republic in the future. It defines road surface noise corrections for entire EU globally but these pavements don't correspond to wearing course mixtures in individual EU member states, not only by their parameters but also by their composition or noise levels (each country uses mixtures adapted to local conditions). So implementation of CNOSSOS-EU in national Czech conditions is ongoing.

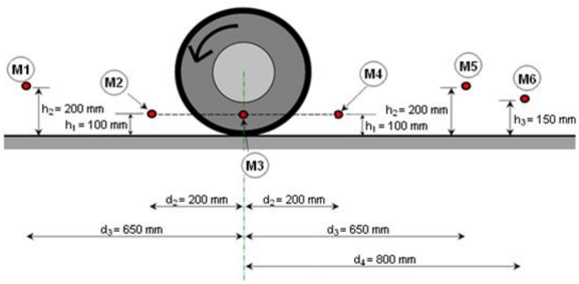
Average noise level of the typical road section (ACO 11) in the CZE at the second year of lifetime was determined as reference value for tire/road contact noise of wearing courses in the CZE. Reference value is set at 90 dB for velocity 50 km/h and 98 dB for velocity 80 km/h (for tyre P1 according to ISO 11819-3 (2017)) based on measurements performed in the Czech Republic and foreign experiences. Pavements with lower values have noise reducing properties. The low-noise surface is a surface that can reduce tire rolling noise by at least 3 dB compared to an average noise level of typical road surface. Newly laid low-noise wearing courses can reduce noise by up to 5 dB. They are divided into low-noise wearing courses which meet TP 259 (with abbreviation NH) and into wearing courses outside of TP 259 but still reducing noise due to small aggregate fraction since 2017 when TP 259 was issued. The first experimental implementations of low-noise wearing courses in CZE were realized in 2010, approximately twenty of these test sections were layed until 2012 (porous asphalt (PA) primarily). PA has excellent properties immediately after laying but clogging and loss of surface macrotexture which decrease noise reduction properties, occur very quickly, especially in urban areas, PA wearing course is also damaged by freeze-thaw cycles (Kwiatkowski et al., 2014). Wearing courses SMA LA (from originally Austrian or German "lärmarm") or BBTM have been used since 2014 based on foreign experiences (especially from Germany) which were incorporated to TP 259 so they are named SMA NH and BBTM NH (NH is Czech abbreviation for low-noise and has the same meaning as LA) since 2017. Grinding technology based on diamond grinding of concrete surface is being tested but for low-noise wearing courses, asphalt mixture are prioritized.

4. MEASURING PROCEDURE

4.1. CPX method

Measurements have been performed on an open CPX trailer designed by CDV with approximately width 2.5 m and length 5 m (see Fig. 1) which meets requirements of the international standard ISO 11819-2 (2017) describing close-proximity method. Standard is supplemented by ISO 11819-3 (2017) that defines reference tires for use in the CPX method (the standard was revised in January 2021). Locations of individual measuring microphones on a CPX trailer according to ISO 11819-2 are shown in Fig. 2 and the real placement can be seen in Fig. 3. The standard ISO 11819-2 specifies two mandatory microphones for CPX measurements, but more microphones can be used where more complete processing of tire/road noise is required as for low-noise surfaces. So, the system was upgraded from 5 microphones to 6 microphones (supplemented, tested and verified) according to the standard during winter 2018/2019.

Figure 1: Open CPX trailer designed by Transport Research Centre.



Source: ISO 11819-2 (2017).

Figure 2: Diagram of the locations of individual measuring microphones on a CPX trailer.

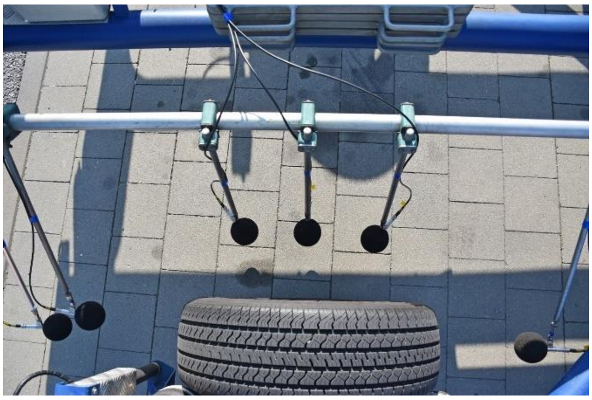


Figure 3: Real placement of individual measuring microphones on the CPX trailer.

Measurements were performed at a speed of 50 km/h for rural areas, and at 80 km/h for other categories of roads (eg. highways, bypass). Velocity of vehicle with CPX trailer (GPS module UA-9004 RLVBSS1), air temperature (sensor SENECA PT100 probe) and surface temperature of measured road (infrared sensor CALEX – PC21MT-0) were continuously recorded during each measurement, see Fig. 4. The measuring multianalyzer Brüel&Kjær was composed of PULSE LAN-XI (module 3050-B-060 and module 3056-A-040). The software PULSE LabShop which can work and synchronize all signals in one timeline, even if they have different sampling frequencies, was used. Pre-polarized free field microphones Brüel&Kjær type 4189, in accuracy class 1 and with sensitivity 50 mv/Pa, were used for self-sensing noise.

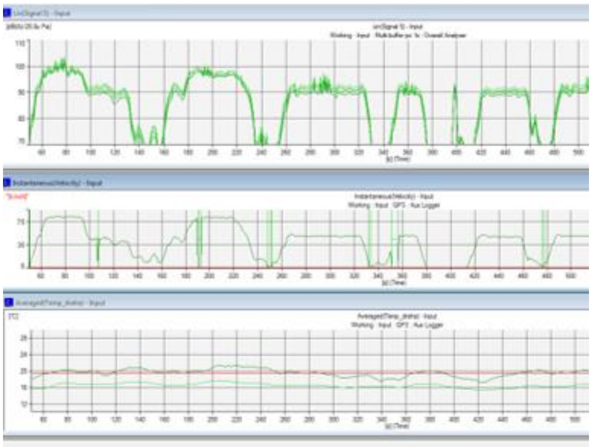


Figure 4: L_{Aeq}, velocity and temperature on a selected section on synchronized timeline.

The effect of the measuring tire represents the greatest uncertainty in the noise measurement at the tire/road contact (Licitra et al., 2017; Buhlmann et al., 2007), which means the same type of tire should be used for long term measurements. Rubber hardness and temperature of the tire have to be within a specified range (Sandberg et al., 2016) and the standard ISO/TS 13471-1 (2017) determines procedures for defining effect of temperature on tire/road noise emissions. Only one tire should be selected for measurements. Commonly manufactured tires don't have adequate production stability and the results must be comparable during time. Tire Uniroyal Tigerpaw 225/60 R16 SRTT (in ISO 11819-3 marked as P1) is recommended for passenger cars in the automotive industry as a standard tire for reference tests according to ASTM F2493 (ASTM, 2020), so the tire SRTT P1 was used for measurements, see Fig. 5. Every year is used newly acquired reference tire in order to minimize the effect of degradation (measurements are performed in the period from about April to October). Impurities in the tread were removed before the actual testing. Hardness has been tested every two months.

Figure 5: Used reference tire Tire Uniroyal Tigerpaw 225/60 R16 SRTT.

4.2. Measured data

The main output of the measurements is the equivalent sound pressure level of tire/road contact and third-octave characteristic of section according to ISO 11819-2. All measurements have been adjusted by A-weighting filter. Data obtained by synchronic measurements from the 6 microphones postprocessed within the PULSE LabShop program. CPX measurement has prescribed spectrum range that is recorded but only the required part of spectrum is evaluated. Disturbing influences, as crossing over the sewerage cover or high peaks caused for example by the rescue siren, were removed manually. Arithmetic average of noise levels values at the 6 microphones was performed by Excel for all valid measurements, it is given as the final result rounded to one decimal place.

5. RESULTS AND DISCUSSION

Results from two different measurement sites are selected as partial examples. Comparison of common and low-noise asphalt mixture (wearing course) from the same locality is made. The sections follow immediately after each other, i.e. sections are affected by almost the same meteorological conditions, traffic intensity and composition as well as the velocity of the traffic flow. Periodic measurements of noise verification were madein summer annually, the wind speed was below 5 m/s and obtained noise values were recalculated according to the standard thus all results were referenced under comparable conditions; e.g. noise value measured under condition of velovity 78.9 km/h and temperature 23.8 °C was corrected to noise value under conditions 80 km/h and 20 °C. Year 0 of wearing course in legend of results (Fig. 6 and Fig. 9) represents the year of laying this pavement, eg. if year 2015 is year 0, then 2016 is year 1, 2017 year 2 etc.

The acoustic lifetime of the low-noise wearing course is characterized as the period during which monitored type of wearing course achieves an improvement in acoustic attenuation compared to reference value. When the wearing course reaches its acoustic life limit, this doesn't mean that the life limit of such wearing course is reached. Acoustic life of wearing course can be very different and can be fundamentally affected by appropriate maintenance.

5.1. Rural road

The first example represents a rural area with a velocity limit of 50 km/h. Industrial (heavy freight) and agricultural activities are very active in this area. It is a class II road with traffic flow above 3,000 vehicles per day. Differences between noise development of low-noise (PA 8 CRmB, section length 780 m) and common (ACO 11, section length 260 m) asphalt mixture is shown in Fig. 6. The most significant increase of noise (no road's defects) occurs in the first three years after laying which is related to the gradual "taxiing" of the surface.

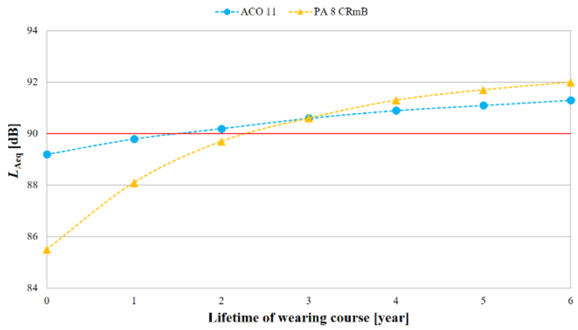


Figure 6: Comparison of tyre/road noise of mixture ACO 11 and low-noise mixture PA 8 CRmB in rural area under heavy transport and agricultural machinery during six years of lifetime (measured with velocity 50 km/h).

Red line in Fig. 6 symbolizes the reference equivalent sound pressure level 90 dB for velocity 50 km/h. Common (ACO 11, blue color) and low-noise (PA 8 CRmB, yellow color) surfaces have approximately the same noise level about 2–3 years after laying without regular maintenance in this measured location. Noise of ACO 11 was 89.2 dB after laying and after two years it was 90.2 dB, it is an increase of only one decibel. Compared to that, noise increase of PA 8 CRmB during two years of lifetime is 4.2 dB, from 85.5 dB to 89.7 dB. Low-noise asphalt mixtures are mostly surfaces with high void content, especially porous asphalt has higher void content then other low-noise asphalt mixtures. In order to achieve the maximum void content of asphalt, the individual grains are bonded in points and there are high requirements for the binder, e.g. apply suitable additives such as mineral or cellulose fibers (Valentin et al., 2017). Nevertheless, there is a risk of gradual weakening of the contact surface which leads to a higher susceptibility to damage of the pavement – ripping out of aggregates from wearing course (Sandberg & Ejsmont, 2002). In addition, clogging occurs at these surfaces (Fig. 7). The self-cleaning effect by the traffic is minimal at lower speeds, so a specific treatment is required (Krivanek, Pavkova et al., 2016). Regular maintenance of all low-noise wearing courses is rather not performed in the CZE. Less significant noise increase of the low-noise surface may be due to their a shorter service life compared to common surfaces. Lifetime of low-noise surfaces is set at 8 years and lifetime of common wearing courses is 12 years according to the Czech experiences and informations from Directorate of Roads and Motorways. After three years of lifetime was noise of wearing courses ACO 11 and PA 8 CRmB 90.6 dB for both of them and then noise of porous asphalt increased faster than noise of common asphalt ACO 11, so in six year of wearing course lifetime was noise of PA 8 CRmB 92 dB, resp. 91.3 dB for ACO 11. So, defects of low-noise wearing course show earlier than on the common surface, especially in the experimental and the first realized sections of these asphalt mixtures.

Figure 7: Clogging of road surface.

Third-octave frequency spectrum of ACO 11 and PA 8 better expresses noise emission changes (Fig. 8). More detailed differences between the surfaces can be traced by this. PA 8 CRmB after laying (red color) has the lowest values and they are significantly lower than others for frequencies above 800 Hz. In porous asphalt, sound wave propagates into the gap, releases part of its acoustic energy (with the generation of heat) and thus reduces the noise reflected back out. Frequencies of ACO 11 (orange color) and PA 8 CRmB (green color) after two years of lifetime have very similar course which is in agreement with the measured noise values in the Fig. 6. The situation is opposite to the previous case of PA, the gaps are clogged, sound waves don't get into them and are immediately reflected.

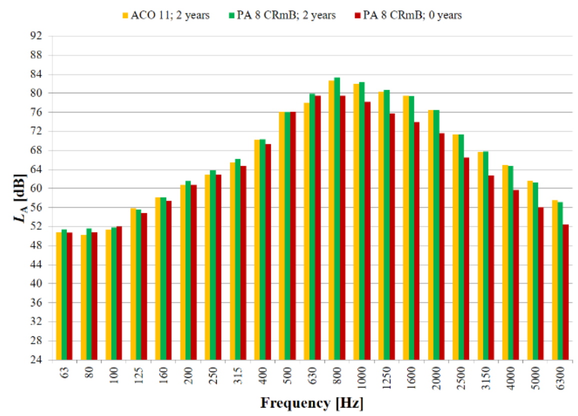


Figure 8: Third-octave frequency spectrum of A-weighted signal of surfaces ACO 11 (2 years) and PA 8 CRmB (0 and 2 years) for velocity 50 km/h.

5.2. Highway

Fig. 9 shows results for highway road outside of rural area with traffic flow above 30,000 vehicles per day and higher velocity 80 km/h which means that the self-cleaning effect applies here. The absence of agricultural machinery on the road represents lower mud pollution and thus less clogged surface. Red line in Fig. 9 symbolizes the reference equivalent sound pressure level 98 dB for velocity 80 km/h. There are results of low-noise with section length 2,500 m (BBTM 8 NH, yellow color) and 4,100 m long section of common (ACO 11, blue color) wearing course.

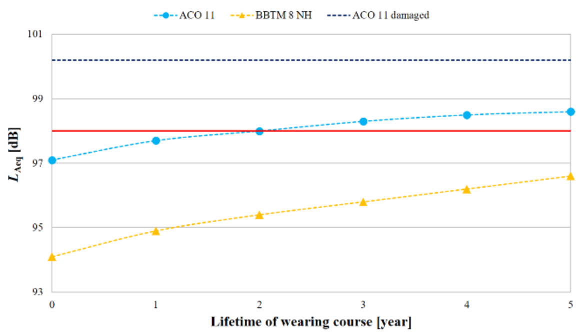


Figure 9: Comparison of tyre/road noise of low-noise mixture BBTM 8 NH and common mixture ACO 11 during five years of lifetime (measured with velocity 80 km/h).

Dark blue dashed line (100.2 dB) represents noise level of the damaged common pavement before replacement, the value 97.1 dB was measured after laying the new pavement (the same asphalt mixture) at the same place so it is a decrease of 3.1 dB. The renewing of wearing course even with a common asphalt is also a good solution for road noise reduction. This exchanged surface reaches reference value 98 dB after two years and then the noise is slowly increasing to 98.6 dB in five year of lifetime of pavement ACO 11. Low-noise surface has the biggest influence on noise reduction in this location. Noise level of noise is 94.1 dB after laying which is 3 dB lower than common pavements after laying. Noise continuously increases to value 96.6 dB in five year of lifetime, so even after five years the noise level doesn't reach the reference level.

The most significant increase in wearing course noise occurs in the first few (2–4) years after laying. Subsequently, the increase of noise gradually slows down provided that no distresses occur. Behavioral curve about the long-term development of road surface noise corresponds to logarithmic rather than a linear course, which is also generally known fact (Haider et al., 2016). Initial void content of asphalt mixtures should be in the range 8–15 vol. % according to TP 259. This value decreases by gradual presence of dust particles, grease and other impurities until the asphalt course approaches void content of a common asphalt mixture with value below 4 vol. %. Technically, this asphalt course continues to function for several years as a quality wearing course, but from an acoustic point of view, its acoustic life will be gradually achieved. So, the results from noise measurements performed in the Czech Republic correspond with the knowledge from foreign countries.

Open wearing course of porous asphalt type is less suitable to use in the Czech Republic due to meteorological conditions (freeze-thaw cycles) primarily and in terms of non-mainenance (clogging). Their initial benefits may be large, but air gaps become clogged faster and if more frequent regular cleaning is not chosen, they quickly lose their acoustic properties. Asphalt mixture BBTM 8 NH (and also SMA 8 NH) is more suitable for using and higher velocity of vehicles is also benefit due to self-cleaning effect. Tyre/road noise changes over time, so it is not possible to stay at the first measured value as given value for wearing course, but it is necessary to perform repeated measurements.

In the area of evaluation of road surface properties at the international level within CEDR the need to introduce and monitor noise in the new road management system is discussed. At the national level, a decision scheme for the classification of the overall condition of road is proposed, listing the five main criteria for monitoring condition of the road network: anti-skid properties, roughness, surface noise, road load capacity and disorders (damages).

The noise should be monitored but no specific indicators are prescribed. Determination of noise classification of road surfaces is the main aim of the solving project in the national level in the Czech Republic for now.

6. CONCLUSION

The CNOSSOS-EU methodology requires for processing of strategic noise maps to work with individual types of road surfaces (e.g. SMA, ACO, BBTM, PA etc.), including its age. Each country has its own recipe for each type of asphalt mixture. The accompanying table F-4 shows the road surface coefficients which are based on data obtained in the Netherlands. Many types of these surfaces are not used in the Czech Republic, the composition of the wearing courses used in CZE is similar to those used in Germany, yet there are differences. However, CNOSSOS-EU allows the setting of national coefficients. The exact determination of values (especially for low-noise wearing course) is up to the individual EU members due to the high diversity of used technologies and local meteorological conditions. Roads are in the CZE classified into five categories (1–5) on the basis of skid resistance, pavement surface roughness, a load of road surfaces and damages. Road surface noise has no classification and there is already a demand for noise assessment for e.g. manufacturers or road managers. Tire/road noise is a variable parameter and it is important to perform many CPX measurements and data evaluations of different wearing courses to create classification levels for the CZE. Partial results could then be used as input data for modeling within the CNOSSOS methodology, so as for adjustment of the Czech tire/road noise coefficients.

The noise changes over time and differently for each wearing course. These changes of noise in tire/road contact can be monitored by regular measurement with the CPX method. Low-noise wearing courses have a durable efficiency on highways due to the self-cleaning effect thru the velocity of vehicles. Whereas on rural roads, pavements may lose rapidly their acoustic properties and behave as a common wearing course faster due to clogging which can be better observed using the noise spectrum. So, if we want to keep acoustical function of low-noise surfaces longer, it is necessary to perform proper maintenance/cleaning, which is not yet the rule in the Czech Republic.

Acknowledgments

This article is financed from the state budget by the Technology agency of the Czech Republic and the Ministry of Transport of Czech Republic under the DOPRAVA 2020+ Programme, project CK02000121 Determination of values of classification levels for evaluation of road surface noise in the Czech Republic.

REFERENCES

[1] Aksnes, J., Saba, R. G., & Berge, T. (2009). Development and testing of low noise pavements in Norway. International Conference BCR^{2}A’09 (pp. 513–520), Champaign, USA. Leiden: CRC Press (Taylor & Francis Group).

[2] Anfosso-Ledee, F., & Brosseaud, Y. (2009). Acoustic monitoring of low noise road pavements. Noise Control Eng., 57(2), 50–62.

[3] ASTM International. F2493-20: Standard Specification for P225/60R16 97S Radial Standard Reference Test Tire. West Conshohocken, PA, USA, 2006(2020).

[4] Babisch, W. (2008). Road traffic noise and cardiovascular risk. Noise Health, 10, 27–33.

[5] Bluhm, G. L., Berglind, N., Nordling, E., & Rosenlund, M. (2007). Road traffic noise and hypertension. Occup. Environ. Med., 64, 122–126.

[6] Buhlmann, E., Schulze, S., & Ziegler, T. (2013). Ageing of the new CPX reference tyres during a measurement season. Internoise (pp. 369–376), Innsbruck, Austria. Austrian Noise Abatement Association.

[7] Campillo-Davo, N., Peral-Orts, R., Campello-Vicente, H., & Velasco-Sanchez, E. (2019). An alternative close-proximity test to evaluate sound power level emitted by a rolling tyre. Appl. Acoust., 143, 7–18.

[8] Castaneda, E., Leavings, V. R., Noss, R. F., & Grace, M. K. (2020). The effects of traffic noise on tadpole behavior and development. Urban Ecosyst., 23, 245–253.

[9] Cesbron, J., & Klein, P. (2017). Correlation between tyre/road noise levels measured by the Coast-By and the Close-ProXimity methods. Appl. Acoust., 126, 36–46.

[10] CESTI, Centre for Effective and Sustainable Transport Infrastructure, project No. TE01020168 of Technology Agency of the Czech Republic. Available from: http://www.cesti.cz/index.php?page=home&en=en.

[11] Cholava, R., & Krivanek, V. (2009). Optimization of parameters of road surfaces to reduce noise burden from transport. International Congress on Sound and Vibration (pp. 1408–1415), Krakow, Poland. Auburn: International Institute of Acoustics and Vibration.

[12] Commission Directive (EU) 2015/996 of 19 May 2015 establishing common noise assessment methods according to Directive 2002/49/EC of the European Parliament and the Council. Off. J. Eur. Comm., L 168, 1–823.

[13] Czuka, M., Palla, M. A., Morgan, P., & Conter, M. (2016). Impact of potential and dedicated tyres of electric vehicles on the tyre-road noise and connection to the EU noise label. Transp. Res. Proc., 14, 2678–2687.

[14] Descornet, G. (2000). Low-noise road surface techniques and materials. Paper presented at Internoise 2000, Nice, France.

[15] Directive 2002/49/EC of the Europen Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise. Off. J. Eur. Comm., L 189, 12–25.

[16] Finch, D., Schofield, H., & Mathews, F. (2020). Traffic noise playback reduces the activity and feeding behaviour of free-living bats. Environ. Pollut., 263(B), 114405.

[17] Francis, C. D., Ortega, C. P., & Cruz, A. (2009). Noise pollution changes Avian communities and species interactions. Curr. Biol., 19(16), 1415–1419.

[18] Fyhri, A., & Klæboe, R. (2006). Direct, indirect influences of income on road traffic noise annoyance J. Environ. Psychol., 26(1), 27–37.

[19] Haider, M., Conter, M., Green, M., Schmidt, B., Sandberg, U. (2016). Status of the EU-Project ROSANNE. Transp. Res. Proc., 14, 2946–2955.

[20] ISO 11819-2:2017. Acoustics – Measurement of the in-fluence of road surfaces on traffic noise – Part 2: The close-proximity method. Geneva: International Organization for Standardization, 2017.

[21] ISO 11819-3:2017. Acoustics – Method for measuring the influence of road surfaces on traffic noise – Part 3: Reference tyres. Geneva: International Organization for Standardization, 2017.

[22] ISO/TS 13471-1:2017. Acoustics – Temperature influence in tyre/road noise measurements – Part 1: Correction for temperature when testing with the CPX method. Geneva: International Organization for Standardization, 2017.

[23] Jafari, Z., Kolb, B. E., & Mohajerani, M. H. (2018). Chronic traffic noise stredd accelerates brain impaitment and cognitive decline in mice. Exp. Neurol., 308, 1–12.

[24] Jafari, Z., Kolb, B. E., & Mohajerani, M. H. (2020). Noise exposure accelerates the risk of cognitive impairment and Alzheimer's disease: Adulthood, gestational, and prenatal mechanistic evidence from animal studie. Neurosci. Biobehav. Rev., 117, 110–128.

[25] Kephalopoulis, S., Paviotti, M., & Anfosso-Lédée, F. (2012). Common noise assessment methods in Europe (CNOSSOS-EU). Luxembourg: Publications Office of the European Union.

[26] Kim, K., Shin, J., Oh, M., Jung. J-K. (2019). Economic value of traffic noise reduction depending on residents' annoyance level. Environ. Sci. Pollut. Res., 26, 7243–7255.

[27] Klæboe, R. (2011). Noise and health: annoyance and interference. In J. O. Nriagu (Ed.), Encyclopedia of Environmental Health (Vol. I, pp. 152–163). Amsterdam, London: Elsevier Science.

[28] Kleiziene, R., Sernas, O., Vaitkus, A., & Simanaviciene, R. (2019). Asphalt pavement acoustic performance model. Sustainability, 11(10), 2938–2952.

[29] Klompmaker, J., Hoek, G., Bloemsma, L. D., Wijga, A. H., Van den Brink, C., Brunekreef, B., Lebret, E., Gehring, U., & Janssen, N. A. H. (2019). Associations of combined exposures to surrounding green, air pollution and traffic noise on mental health. Environ. Int., 129, 525–537.

[30] Kragh, J., Iversen, L. M., & Sandberg, U. (2013). Road surface texture for low noise and low rolling resistance. Copenhagen: Vejdirektoratet.

[31] Krivanek, V., & Effenberger, K. (2014). Comparison of noise level measurement of road surfaces in the Czech Republic using the Close Proximity method. Transactions on Transport Sciences, 7(3), 109–116.

[32] Krivanek, V., Markova, P., & Spicka, L. (2016). Verification of the acoustic parameters of Close-Proximity method. Akustika, 26, 13–19.

[33] Krivanek, V., Pavkova, A., Togel, M., Jedlicka, J. (2016). Cleaning low-noise surfaces as a basic condition for improving pavement's acoustic absorption capability. Arab. J. Sci. Eng., 41, 425–431.

[34] Krivanek, V., Markova, P., & Hablovicova, B. (2020). An economic comparison of the ordinary and low-noise wearing course during the lifetime of roadway. Akustika, 36, 2–15.

[35] Kwiatkowski, K., Oslakovic, I. S., Hartmann, A., ter Maat, H. (2014). Potential impact of climate change on porous asphalt with a focus on winter damage. Paper presented at Transport Research Arena 2014, Paris, France.

[36] Larsen, L. E., & Bendtsen, H. (2001). Costs and perceived noise reduction of porous asphalt pavements. Internoise (pp. 1859–1864), Hague, Netherlands. Red Hook, NY: Curran Associates, Inc.

[37] Licitra, G., Teti, L., & Cerchiai, M. (2014). A modified Close Proximity method to evaluate the time trends of road pavements acoustical performances. Appl. Acoust., 76, 169–179.

[38] Licitra, G., Teti, L., Cerchiai, M., & Bianco, F. (2017). The influence of tyres on the use of the CPX method for evaluating the effectiveness of a noise mitigation action based on low-noise road surfaces. Transp. Res. D Transp. Environ., 55, 217–226.

[39] Licitra, G., Moro, A., Teti, L., Del Pizzo, A., Bianco, F. (2019). Modelling of acoustic ageing of rubberized pavements. Appl. Acoust. 146, 237–245.

[40] Miljkovic, M., & Radenberg, M. (2012). Thin noise-reducting asphalt pavements for urban areas in Germany. Int. J. Pavement Eng., 13(6), 569–578.

[41] Ohrstrom, E., Hadzibajramovic, E., Holmes, M., & Svensson, H. (2006). Effects of road traffic noise on sleep: Studies on children and adults. J. Environ. Psychol., 26(2), 116–126.

[42] Okokon, E. O., Yli-Tuomi, T., Turunen, A. W., Tiittanen, P., Juutilainen, J., & Lanki, T. (2018). Traffic noise, noise annoyance and psychotropic medication use. Environ. Int., 119, 287–294.

[43] Peeters, B., Reinink, F., & Van Vliet, W. (2018). Close proximity (CPX) round robin test 2017. Paper presented at Euronoise 2018, Heraklion, Crete – Greece.

[44] PIARC, Working Group II of the Technical Committee TCE2 Environmental Considerations in Road Projects of the World Road Associacion (2019). Traffic noise – best practice guide. France: PIARC.

[45] Pirrera, S., De Valck, E., & Cluydts, R. (2014). Field study on the impact of nocturnal road traffic noise on sleep: The importance of in- and outdoor noise assessment, the bedroom location and nighttime noise disturbances. Sci. Total Environ., 500–501, 84–90.

[46] Sandberg, U. (1999). Low noise road surfaces – A state-of-the-art review. J. Acoust. Soc. Jpn. (E), 20(1), 1–17.

[47] Sandberg, U., & Ejsmont, J. A. (2002). Tyre/road noise reference book. Kisa: Informex.

[48] Sandberg, U., Buhlmann, E., Conter, M., & Mioduszewski, P. (2016). Improving the CPX method by specifying reference tyres and including corrections for rubber hardness and temperature. Internoise (pp. 4913–4923), Hamburg, Germany. Red Hook, NY: Curran Associates, Inc.

[49] Stansfeld, S. A., Clark, C., Cameron, R. M., Alfred, T., Head, J., Haines, M. M., Van Kamp, I., Van Kempen, E., & Lopez-Barrio, I. (2009). Aircraft and road traffic noise exposure and children’s mental health. J. Environ. Psychol., 29(2), 203–207.

[50] Tonin, R., & Szabo, A. (2014). RONDA – CPX trailer initial test results. Internoise (pp. 711–716), Melbourne, Australia. Red Hook, NY: Curran Associates, Inc.

[51] Transport Research Centre (2011). Methodology for measurement using Statistical Pass-By method and Close Proximity method. Brno: Transport Research Centre.

[52] Transport Research Centre (2014). Methodology for measuring and evaluation of noise on roads. Brno: Transport Research Centre.

[53] Transport Research Centre (2017). Long-term noise assessment of road pavements. Brno: Transport Research Centre.

[54] Vaitkus, A., Vorobjovas, V., Jagniatinskis, A., Andriejauskas, T., & Fiks, B. (2014). Peculiarity of low noise pavement design under Lithuanian conditions. Balt. J. Road Bridge Eng., 9(3), 155–163.

[55] Valentin, J., Mondschein, P., Bures, P., & Krivanek, V. (2017). Technical specification 259 – Asphalt mixtures for wearing courses with reduced noise levels. Prague: Ministry of Transportation.

[56] Van Essen, H., Van Wijngaarden, L., Schroten, A., Sutter, D., Bieler, C., Maffii, S., Brambilla, M., Fiorello, D., Fermi, F., Parolin, R., & Beyrouty, K. E. (2019). Handbook on the external costs of transport. Luxembourg: Publications Office of the European Union.

[57] Van Leewen, H., Kok, A., & Reubsaet, J. (2007). The uncertainty of acoustical measurements on road surfaces using the CPX-method. Internoise (pp. 4394–4399), Istanbul, Turkey. Red Hook, NY: Curran Associates, Inc.

[58] Vazquez, V. F., Teran, F., Luong, J., & Paje, S. E. (2019). Functional performance of stone mastic asphalt pavements in Spain: Acoustic assessment. Coatings, 9(2), 123–137.

[59] Vieira, T., & Sandberg, U. (2019). Close proximity (CPX) round robin test: Comparison of results from four different CPX trailers measuring noise properties of 10 Swedish road surfaces. Internoise (pp. 7176–7187), Madrid, Spain. Red Hook, NY: Curran Associates, Inc.

[60] World Health Organization (2018). Environmental Noise Guidelines for the European Region. Copenhagen: WHO Regional Office for Europe.

INTERVENTIONS TO IMPROVE PUBLIC TRANSPORT TRAVEL TIME AND RELIABILITY IN A NORWEGIAN CITY, AN ECONOMETRIC EVALUATION*

Knut Johannes L. Hartveit, Researcher
Askill H. Halse, Chief Research Economist
Nina Hulleberg, Researcher
Institute of Transport Economics
Gaustadalléen 21, 0349 Oslo, Norway
klh@toi.no

ABSTRACT

Policy makers in urban areas are in continuous search for policies that can improve the competitiveness of public transport vs. private cars. In this paper, we propose a framework for evaluation of smaller interventions and study the construction of a new bus lane, which was built to improve travel time and reliability of city buses in Bergen, Norway. To identify the effects of such an intervention, we use disaggregated operational data on travel time measured at each bus stop. This high data resolution enables us to identify the effects exactly where and when they occur.

We use different measures of travel time delay and variability to assess whether there are any problems before the intervention was done, and if these vary in certain areas or time periods. Then, we utilize the fact that the intervention is done in only one of the bus routes' directions to create a control group. That enables us to conduct difference-in-differences analyses to estimate the causal effects of the intervention.

The analyses show no apparent effect from the new bus lane. This is largely due to an installation of new active signal prioritizing technology in a signal-controlled intersection just over a month after the opening of the bus lane. This means that if the bus lane had a longer-term effect, this effect would be lost in the control variable we use to account for this technology.

Keywords:
Public transport, transport economics, econometrics, travel time reliability

1. BACKGROUND

In Norway there is a zero-growth goal which states that all growth in person transport in the larger urban areas is to be absorbed by public transport, bicycling and walking (Tønnesen et al. 2019). Therefore, it is essential that policy makers and public transport authorities makes the public transport as attractive as possible for users of private cars.

Time is a scare resource, so if a travel has longer duration, that has a cost for the traveller because that is time the traveller could have spent doing something else. This cost of travel time is normally called value of travel time, which depends on many factors, like delays or congestion. Time spent in congestion has a higher value of travel time compared to free-flowing traffic (see e.g. Wardman et al. 2016). A public transport sector with more delays and congestion driving, will hence be less appealing to people.

Additionally, travel time seems to be the most important aspect of inter-modal competition, followed by fare, access/egress/transfer time, and number of interchanges (Fearnley et al. 2018; Lunke et al., 2021a; Lunke et al., 2021b).

So, in order to achieve the zero-growth goal, it is paramount to be able to identify projects that reduce travel time and congestion, while at the same time being economically viable. If the authorities can choose projects that improve the attractivity of public transport and active transport in an economical sustainable way, the zero-growth goal is easier achievable.

In this paper, we propose a framework for evaluating of smaller interventions using disaggregated

* The paper departed from a larger, but more general report; Hartveit et al. (2020).

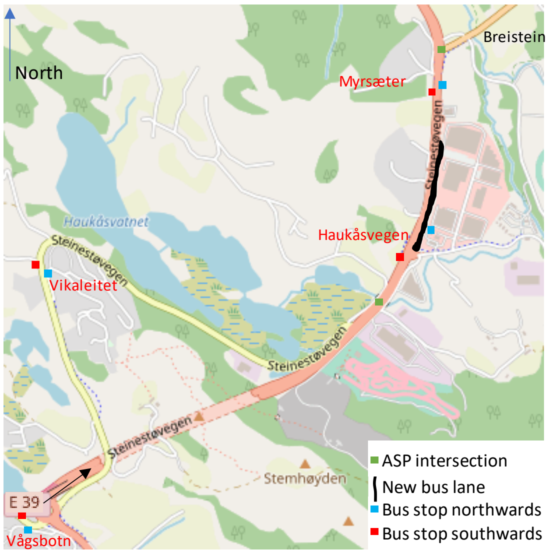
operational data for buses on travel time measured at each bus stop.

This high data resolution enables us to identify the effects exactly where and when they occur, compared to if we were to use total travel time, where the effects could be hidden due to timetable slack. It also enables us to illustrate and test which measures of travel time and reliability that should be used in these analyses.

The intervention we evaluate was done in Bergen, Norway to improve the travel time and reliability for city buses. We do this to give insights into how one can measure, and possibly select, projects for upgrades in a public sector with limited funds.

More specifically, we study if the construction of a dedicated bus lane at Haukås, north of Bergen has an effect on travel time and reliability for city buses.

The new bus lane goes through an intersection along the E39 in the northbound direction, and continues for approximately 350 meter north of the intersection, illustrated by the black marking in Figure 1 below.



Source: Map from OpenStreetMap.

Figure 1: Illustration of the influenced area.

There was no dedicated bus lane before the construction, so the buses are now driving adjacent to mixed traffic for the 350 meters opposed to earlier driving in the same lane as the other traffic.

If there was a congestion challenge or travel time challenge in the area of analysis, we would expect the buses to drive faster after the construction of the dedicated lane and reduce the travel costs for its passengers.

2. DATA AND DESCRIPTIVE ANALYSIS

The main data source we use is disaggregated operational data on travel time measured at each bus stop received from the public transport authority (PTA). This high-resolution data enables us to identify the effects exactly where and when they occur, compared to if we were to use total travel time, where the effects could be hidden due to timetable slack.

The data consists of when the buses are scheduled to arrive and depart bus stops (time and date), when the buses actually arrive and depart bus stops, journey ID, stop ID and information about passengers getting on and off the buses.

To ensure that we capture all seasonal effects, like holidays and vacations, we have chosen to have a year of data before the construction period (the before period) and a year of data after the construction period.

The bus lane was under construction from 01.09.2017-30.09.2018 (the construction period). Hence, our dataset spans from 01.09.2016 to 01.10.2019, with the before period being from 01.09.2016 to 31.08.2017 and the after period being 01.10.2018 to 01.10.2019.

We study two of the bus lines that are influenced by the intervention, called line 36 and line 37. These lines are chosen because they are similar in terms of frequency and passenger numbers, but they also have some differences in their routes, which adds to our understanding of the intervention's effect. Travelling northwards, line 36 drive along E39 the whole way, while line 37 drive via Vikaleitet, then onto E39, and then drive off E39 towards Breistein (reverse order when driving southwards).

In the influenced area, there are four bus stops we include in our analysis (Vikaleitet, Vågsbotn, Haukåsvegen and Myrsæter), illustrated in Figure 1. We choose these stops because they are the stops that give us the information about travel time and reliability we need for our analysis.

The road section northwards from the Vågsbotn/ Vikaleitet stop (south) to the Myrsæter stop (north) is the road section being affected by the new bus lane. The stop Haukåsvegen northwards (blue square in Figure 1) is situated along the new bus lane. Hence, the road section that is affected by the bus lane can be divided into two smaller sections: Vågsbotn/Vikaleitet to Haukåsvegen (southern) and Haukåsvegen to Myrsæter (northern).

The northern one is significantly shorter than the southern one, while the southern one also has a signalled intersection, a junction and different routes for the buses. Line 36 drive on the European route

E39, while line 37 drive on a slower, tighter road and has to enter E39 through the intersection.

2.1. Selection of control group

Choosing or finding a suitable control group is unfortunately not easy, and this applies here as well. There are multiple aspects and reasons for this. First, there is a geographic challenge, where types of roads and infrastructure could be different between the control and treatment lines.

There is also a demographic challenge, where different demographic areas could have different demand for public transport. It could also be temporary changes along one of the bus lines, like road maintenance or re-routing, that could interfere with the comparability of the bus lines.

We entertained the idea of using a similar line as a control group, but there were no other lines with the same frequency, number of passengers, demographics and driving conditions.

Therefore, we utilize the fact that the intervention was only done in one of the directions (northwards) and use the other direction (southwards) as the control group. This means that both groups are expected to have the same passenger growth and being affected by other factors, like weather conditions, in the same way.

At the same time, this means that the control and treatment group could have different driving conditions depending on time of day, but this is can be accounted for in the analyses.

2.2. Other relevant factors

It is important to know whether there has been other interventions or occurrences that happened in the period we are analysing that could affect driving time and reliability. Knowing this can enable us to control for these, and hence isolate the effect of the intervention we are analysing.

Therefore, in addition to the operational data, we have collected information and knowledge about overlapping events from the PTA, but also other official reports and information channels.

The first important other factor we need to be aware of, is the installation of technology for active signal prioritizing (ASP) of public transport in two signal-controlled intersections just south and north of the bus lane (the two green squares in Figure 1). With the new technology, sensors before the intersection pick up when buses are approaching and ensure that the buses get a green light earlier than with the older traffic adjusted signals.

The installation happened almost two months after the opening of the bus lane, so if ASP works and reduce the wait time for buses in signal controlled intersections, this would also affect the results in our analysis if we do not account for it.

Even though the installation happened at the same time, we would expect different effects from the southern and northern ASP. The one south of the bus lane is between the two stops on the southern sub-section, so if ASP has an effect, it will directly affect the driving time there, and influence our estimates of the new bus lane. The one north of the bus lane can affect the general traffic flow in the area, so it should be accounted for, even though we do not expect it to have nearly as strong an impact as the southern one.

Another factor is road tolls. Road tolls, or road use pricing, charges the people driving through certain areas a charge to pass through, and reduce traffic congestion (Hosford et al. 2021).

Three different changes in the road tolls happened in Bergen from 2016 to 2020. Time differentiated road tolls were implemented 01.02.2016, environmentally differentiated road tolls were implemented in 01.06.2018 and there were established more road toll stations which were operational from 06.04.2019. If these toll schemes have the anticipated effect, they will disturb the effect of the measure in our analysis if we do not account for it.

Since new toll stations or changes in the tolls are operational from a specific date, we assume that most of the effects arise close to the date they are operational. With this assumption, we do not include the time differentiation change in our analyses, as it happened more than eight months before the before period. The other two, on the other hand, are in the time period we are analysing, and are therefore included in the analyses.

2.3. Descriptive analysis

In the descriptive analysis, we use the whole period, meaning the before period (09.2016-09.2017), the construction period (09.2017-10.2018) and the after period (10.2018-11.2019).

Utilizing the richness of the data, we calculate the variables *travel time* and *driving time.* Driving time is the time the bus takes driving between the bus stops. Travel time is the time the bus takes driving from the first bus stop to the last stop in the affected area, including driving time and time spent on the bus stops.

This distinction is made due to the fact that the driving time variable could tell us about the challenges along the road, while the travel time also says something about the accessibility of the bus stops in terms of stopping, letting passengers on and off, starting to drive and get back onto the road with much traffic. Thus, we might capture some slack using the travel time variable.

The travel time and the driving time for line 36 in the three periods of analysis in monthly average, separating on buses going north and south, is illustrated in Figure 2 and Figure 3 below. We see that the red (north) and blue (south) lines seem to follow the same trend in the before period.

The dip we see in July in the before period, is the summer holidays in Norway, so less traffic congestion is to be expected. However, the fact that is seems to give such a strong effect on the travel time of buses, could indicate that there are some challenges.

Comparing the before and after period visually, it looks like the lines are flatter, converging and at a lower average in the after period. This could indicate that the bus lane has an effect on the travel time. It also seems to be a general downwards going trend (shorter travel times), so we should include a control for this in the regression analysis.

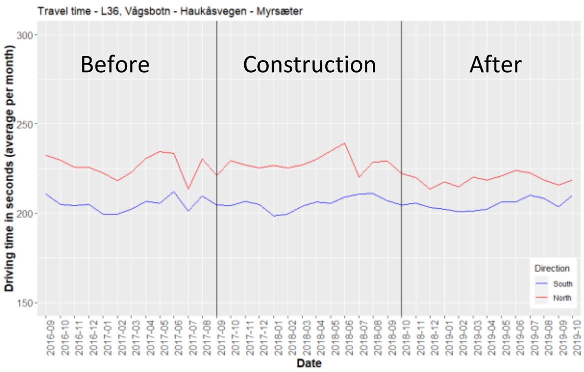


Figure 2: Travel time of line 36 in seconds, monthly average, over the course of the analysis period

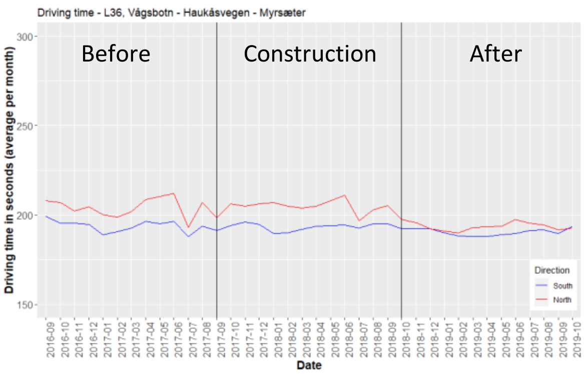


Figure 3: Driving time of line 36 in seconds, monthly average, over the course of the analysis period

For line 37, we see the travel and driving time in Figure 4 and Figure 5. Again, the control group (south) and treatment group (north) in the before period seems to be following the same trend, but maybe slightly less than line 36.

We see that the travel and driving time of the northwards going buses are more volatile. This could be due to travel time challenges, as the clear reductions are in the summer holidays, which normally has less traffic and travel time challenges.

Looking at the after period, there is a clear reduction in driving and travel time for the treatment group, while the control group looks to be unaffected. This indicates a strong effect from the intervention.

However, it should be noted that the treatment group in the first month in the after period seemingly has a slight reduction in driving and travel time, before the second month has a large reduction. We examine this further in the regression analyses.

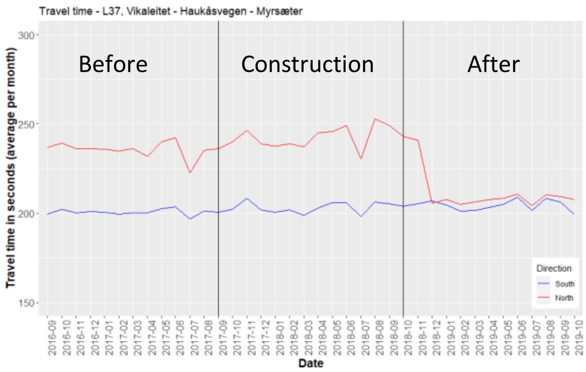


Figure 4: Travel time of line 37 in seconds, monthly average, over the course of the analysis period

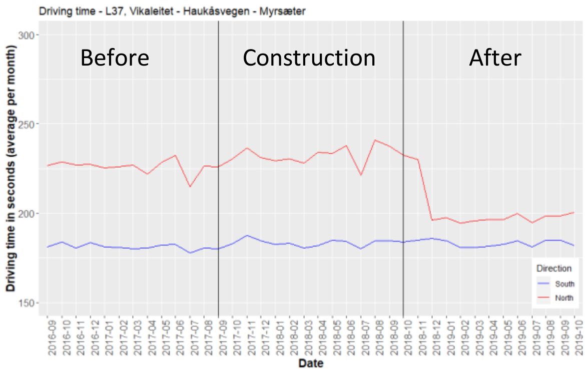


Figure 5: Driving time of line 37 in seconds, monthly average, over the course of the analysis period

3. METHODOLOGY

In this project, we suggest a framework for an ex ante evaluation of whether an intervention in a certain area could be beneficial and for an ex post evaluation of whether the intervention did have an effect. We call this the *before analysis* and the *regression analysis*, respectively.

In the before analysis, we want to figure out whether there are some travel time or reliability challenges in the area, before the bus lane was constructed. We also explore which measures of travel time or reliability are the best suited for these sorts of evaluations.

The regression analysis helps explaining what the patterns we saw in the descriptive analysis are caused by. Like with the descriptive analysis, we use the whole period (before, construction, after).

More specifically, the regression analysis we do is called Differences-in-Differences (DinD). DinD is a statistical method used for evaluating the effect of an intervention, a new law or regulation, or a natural event. A classic example of a DinD-analysis is Card and Krueger (1994). They studied how an increase in the minimum wage affected employment by utilizing that New Jersey's minimum wage rose, while neighbouring Pennsylvania had a constant minimum wage. Card and Kruger did the DinD-analysis by comparing the employment in fast food restaurants in New Jersey (treatment) with the one in Pennsylvania (control) before and after the increase.

These interventions or events makes it possible to identify a control group and a treatment group, where the treatment group is the group that was affected by the event (treatment). Observing the two groups before and after the treatment, we can measure the change in outcome of the two groups.

Furthermore, it is essential that the control group and the treatment group are as similar as possible. This is due to a fundamental assumption in this methodology, called the assumption of a common trend. It says that the change in outcome for the control group would be the same as the treatment group if it had received the same treatment. If it holds, we can estimate the causal effect of the intervention using DinD.

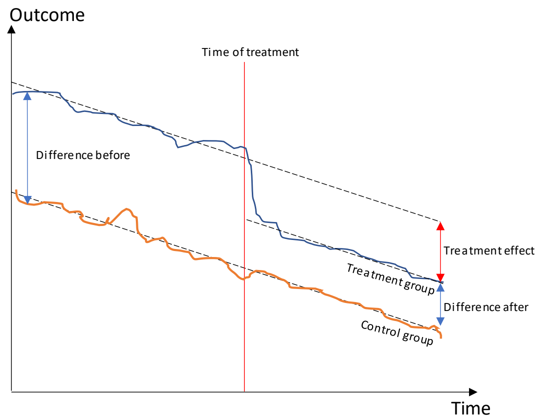


Figure 6: DinD illustration (inspired by Angrist and Pischke (2008))

An illustration of a DinD-analysis is showed in Figure 6. The dotted lines express an average level of driving time for the treatment group and the control group. If the treatment did not have any effect, we would expect to have the same difference between the treatment group and the control group. If the treatment has an effect, we could get something displayed in the graph, where we see there has been a downwards shift in the average travel time. The difference between the difference before and difference after is then the treatment effect (the red arrow).

In our analysis, we also have a construction period between the before and after period. We control for this in our analysis and compare the before and after period.

In the following, we present the before analysis and then the results from the regression analysis.

4. BEFORE ANALYSIS

In this analysis, we study how the different measures on travel time and reliability fare and what the different measures highlight. We also use these different measures to study whether it seems to be challenges related to driving time and reliability in the before period. These are presented in the following sub-sections.

The bus lane was constructed in the northwards facing direction to improve travel time, so if the bus lane indeed was constructed in an area with travel time challenges, we would expect to see this in our before analysis.

4.1. Delay relative to the timetable

A way of studying reliability is to see if the buses drive according to the timetable. A challenge we have here is that the stops according to the timetable are in minutes while we have actual arrival or departure times at bus stops down to milliseconds.

This means that we must use minutes as our analysis, and because of that, we do not capture variations in seconds due to rounding. When we consider that the average trip length between these stops are between one and five minutes, it is clear that this is not an optimal measure to use in our regression analysis where we expect changes in travel time of seconds, not minutes.

However, it does give a general overview of how many buses that are delayed at the bus stops. In Figure 7, we show the share of buses being on time (delay of 0-3 minutes) and the share of different levels of being delayed at the bus stops compared to the time table.

It shows that most of the departures are recognized as being on time when looking at the before period, however, we do see some delays with around 20 % of the buses being delayed three or more minutes.

Comparing the buses going northwards and southwards, we do not see a big difference.

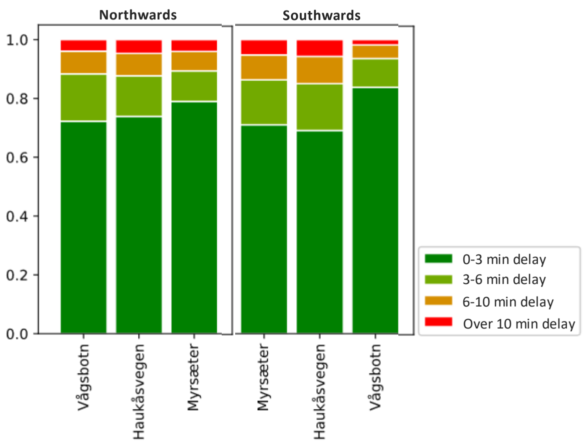


Figure 7: Share of departures that are delayed at departure time from the studied bus stops compared to the timetable of line 36

Figutre 8 shows the share of departures that are delayed at departure time from a certain bus stop compared to the timetable of line 37.

Like we saw for line 36, we see that a large proportion of the buses are departing on time, but there is now a large difference between buses driving northwards and southwards.

The buses going north from Vikaleitet are almost at the same level of delay as the buses going southwards, but when the buses leave the next stop (Haukåsvegen), the buses are significantly more delayed, and keep almost the same delay to the next stop (Myrsæter). In other words, the buses seem to get delayed when going from Vikaleitet to Haukåsvegen.

Looking at the same area for the southwards facing buses, they also have an increase in delay, going from Haukåsvegen to Vikaleitet, but it is smaller.

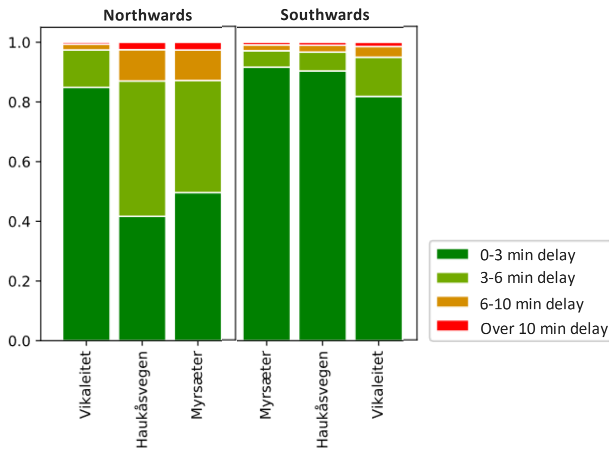


Figure 8: Share of departures that are delayed at departure time from the studied bus stops compared to the timetable of line 37

We have thus found that a significant share of the northwards facing buses on line 37 is delayed.

4.2. Variation in driving time

Another way to analyse reliability and travel time challenges is to see how the actual driving time of the buses between stops vary. This gives us a more precise way of looking at challenges along the road, as discussed in the descriptive analysis.

Figure 9 shows the variation in driving time between bus stops for line 36 northwards and southwards in a box plot for the whole day.

A box plot is a diagram that illustrates the median driving time (the horizontal line inside the coloured box) and how much variation there is in the data. The whiskers of the boxes show the 90th (top) and 10th (bottom) percentile, while the top of the box shows the third quartile and the bottom shows the first quartile. The coloured box hence shows the interquartile range.

Looking at the figure, we see that the road section between Vågsbotn and Haukåsvegen (blue and brown box) have the highest driving time. This is the longest road section, so that is expected. However, we see that the variation in the northwards driving buses is larger than the ones driving southwards, indication some challenges here.

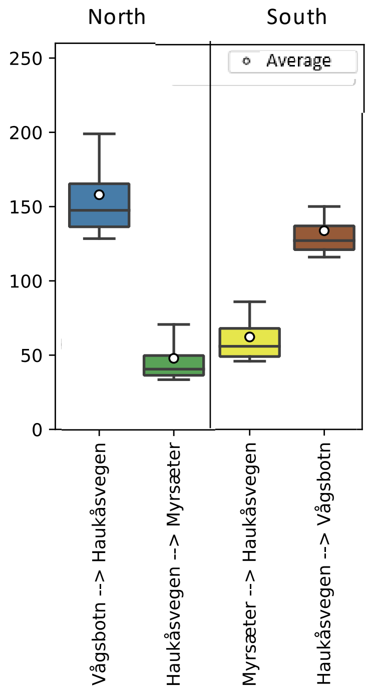


Figure 9: Variation in driving time (in seconds) between bus stops for line 36

Furthermore, we observe that the average value is somewhat higher than the median value. This means that there are some observations in the data set with a very long travel time, which draw up the average value of the section. This is not an unusual result, but it illustrates that one needs to be aware of the difference of using median and average values.

Figure 10 shows the variation in driving time between stops for line 37. We see that the road section Vikaleitet to Haukåsvegen (blue box) in the northward direction that has a large variation in driving time, compared to the other. This indicates quite clearly that there are driving time challenges on this road section.

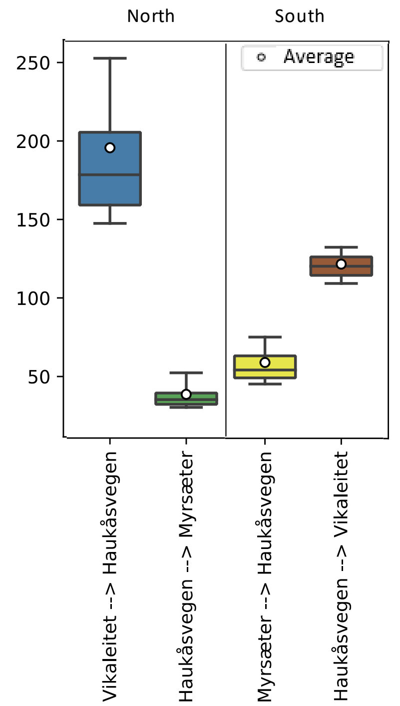


Figure 10: Variation in driving time (in seconds) between bus stops for line 37

For both line 36 and line 37, we observe a larger variation in driving time on the southern sub-section (Vågsbotn/Vikaleitet to Haukåsvegen), compared to the northern sub-section (Haukåsvegen to Myrsæter). Additionally, this variation is larger for line 37 than for line 36, and the difference between the northern route and southern route is also larger for line 37, which correspond well with what we found when using the measure *delay relative to the timetable*.

4.3. Driving time compared to zero-driving

Here, we focus on driving time and zero-driving. By zero-driving, we mean driving time without delays, which in our analysis is defined by the driving time of the 10th percentile.

The advantage of measuring deviations from the zero-driving the fact that we have this data down to seconds, and that median or average driving time will include some delays (to different extents) while zero-driving will be close to or equal to free flow conditions.

This is illustrated in Figure 9 and Figure 10, and we see that the sub-sections we previously identified with the largest challenges also has the largest difference between P10 and the average driving time.

4.4. Summary of the before analyses

In this chapter, we have focused on using different measures to discover whether there are driving time challenges in the before period in the areas that would later be affected by the new bus lane.

Both delay in relation to the time plan, variation in driving time and driving time in relations to zero driving suggest that there are travel time challenges for line 37 in the northwards facing direction on the first section (Vikaleitet – Haukåsvegen).

There are some signs of the same for line 36, but this is to a lesser extent, especially if we consider share of departures delayed compared to the time plan.

The second section (Haukåsvegen – Myrsæter) does not have the same clear signs of challenges, for both line 36 and line 37.

From our before-analysis it does look like there are clear reasons for trying to improve the travel times of the city buses in this area.

5. REGRESSION ANALYSIS

In this chapter, we perform regression analyses to assess if the intervention had an effect on travel time and improved the situation. Like with the descriptive analysis, we use the whole period (before, construction, after) and define the treatment group as the buses going northwards and the buses going southwards as the control group.

Our base Differences-in-Differences (DinD) regression is the following:

$$Driving\ time = \alpha + \beta_1 North + \beta_2 After + \beta_3 North * After + \epsilon \tag{1}$$

Where α is the constant, β_1 shows the effect of driving northwards or southwards, β_2 shows the effect of it being after the construction or not, β_3 is the DinD estimator which shows the effect of the intervention and ϵ is the error term.

As discussed earlier, there have been other interventions and changes happening in Bergen at the same time. To correct for the possible effects from changes in the road toll structure, we include indicator variables for new road toll stations (*New toll stations*) and environmentally differentiated road tolls (*Environmental differentiation*).

Another change happening in the period of this analysis is the installation of new technology in signal-controlled intersections. This happened during the second the month after the opening of the bus lane, so it must be accounted for, and is included as an indicator variable (*ASP*).

Furthermore, knowing the transport system, we know that there normally are daily variations, weekly variations and season effects, so this should also be accounted for. We include indicator variables for peak hours (Rush hours) with more traffic. The weekends (Weekend) are normally less trafficked and, like illustrated earlier, the summer vacation in July (July) is also likely to have an effect. Additionally, like we saw in the descriptive analysis, a time trend is likely, so we include that as well as a control variable (Time trend).

The regression results for line 36 and 37, with driving time and travel time, are displayed in Table 5. The most important numbers in the table is marked with a bold font.

The variable "*North · After*" shows the estimated effect of the intervention. It tells us that there was a significant reduction in driving time and travel time for line 36 northwards as a consequence of the new bus lane, 7.3 and 6.8 seconds respectively. Somewhat surprisingly, we see that the driving time and travel time for line 37 northwards seem to increase, 3.1 and 3.6 seconds respectively.

However, because we know that the ASP-technology was introduced just after the opening of the bus lane, we need to be careful in our interpretation of the *North · After*-variable.

The variable "*North · ASP*" shows whether something happened with driving time and travel time northwards in the period after the ASP-technology was installed. This is an important control of this analysis, as both line 36 and line 37 drive through the southern signal-controlled intersection with the new ASP-technology between the stops Vikaleitet/Vågsbotn and Haukåsvegen.

It shows that there was a large and significant reduction in driving and travel time for line 37, 33.8 and 35.4 seconds respectively. No significant effect for line 36 is found. This gives us a strong reason to believe that the ASP-technology helps line 37, that is driving onto E39 from a side road, to get through the intersection faster than before (see Figure 1 for repetition of locality). Line 36, on the other hand, was already driving on E39, and probably did not have to wait much for a green light before the measure, so even if the ASP did increase the efficiency of the intersection, it would not affect the driving time.

This means that if we had not controlled for the installation of ASP, we would have overestimated the effect of the bus lane. At the same time, if the effect of the new bus lane is not immediate, some or all of the effect of the bus lane will be caught in the *North · ASP* variable. Due to line 36 having a reduction of travel time from the bus lane and no effect from the ASP, supports this possibility.

The variable "*North · Construction*" shows whether there was a change in the construction period. We do not get a significant result for line 36, but we get a significant increase in driving and travel time for line 37, approximately 3.8 seconds for both. Looking at the descriptive graphs in Figure 4 and Figure 5, we can see traces of this.

Time dependent variables act as expected; driving and travel time is lower in July and in weekends, while there are significant increases in driving and travel time during rush hours.

If we assume that the bus lane has an immediate effect, and then can use the *North · After*-coefficients, line 36 seems to have a reduction in driving time, while line 37 has an increase in driving time. This seems like an implausible result due to the fact that the two lines are driving through the same bus lane, even though they have some differences before the junction and bus lane.

On the other hand, if we assume that the effect of the bus lane has a more gradual effect, it is more likely that both of the bus lines has a reduced driving and travel time, but that the *ASP*-coefficient includes some of this effect.

Lastly, it could very well be that these two interventions are multiplicative, meaning that the reduction in driving time from one of the interventions makes the other intervention more effective, but it could also be that they reduce each other's effect. This is something we have not studied, but it could be a factor in our results.

Table 1: Regression results

	Line 36		Line 37	
	DEPENDENT VARIABLE		DEPENDENT VARIABLE	
INDEPENDENT VARIABLE	Driving time	Travel time	Driving time	Travel time
North	11.507***	22.501***	46.798***	37.314***
	(0.694)	(0.844)	(0.453)	(0.537)
After period	-0.739	-2.687	-3.735***	-5.946***
	(1.475)	(1.794)	(1.195)	(1.417)
North · After	**-7.278*** **	**-6.750*** **	**3.121*** **	**3.559** **
	(1.466)	(1.783)	(1.183)	(1.403)
Construction period	1.007	0.715	-0.786	-1.656**
	(0.919)	(1.118)	(0.699)	(0.829)
North · Construction	**0.415**	**0.834**	**3.825*** **	**3.766*** **
	(0.877)	(1.066)	(0.625)	(0.742)
ASP	-2.714***	-2.267*	-2.329**	-1.574
	(1.049)	(1.276)	(0.909)	(1.078)
North · ASP	**-1.756**	**-2.530**	**-33.800*** **	**-35.382*** **
	(1.398)	(1.701)	(1.181)	(1.401)
New toll stations	0.813	1.955**	2.131***	3.206***
	(0.684)	(0.832)	(0.552)	(0.655)
Environmental differentiation	0.786	3.123***	2.971***	4.256***
	(0.754)	(0.918)	(0.585)	(0.694)
July	-7.170***	-7.446***	-8.300***	-10.527***
	(0.773)	(0.940)	(0.548)	(0.649)
Time trend	0.002	0.005*	0.005***	0.007***
	(0.002)	(0.003)	(0.002)	(0.002)
Weekend	-7.097***	-10.223***	-13.928***	-15.107***
	(0.441)	(0.536)	(0.324)	(0.384)
Rush hours	12.592***	18.695***	22.976***	26.387***
	(0.327)	(0.398)	(0.264)	(0.313)
Constant	157.161***	106.264**	86.897***	68.092**
	(37.912)	(46.112)	(27.773)	(32.938)
Observations	42,223	42,223	61,040	61,040
R^2	0.069	0.113	0.391	0.268
Adjusted R^2	0.069	0.113	0.390	0.268

** illustrate 10 % significance level, ** 5 % significance level, *** 1 % significance level*

6. SUMMARY

In this paper, we have proposed a framework for how to decide where to do an effort in improving travel times and reliability and how to evaluate these efforts or improvements.

The framework is utilized on an intervention to improve the travel time and reliability of city buses at Haukås north in Bergen, Norway.

Highly detailed operational data on travel time for buses allowed us to use different measures of travel time and reliability, and to capture small deviations in driving time.

The before analysis shows that the selected area seems like sensible area to try to improve the travel time. It demonstrates that line 37 has travel time challenges in the northwards facing direction, especially southern sub-section of the influenced area. Line 36 seems to have some challenges in the same area, but not as large.

This sub-section is the part of the area where the buses are influenced by the installation of the ASP-technology in the after period. The other section of the affected area does not seem to have any particular challenges in terms of travel time and delays. None of the bus lines seem to have any challenges when driving southwards.

The regression analysis shows that the driving and travel time is significantly lower after the opening of the new bus lane. However, due to the installation of the ASP-technology that got installed just over a month after the opening of the bus lane, we struggle to identify whether the reduction in driving time is due to the bus lane, the ASP-technology or a combination.

Combining the effect sizes of the intervention and the ASP, we get a reduction in driving time and travel time for the two bus lines. Line 37 has the largest reduction, which is not surprising due to 37 being the line with the most driving time variance and delays in the before period.

6.1. Reservations and further research

The data and methods used in this article seem to be well suited to measure travel time and delays, and to identify effects of interventions on short road sections.

What we do not know as much about is how possible time savings would be realized. Would it result in more waiting time at bus stops or slack in the timetables? The first does not seem to be the case, because then the variable travel time would have been more constant while the variable driving time would have been more varying, however, we cannot be sure.

This is possible to study more but could be demanding due to the nature of the changes in the time plan; they are adjusted more sporadically and is not easy to pin to a certain intervention.

Furthermore, we have in this article only studied a single intervention, so the external validity of this study could possibly be challenged. We have focused on developing a framework for evaluation and would strongly recommend doing more analyses of effects and economic profitability. This would be important both for getting better knowledge about how the public transport system works, and to ensure that the best steps are being taken.

REFERENCES

[1] Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

[2] Fearnley, N., Currie, G., Flügel, S., Gregersen, F. A., Killi, M., Toner, J., & Wardman, M. (2018). Competition and substitution between public transport modes. *Research in Transportation Economics*, *69*, 51-58.

[3] Card, David & Krueger, Alan B, 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," American Economic Review, *American Economic Association, vol. 84*(4), pages 772-793, September.

[4] Hartveit, K. J. L., Hulleberg, N., Halse, A. H., Fearnley, N., & Ingebrigtsen, R. Effekter av framkommelighetstiltak for kollektivtransport. *TØI rapport 1811/2020*

[5] Hosford, K., Firth, C., Brauer, M., & Winters, M. (2021). The effects of road pricing on transportation and health equity: a scoping review. *Transport Reviews*, 1-22.

[6] Lunke, E.B., Fearnley, N., Aarhaug, J., 2021a. Commute trips in Norwegian cities: Data combining trip characteristics and revealed mode choice, *Data in Brief*, https://doi.org/10.1016/j.dib.2021.107319

[7] Lunke, E.B., Fearnley, N., Aarhaug, J., 2021b. Public transport competitiveness vs. the car: Impact of relative journey time and service attributes, *Research in Transportation Economics*, https://doi.org/10.1016/j.retrec.2021.101098

[8] Tønnesen, A., Krogstad, J. R., Christiansen, P., & Isaksson, K. (2019). National goals and tools to fulfil them: A study of opportunities and pitfalls in Norwegian metagovernance of urban mobility. *Transport Policy*, *81*(C), 35-44.

[9] Wardman, M., Chintakayala, V. P. K., de Jong, G. 2016. Values of travel time in Europe: Review and meta-analysis. *Transportation Research Part A, 94*, 93-111.

A REVIEW OF B-WIM AS A SHM APPLICATIONS ON BRIDGES

Doron Hekič
Andrej Anžlin, PhD
Assist. Prof. Aleš Žnidarič
ZAG Slovenian National Building and Civil Engineering Institute
Dimičeva 12, Ljubljana, Slovenia
doron.hekic@zag.si, andrej.anzlin@zag.si, ales.znidaric@zag.si

ABSTRACT

Bridge Weigh-in-Motion (B-WIM) systems are installed on the bridges to weigh the passing vehicles in motion without traffic interruption. Besides weighing the freight traffic with equivalent accuracy to the pavement WIM, they can be designed to monitor bridge performance under traffic loading simultaneously. This is characterised by performance indicators (PI), most often influence lines, dynamic amplification factors (DAF) and load distribution factors. If suppose these PIs are adequately extracted from B-WIM measurements, the B-WIM takes the role of a Structural Health Monitoring (SHM) tool, in which the bridge performance monitoring can be correlated by the known axle loads of the crossing vehicles. While the traditional B-WIM systems measure strains in the bridge superstructure, some recent methods use accelerometer measurements. The conventional SHM systems typically include various types of sensors, including accelerometers, strain and temperature gauges, inclinometers, etc. They are generally installed on the most important or critical structures only, while B-WIM systems have the potential also for smaller, less important bridges. This paper aims to present the B-WIM capabilities, which are developing towards SHM and are no more limited to just weigh-in-motion the heavy traffic. In the beginning, the paper presents SHM and B-WIM definitions and reviews the PIs that can be extracted from B-WIM measurements. Later, selected studies where B-WIM is used for fatigue lifetime assessment are presented, and finally, two applications of B-WIM for damage detection reviewed.

Keywords:
B-WIM, SHM, sensors, bridges, viaducts

1. INTRODUCTION

The main part of the Structural Health Monitoring (SHM) literature was published in the last 30 years. Over these years, it came to the stage when these systems are installed on the majority of the extraordinary bridges, with respect to their dimensions, structural system or strategic position. SHM systems are nowadays due to misunderstandings of decision-makers about their complexity, installation and maintenance costs still limited to these special structures. One way to extend applications to the smaller bridges is to use Bridge Weigh-in-Motion (B-WIM) systems with some recently developed features, which are extending its primary purpose of weighing towards the structural health monitoring of the bridging structures.

2. SHM AND B-WIM DEFINITIONS

2.1. SHM definitions

SHM, in its most basic form, appeared parallel with structural engineering, although this expression was not used at that time and began to appear widely in the 1980s. The definition of the SHM nowadays, as stated by Boller (2009), is: "SHM is the integration of sensing and possibly also actuation devices to allow the loading and damaging conditions of a structure to be recorded, analysed, localised, and predicted in a way that nondestructive testing (NDT) becomes an integral part of the structure and a material". Wenzel (2009), when focusing on SHM of bridges, describes SHM as "implementation of a damage identification strategy to the civil engineering infrastructure. Damage is defined as changes to these systems' material and/or geometric properties, including changes to the boundary conditions and system connectivity. Damage affects the current or future performance of these systems".

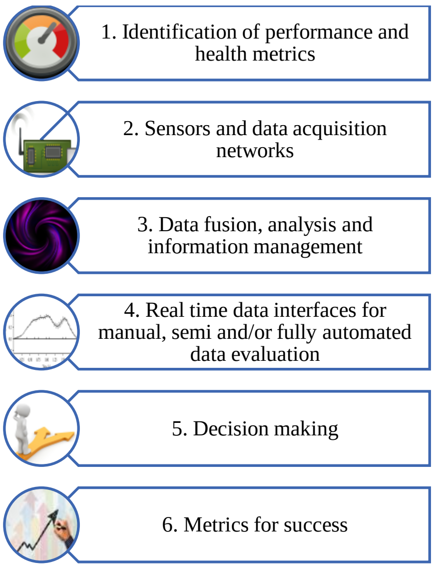
As stated by Catbas (2009), the following challenges are interrelated and have to be carefully addressed:

- fundamental knowledge needs,
- technology needs and
- socio-organisational challenges for routine applications.

Moreover, these challenges are linked with the main components of a complete SHM system, shown in Figure 1.

Similar to Rytter (1993), Catbas (2009) proposed to analyse SHM data at different levels to generate information about the health and performance of structures. Four levels were considered, which can be briefly described as:

- Level 1: Information obtained from the raw data without detailed analysis i.e. "Raw data indicators"
- Level 2: Usage of different analysis methods and different metrics to detect any damage, deterioration or "considerable change" in measurements over the monitoring period i.e. "Identification stage".
- Level 3: Localisation and quantifications of "considerable changes", damage or deterioration.
- Level 4: Prediction of future behaviour.



Adapted from:(Catbas., 2009)

Figure 1: Main components of a complete health monitoring design and challenges

2.2. B-WIM definitions

Žnidarič (2017) defines B-WIM as a specific method that uses an instrumented bridge or culvert to weigh in motion the crossing vehicles. The beginnings of the B-WIM date in the 1970s, when Goble et al. (1974) investigated measurements and processed bridge strain histories to analyse bridge fatigue life. After a comparison between the weights, processed from strain signals and weights, obtained from static weighing (Goble et al., 1976), it was evident that this method could weigh the passing vehicles. Soon after, Moses (1979) laid the foundations for B-WIM general principles that are still valid in today's B-WIM systems. Considering that the bridge during a vehicle passage oscillates about a static displacement position and linear behaviour of the bridge under traffic loading, then the resulting bending moment is described by the following equation:

$$M(t_j) = \sum_{i=1}^{N} A_i I(x) = \sum_{i=1}^{N} A_i I\left(v_i(t_j - t_i)\right); j = 1 \dots N_M \quad (1)$$

where:

$M(t_j)$ is bending moment at a time t_j,

A_i are (unknown) axle loads,

$I(x)$ are (known) values of bending moment IL; where x is location, calculated from measured axle velocity and axle arrival times,

v_i are axles' velocities,

t_i are arrival times of individual axles,

N is the number of axles,

N_M is the number of times in which bending moments are calculated.

Assuming a linear relationship between the moment M and measured strains ε_g overall N_G girders or sections of a slab, we can write the bending moments as:

$$M = CF \sum_{g}^{N_G} \varepsilon_g \quad (2)$$

where the average stiffness of the entire structure is represented by calibration factor CF, obtained experimentally using the calibration vehicles with their axle loads measured on the static scales.

The goal of the B-WIM analysis is to calculate the unknown axle loads A_i by minimising the difference between the measured and modelled bending moments.

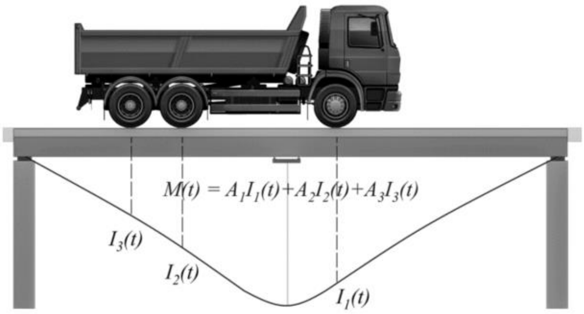
3. B-WIM PERFORMANCE INDICATORS (PI)

Evaluation of several structural performance indicators is essential to perform efficient B-WIM measurements that provide accurate weighing results. Over recent years, the calculation of these PIs has been elaborated to be suitable for SHM purposes. These PIs are influence lines, load distribution factors and dynamic amplification factors.

3.1. Bridge influence line (IL)

The bridge influence line describes the bridge response, expressed in strain, deflection, inclination etc., at a selected location under a concentrated unit load passing the structure (Figure 2).

Influence line was introduced by E. Winkler in 1867 as "a graph of a response function of a structure as a function of the position of a downward unit load moving across the structure" (Kassimali, 2009). For B-WIM analysis, an influence line is the primary structural parameter. It is multiplied by the axle loads, which after the summation of all axles' contribution gives the modelled bending moment caused by the vehicle, as shown by Equation (1).



Source: (Žnidarič & Kalin, 2020)

Figure 2: Using influence line to calculate the bending moment of a 3-axle vehicle

The first B-WIM systems used a theoretical influence line, which has for the simply supported superstructure and observed strains at the midspan, the well-known triangular shape. Nowadays, influence lines for B-WIM applications are always calculated from the measured bridge response to traffic loading. Two EC projects (WAVE (2001) and COST 323 (2002)) initiated this approach. Influence line can be obtained either using passages of the trucks with known axle loads (matrix method) or crossing trucks of unknown axle loads that are part of the B-WIM process (ZAG method). The matrix method (OBrien et al., 2005) performs a linear optimisation to find the best fit of the influence line ordinate at each location across the bridge, for which vehicles with known axle loads are needed. The ZAG method (Žnidarič & Kalin, 2020) uses random vehicles from the ongoing traffic flow and performs a nonlinear optimisation to calculate the shape of the influence line. One should note that this procedure is the same as when weighing, except that while weighing, the influence line is known, whereas when calculating it, the influence line is a part of the unknowns to be calculated.

ZAG method, also implemented in the SiWIM® system (Žnidarič et al., 2011), is constructed from cubic splines. These are defined by the following control points:

- 'Support' points with abscissas defined by bridge geometry and zero ordinates,
- 'M' point, located typically where the highest strains are expected,
- 'Radius' points, which define the radius *d/*2 of the rounded cap in the vicinity of the 'M' point,
- Optional points, with abscissas defined and ordinates to be optimised, for fine-tuning the match between the measured and modelled response.

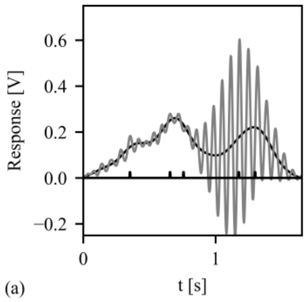
Such an influence line, described in detail in (Žnidarič & Kalin, 2020), has the advantage of using random traffic flow instead of a limited set of preweighed vehicles. Furthermore, this paper also demonstrates a promising potential to detect structural damages based on permanent monitoring of their shape over the structural lifetime.

3.2. Girder distribution and Lane factors

Girder distribution factor (GDF), traditionally used in the USA, describes how loading is divided to individual beams, girders or sections of a slab. In bridge design codes, such as AASHTO (2012), values of GDFs are specified for different types of beam-slab bridges. When performing B-WIM measurements on beam-slab bridges, GDFs for girders, with B-WIM sensors installed, can be obtained. This performance indicator is calculated from the measured responses of the bridge caused by individual vehicles to obtain statistically defined contributions of traffic loads from various traffic lanes on the measured girder. For example, if the bridge carries two lanes of traffic and has four instrumented girders, then 8 GDFs are calculated. While assessing the safety of existing bridges, the engineers are looking for hidden reserves in structural performance. By considering the GDFs from the in-situ measurements rather than the design codes, they more realistically assess the structural behaviour, which allows them to optimise the structural models. In a similar way, GDF under certain lanes can be merged into Lane factors (Enright & OBrien, 2013). A more detailed explanation of how GDFs can be obtained from B-WIM measurements can be found in (Žnidarič, 2017) and (Žnidarič & Kalin, 2020).

3.3. Dynamic amplification factor (DAF)

The dynamic component of bridge traffic loading is commonly taken into account with a Dynamic Amplification Factor (DAF) – the ratio between the dynamic and static load effects on a bridge. In some cases, a resonance effect can drastically amplify the response of a bridge, as can be seen in Figure 3. Therefore, this coefficient is one of the most relevant performance indicators when reliability analysis of existing structures is considered and when some hidden reserves from the traffic load perspective need to be taken into account.



Source: (Kalin et al., 2021)

Figure 3: Example of a high dynamic response of a bridge under traffic loading

Replacing conservative DAF values, which are defined in some current provisions (AASHTO, 2017), with measured data, can significantly improve bridge performance assessment. Recently a new method of calculation of this factor based on B-WIM measurements had been developed (Kalin et al., 2021). An automated 2-phase method that calculates DAF values from all heavy vehicle crossing events was developed. In the first phase, the optimum filtering frequency for each event is calculated by minimising the differences between the filtered measured and calculated response. Based on the complete population of calculated optimum frequencies, the mean cut-off frequency f_0 is calculated. In the second phase, the complete dataset of events is filtered with the frequency f_0, and the DAF of any event k is calculated as:

$$DAF_k = \frac{max|response_k|}{max|filtered\ response_k(f_o)|} \quad (3)$$

This approach removes the need for an expert who selects the parameters based on personal experience. The proposed method is based on data from 15 different bridges, which invariably demonstrates that a DAF decreases with increasing bridge traffic loads.

4. FATIGUE LIFE ASSESSMENT OF BRIDGES USING B-WIM

4.1. Virtual monitoring to predict the remaining life of steel bridges

As a part of EC 7th Framework Programme project Bridgemon, B-WIM and SHM were for the first time coupled to evaluate the fatigue life of a cable-stayed bridge (Hajializadeh et al., 2017). The idea of the virtual monitoring approach is to use a B-WIM system to weigh the axles of vehicles that pass the bridge and then use this information as an input to the calibrated numerical model, in which strains at the desired points (in this case, near fatigue-prone locations) are obtained.

In this study, an orthotropic deck bridge over a canal was installed with a B-WIM system that collected the traffic loading data for 38 days. Due to the bridge complexity, the B-WIM sensors were installed only on the simply supported span part of the bridge. 24 sensors were installed near the midspans and 4 sensors (for axle detection) near the abutments. Simultaneously, a set of additional strain gauges were installed on the main span and on the pylon to calculate the correlation between traffic loads and the fatigue of the orthotropic deck. The authors concluded that in the ideal circumstances, the same sensors would serve both to weigh-in-motion the traffic and for finite element (FE) model calibration required for the implementation of virtual monitoring.

Iterative FE model calibration consisted of calibration of the total response, where model parameters, such as material properties, were varied and final calibration to match the measured strain records at various locations on the bridge. After that, traffic simulation from 38 days of measurements to 15 years was performed using Monte Carlo simulation, and fatigue lifetime estimation for fatigue-prone locations was done.

In addition to confirmation of conservativism in the Eurocode's fatigue load model, the authors concluded that the virtual monitoring concept has the potential to optimise maintenance costs due to avoidance of unnecessary repairs or even replacements of the bridge structures.

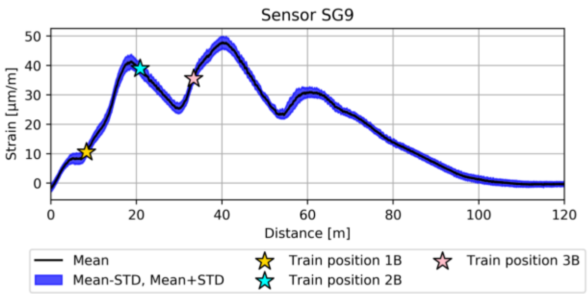
Similar study on the Orthotropic Steel Deck (OSD) bridge was performed by Zhang et al. (2019). To predict the strains at the locations where sensors are not installed, the authors used the bridge's finite element model, which was updated based on modal testing. With such a model, it was possible for every point on the bridge to determine the influence surface, which, together with known axle positions and loads, served to calculate the fatigue damages without a sensor installed at the specific location.

For fatigue evaluation, authors used AASHTO codes and assumed that annual average daily traffic would remain constant (as recorded during the experimental campaign) over its design life, which was 100 years. It was found that the considered bridge would not fail in fatigue in the most fatigue-prone measurement point, not even after crossing of a recorded 148 t vehicle. This (overweight) truck was almost 2.5 times heavier than the design truck. The authors concluded that the proposed approach of virtual monitoring was feasible for fatigue evaluation of orthotropic steel bridges.

4.2. Improved fatigue consumption assessment through structural monitoring

In a recent study performed within the Assets4rail project, funded by EU Horizon 2020 Shift2Rail Joint Undertaking (Anžlin et al., 2019; Anžlin et al., 2020), authors used B-WIM together with additional sensors for estimating the remaining fatigue lifetime of a railway bridge. While for fatigue lifetime estimation similar concept of virtual monitoring as by Hajializadeh et al. (2017) was used, both response (strains) of the bridge during the train passage and response (natural frequencies) during shaker test were used to calibrate the numerical model.

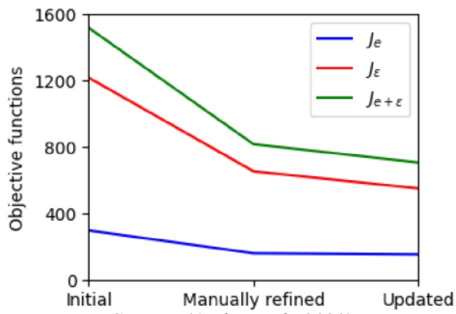
For B-WIM purposes, two strain gauges were mounted near the midspan of the secondary longitudinal beam (stringer), which supports sleepers, and additional two strain gauges at the bottom of the rails for axle detection. Along with other sensors, B-WIM sensors were used for FE model calibration and estimated fatigue lifetime prediction of the selected fatigue-prone details. In this study, the FE model calibration procedure consisted of manual calibration and calibration using the optimisation solver. Furthermore, the objective function within the optimisation solver has not just considered the natural frequencies from the modal response (during the shaker test), but the response during the soft load test (SLT) (Figure 4) as well. A detailed description of the SLT can be found in (SAMARIS D30, 2006).



Source: (Anžlin et al., 2020)

Figure 4: Measured response in the space domain for strains at selected sensors during the SLT

Figure 5 shows the evolution of the objective functions J_e (modal response), J_ε (response during the soft load test) and $J_{e+\varepsilon}$ during model updating.



Source: (Anžlin et al., 2020)

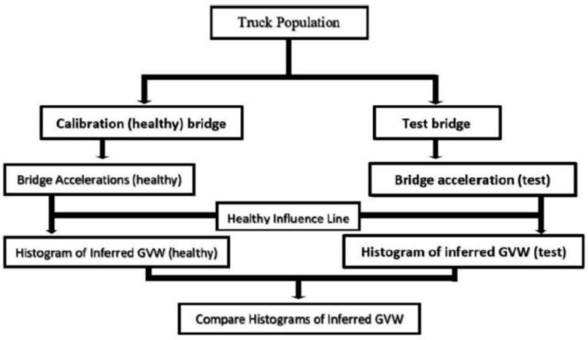
Figure 5: Evolution of the objective functions J_e, J_ε and $J_{e+\varepsilon}$ during model calibration

This study confirmed that B-WIM system is suitable for weighing in motion the railway traffic. Analogous to the B-WIM for the roadway bridges, COST323 (2002) specifications were used to determine accuracy. Accuracy classes were A(5) for the GVW of carriage A in both directions and the second carriage in one direction, while for the opposite direction, the accuracy class for carriage B was B+(7). Since only the GVWs from the static weighting for carriages A and B but not for C and locomotive (these were obtained from the technical spreadsheets) were known, accuracy classes were defined only for the GVW of carriages A and B. One should note that in this study axial load of traffic could be measured, but unfortunately no static axial load data were provided.

It was also shown that for model calibration, both the modal response and the response from the SLT could be taken into account. Furthermore, all sensors installed for B-WIM, except axle detection sensors, were used for model updating.

5. ACCELERATION-BASED B-WIM FOR DAMAGE DETECTION

Besides conventional B-WIM systems, which measure relative strains at the superstructure to weigh the traffic, new concepts appeared recently, such as acceleration-based B-WIM. To use a bridge acceleration response in a B-WIM system, the relationship between bridge accelerations and vehicle weights has to be considered. In this way, the B-WIM system can become more suitable for monitoring of structures (OBrien et al. 2020) since, in the case of strain measurements, the response only changes if the sensor is located in the vicinity of the damaged point. This also implies that an acceleration-based SHM requires a smaller number of sensors to identify bridge damage. Acceleration based B-WIM uses an interesting concept (Figure 6): the inferred GVWs should be repeatable, but in case of change of global stiffness, different GVWs can be measured. Suppose an adequate deduction of the impact of temperature on the variation of global stiffness can be performed. In that case, one could use a statistical approach and compare the mean inferred GVWs to estimate the health of the monitored bridge. A recent study based on numerical simulations and field testing (OBrien et al. 2020) has shown that a hybrid SHM and acceleration based B-WIM gives promising results.



Source: (OBrien et al., 2020)

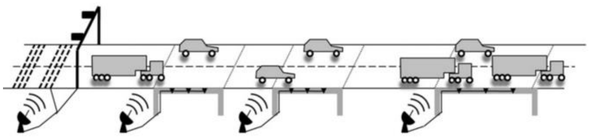
Figure 6: Schematic diagram of the methodology of bridge damage detection using the acceleration-based B-WIM system

6. BRIDGE DAMAGE DETECTION USING WIM TECHNOLOGY

Cantero & González (2015) proposed a novel bridge damage detection method (Level 1) for short to medium span road bridge using both (pavement-based) WIM system located near the bridge, and a B-WIM system, installed on the bridge. Authors showed that the ratio of the estimations of vehicles weights by both WIM and B-WIM system is a reliable and robust indicator of structure's condition and furthermore that this indicator is more sensitive to damage than the traditional method based on variation in natural frequencies.

This study aimed to use pavement-based WIM, installed near the bridge, and B-WIM, installed at the considered bridge, for gross vehicle weight (GVW) estimation (Figure 7). Both systems give GVW estimation, and comparison between them in time could lead to the identification of the occurrence of bridge damage in time.

Using the FE model, the authors simulated a vehicle traversing a bridge. For the bridge, they have considered two types of boundary conditions; simply supported and fixed-fixed. A four degree of freedom two-axle system was used to simulate the vehicle.



Source: (Cantero & González, 2015)

Figure 7: WIM-based damage-identification concept for SHM of bridges

For traffic population generation, a Monte Carlo simulation was performed considering normally distributed parameters of vehicles (mean and standard deviation), typical for two-axle trucks. For damage simulation, authors modelled global damage and local damage separately. Global damage was taken into account as a reduction in stiffness of all elements, and local damage was modelled as a reduction in stiffness at only one particular element.

The authors proposed E_{BWIM} indicator as a new tool to monitor structural changes. This tool compares GVW, estimated on the damaged bridge and GVW of the same vehicle, estimated on the undamaged bridge (i.e. at the pavement WIM station near the bridge). They concluded that this indicator:

- allows distinguishing between global and local damages,
- can roughly estimate the location of damage,
- has both for global and local damage situations greater sensitivity to the damage occurrence than the traditional Level 1 damage-identification technique, which is based on tracking frequency changes.

7. CONCLUSIONS

Comprehensive SHM systems were in the past mainly present in academia and on some landscape structures. On the other side, simpler SHM applications could be extended to ordinary bridges by using recently developed acceleration-based B-WIM systems, where among other things, the automatic evaluation of bridge performance indicators is performed.

To obtain PIs, one does not need (besides B-WIM sensors) to install additional sensors. This approach is different for the "virtual monitoring", where additional sensors are usually needed for sufficient calibration of FEM model. For acceleration-based B-WIM, typical strain sensors are replaced by acceleration sensors, the measurements of which are used as an input for the conventional (strain-based) B-WIM algorithm. For bridge damage detection using WIM technology, besides the B-WIM system installed on the bridge, an additional pavement WIM system should be installed near the bridge to monitor the change of axial load due to damage of the structure.

In general, most SHM's are not capable of adequately detecting bridge damages. But given recent experiences, one can state that acceleration-based B-WIM is more effective in the detection of damages than strain-based one. It was therefore concluded at this point to push the future research activities in the direction towards developing hybrid B-WIM that would, besides strain signals, also consider acceleration signals. However, for more complex bridges such as cable-stayed bridges with orthotropic steel deck, a larger set of sensors than for B-WIM will always be necessary. Particularly if one is interested in observing the total behaviour or FE model calibration for purposes such as fatigue evaluation.

Based on the presented studies, we can conclude that B-WIM systems with PI calculation and some additional sensors could offer an alternative to the conventional SHM systems. They in general require big initial financial bite and can therefore discourage bridge owners for the implementation of these systems. Therefore, the proposed simpler monitoring system could be more extensively used on smaller, often older and deteriorated bridges. Today, actions on those bridges are mostly based on condition assessment and not actual performance under traffic loads.

REFERENCES

[1] AASHTO. (2012). AASHTO LRFD Bridge Design Specifications, 6th Edition 2012. The American Association of State Highway and Transportation Officials, Inc., Washington, D.C.

[2] AASHTO. (2017). AASHTO LRFD Bridge Design Specifications, 8th Edition 2017. The American Association of State Highway and Transportation Officials, Inc., Washington, D.C.

[3] Anžlin, A., Hekič, D., Kalin, J., Ralbovsky, M., & Lachinger, S. (2019). Section-specific fatigue load models: method and application, Shift2Rail (Grant Agreement Nr. 826250 "Assets4rail") Deliverable D3.2.

[4] Anžlin, A., Hekič, D., Kalin, J., Kosič, M., Ralbovsky, M., & Lachinger, S. (2020). Improved fatigue consumption assessment through structural and on-board monitoring, Shift2Rail (Grant Agreement Nr. 826250 "Assets4rail") Deliverable D3.3.

[5] Boller, C. (2009). Structural Health Monitoring—An Introduction and Definitions. In Encyclopedia of Structural Health Monitoring. https://doi.org/https://doi.org/10.1002/9780470061626.shm204

[6] Cantero, D., & González, A. (2015). Bridge damage detection using weigh-in-motion technology [Article]. Journal of Bridge Engineering, 20(5), Article 04014078. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000674

[7] Catbas, F. N. (2009). 1 - Structural health monitoring: applications and data analysis. In V. M. Karbhari & F. Ansari (Eds.), Structural Health Monitoring of Civil Infrastructure Systems (pp. 1-39). Woodhead Publishing. https://doi.org/https://doi.org/10.1533/9781845696825.1

[8] Enright, B. & OBrien, E. J. (2013). Monte Carlo simulation of extreme traffic loading on short and medium span bridges. Structure and Infrastructure Engineering. 9(12), pp. 1267-1282.

[9] Goble, G., Moses, F., & Pavia, A. (1974). Field Measurements and Laboratory Testing of Bridge Components - Final report.

[10] Goble, G., Moses, F., & Pavia, A. (1976). Application of a Bridge Measurement System. Transportation Research Record No. 579, Washington DC, 36-47.

[11] Hajializadeh, D., OBrien, E., & O'Connor, A. (2017). Virtual Structural Health Monitoring and Remaining Life Prediction of Steel Bridges. Canadian Journal of Civil Engineering, 44. https://doi.org/10.1139/cjce-2016-0286

[12] Kalin, J., Žnidarič, A., Anžlin, A., & Kreslin, M. (2021). Measurements of bridge dynamic amplification factor using bridge weigh-in-motion data. Structure and Infrastructure Engineering, 1-13. https://doi.org/10.1080/15732479.2021.1887291

[13] Kassimali, A. (2009). Structural Analysis. Cengage Learning.

[14] Moses, F. (1979). Weigh-in-motion system using instrumented bridges. Journal of Transportation Engineering, 105(3).

[15] OBrien, E., Khan, M. A., McCrum, D. P., & Žnidarič, A. (2020). Using Statistical Analysis of an Acceleration-Based Bridge Weigh-In-Motion System for Damage Detection. Applied Sciences, 10(2), 663.

[16] OBrien, E., Quilligan, M. J., & Karoumi, R. (2005). Calculating an Influence Line from direct measuremetns. Proceedings of the Institution of Civil Engineers, Bridge Engineering 159(1), 31-34.

[17] Rytter, A. (1993). Vibrational Based Inspection of Civil Engineering Structures. Dept. of Building Technology and Structural Engineering, Aalborg University.

[18] SAMARIS D30. (2006). Guidance for the optimal assessment of highway structures. ZAG.

[19] WAVE. (2001). Bridge WIM - Report of Work Package 1.2. University College Dublin & ZAG.

[20] Weigh-in-Motion of Road Vehicles: Final Report of the COST 323 Action. Jacob, B., OBrien, E.J., Jehaes, S. (Eds.). (2002).

[21] Wenzel, H. (2009). Bridge Management and Health Monitoring. In Health Monitoring of Bridges (pp. 7-18). https://doi.org/https://doi.org/10.1002/9780470740170.ch2

[22] Zhang, L., Zhao, H., OBrien, E. J., & Shao, X. (2019). Virtual Monitoring of orthotropic steel deck using bridge weigh-in-motion algorithm: Case study. Structural Health Monitoring, 18(2), 610-620. https://doi.org/10.1177/1475921718764081

[23] Žnidarič, A. (2017). Influence of number and quality of weigh-in-moiton data on evaluation of load effects on bridges, Doctoral disertation. University of Ljubljana.

[24] Žnidarič, A., & Kalin, J. (2020). Using bridge weigh-in-motion systems to monitor single-span bridge influence lines [Article]. Journal of Civil Structural Health Monitoring, 10(5), 743-756. https://doi.org/10.1007/s13349-020-00407-2

[25] Žnidarič, A., & Lavrič, I. (2011). SiWIM Data Post-Processing Manual.

IMPROVING THE ANNUAL AVERAGE DAILY ROAD TRAFFIC ESTIMATION BY USING MULTITEMPORAL HIGH RESOLUTION AERIAL IMAGERY

Jens Hellekes, M. Sc.
Remote Sensing Technology Institute, German Aerospace Center (DLR)
Muenchener St. 20, 82234 Wessling, Germany
jens.hellekes@dlr.de

ABSTRACT

The annual average daily traffic (AADT) is a key indicator for describing transport development and an essential parameter for traffic engineering and transportation planning. Commonly, the AADT is collected by permanent counting stations on selected sections of the major road network and supplemented by temporary counts at multi-annual intervals. Because of the high expenditures for construction and maintenance, dense networks of counting stations are rarely established which leads to uncertainties about the spatial and temporal distribution of vehicle volumes. Remote sensing data can be part of the solution as it can be acquired flexibly and over large areas. This paper utilizes aerial image sequences to capture traffic density as well as vehicle speeds to determine the traffic flow on motorway and federal road segments. The instantaneous observations at different days and times are extrapolated to the AADT using time series from reference counting stations in the study area of Braunschweig (Germany). Discrepancies of up to 11 % between estimated and measured AADT are found at three of five metering sections. For the remaining sections, this paper elaborates why limits of the approach are reached here. The presented method is particularly suitable for motorway sections where no or only insufficient data are available.

Keywords:
Traffic Monitoring, Vehicle Trajectories, Traffic Flow, AADT, Remote Sensing, Aerial Image Sequences

1. INTRODUCTION

Road traffic counts and derived quantities such as traffic volumes or annual vehicle miles travelled are used worldwide to monitor transport development and to accompany measures from design through assessment to their evaluation after implementation. Apart from the collection for statistical purposes (International Transport Forum, 2019), detectors provide essential input data for traffic management systems (Vortisch, 2006) and for quantifying externalities such as noise emissions (Bajdek, et al., 2015). The counts are used by road administrations to schedule maintenance cycles and to develop construction proposals (Bosserhoff et al., 2006). Both infrastructural and policy measures are usually evaluated with transport models, which in turn are often validated or calibrated using traffic counts (Cascetta, 2009). With these models, traffic forecasts can be made with regard to various scenarios (Ortúzar & Willumsen, 2011).

In Germany, about 2,100 permanent counting stations record vehicles on the major road network, with approximately 80 % located on motorways and federal roads. The remaining detectors cover state and municipal road sections that are of particular importance for transport analysis. The continuous metering is supplemented by temporary measurements at five-year intervals: according to schemes, vehicles are counted manually or by measuring devices on several days at specific hours. These data are then extrapolated to the annual average daily traffic (AADT) to obtain an estimate of the vehicle volume until the next counting period (Schmidt et al., 2020). Outside the coverage area of the stationary detectors, uncertainties arise with regard to the growth of the traffic volume and the spatial and temporal distribution (Jiang et al., 2006). This is particularly problematic if a measure for which significant effects are expected is implemented between the surveys: short-term effects are not recorded and medium- to long-term effects can only be partially attributed to the measure due to interference with other factors.

For states that lack detailed count data but still face current traffic problems, exploring new data sources is promising. Traffic-related applications are frequently discussed in the remote sensing community because data can be collected flexibly and cover large areas. Numerous works deal with road traffic monitoring by detecting motorized vehicles on aerial and satellite images or videos (Apeltauer et al., 2015; Azimi et al., 2018; Mishra, 2012) and deriving metrics describing the current traffic situation (Azevedoet al., 2014; Larsen et al., 2013; Yang et al., 2016).

Coifman et al. (2006) have investigated the extent to which extrapolation to unobserved conditions from optical remote sensing data is possible. For this, they used a drone to record vehicle trajectories over a 2-hour period and derived the AADT. Jiang et al. (2006) demonstrated that inferring vehicle volumes from single aerial images can improve the AADT update for motorways based on temporary counts. Kaack et al. (2019) created a processing chain that detects trucks on satellite imagery and uses a regression method to estimate AADT for road freight transport.

The main contribution of the paper consists of the exploration of image sequences for AADT estimation, which are acquired on two different days and times. Imagery of Braunschweig (Germany) is used to count the number of vehicles on five major road sections and extract their trajectories. The traffic situation at the acquisition time is extrapolated to AADT estimators using hourly time series for the corresponding year. The variations and the gain in robustness are examined when the estimators are combined. In addition, different traffic situations are considered: while mainly motorways with low or medium traffic density have been studied so far, here, on the one hand, the end of a motorway controlled by traffic lights is analyzed. On the other hand, the AADT is estimated for an urban state road with high traffic density and low speeds due to a construction site. For a first assessment of the accuracy of the estimated AADT, comparisons are made with the vehicle counts of the corresponding detectors.

The paper is structured as follows: First, the dataset of aerial images leading to the selection of measurement sections to be analyzed is presented. Then, the methods for obtaining trajectory data, deriving traffic flows at the time of observation, and extrapolating the AADT are described. In the results section, similarities and – in case of deviations between estimated and reference AADT – probable reasons are deducted. Identified opportunities as well as limitations of the approach are discussed and finally an outlook on possible improvements is given.

2. DATA ACQUISITION

The German Aerospace Center (DLR) utilizes the city of Braunschweig as a test bed for transportation research. Within various projects, parts of the city were mapped by airplane over the period of several years with the DLR 3K camera system (Kurz, et al., 2012). The flights took place at an altitude of about 700 m above ground, resulting in a resolution of approximately 9 cm. A repetition rate of 1 Hz leads to an overlap of two consecutive images of about 80 %. In several campaigns, flight strips were flown twice within a 15 min interval in order to be able to perform change analysis over short periods of time. Combined with precise ground control points and GNSS/inertial data at the time of recording, the images are georeferenced and projected onto a digital terrain model. Due to the high spatial resolution in combination with the capture frequency, it is possible to determine velocities of moving objects very precisely: in the nadir view, speed differences of up to 0.36 km/h can thus be detected.

Two flights are selected for this paper, differing in date (2019-04-25, 2019-10-30) and time (06:45 to 07:30 pm, 09:45 to 10:30 am). The images were not taken with a focus on capturing the major road network (motorways, federal and state roads), but for mapping urban, suburban and industrial areas of about 40 km². However, due to the building structure, several such roads are covered in the flight area and are selected for analysis if they simultaneously fall within the validity range of a counting station. In

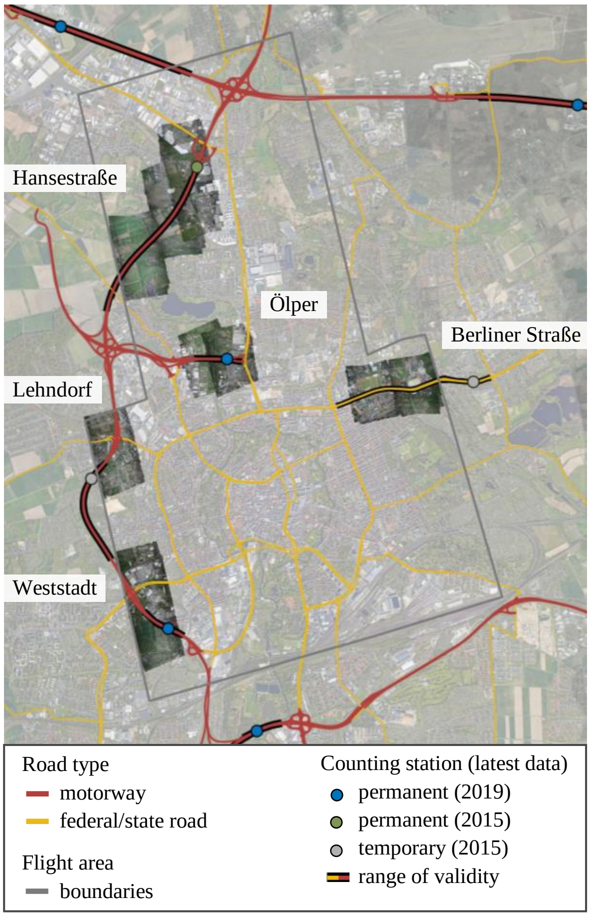


Figure 1: Overview of the study area in Braunschweig (Germany), primary road network, location of counting stations and image sequences for analysis

Figure 2: Motorway end Ölper, controlled by traffic lights, 2019-10-30

Figure 3: Congestion due to construction works at Berliner Straße, 2019-10-30

total, the road sections under analysis are shown in 648 images. An overview of the study area is displayed in Figure 1.

With the exception of *Berliner Straße* (this is a state road), motorway sections are selected. The *Ölper* measuring section is a special case, as the motorway ends here and the transition to the inner-city road network is controlled by traffic lights (see Figure 2). Furthermore, construction work is taking place on *Berliner Straße* during the second flyover period, resulting in narrowing to one lane and congestion (see Figure 3). These features can be considered as challenging when it comes to AADT extrapolation.

For the AADT estimation, hourly time series per year of the counting stations are required; in addition, the measured AADT are used for reconciliation with the estimated values. Hourly loads are only available for the *Weststadt* and *Ölper* sections for 2019. A permanent counting station was operated in the *Hansestraße* section until 2015 – more recent values are not available. For the *Lehndorf* and *Berliner Straße* sections, only AADT values determined on the basis of temporary counts can be obtained (Federal Highway Research Institute, 2020). The range of validity per counting section shown in Figure 1 is determined in the preparations for the national road traffic counts. In the case of motorways, the validity range extends from the on-ramp to the next off-ramp in each direction of travel. As it can be seen in Figure 3, access ways and intersections may occur on state and federal roads where vehicle volumes may change. In general, it is defined that a measurement section on federal and state roads extends to the next junction of a road of equal or higher category. Deviations from this are possible, for example, if a facility with high attraction rate causes a significant change in the traffic volume.

3. METHODS

Cross-sectional counts by stationary induction loops usually deliver aggregated data to traffic management systems. The macroscopic quantities traffic flow Q, occupancy and (in case of double-loop detectors) the mean speed V for a time period are provided. Numerous applications rely on this level of aggregation, so the method developed in this paper should also provide data at this level of granularity.

The movement of individual vehicles expressed by trajectory data can be transformed into macroscopic quantities by referring on the movement of the vehicle collective. This can be described by the hydrodynamic relation:

$$\rho(x,t) = \frac{Q(x,t)}{V\,(x,t)} \qquad (1)$$

The traffic density ρ is defined as the spatial average of the number of vehicles on a given road segment x at a fixed time t. The traffic flow Q can be determined from the number of vehicles crossing x in t. Since ρ is defined as the spatial mean, the mean speed V also enters as spatial mean. However, most permanent counting stations can only measure the temporal average of speeds, which results in imprecise estimates of macroscopic parameters (Treiber & Kesting, 2013).

The continuity equation (1) applies only for the entire unidirectional road section, not to individual lanes, given that lane changes are possible. The conservation property of (1) is violated when substantial changes in ρ and V happen. This is the case, for example, when the number of vehicles changes as a result of accesses and egresses, or road gradients or regulation significantly alternate the mean speed V. This leads to the condition that the road section under analysis must be microscopically small. At the same time, the section cannot be chosen arbitrarily small, since macroscopic quantities can be derived meaningfully only in the case of several vehicles (Treiber & Kesting, 2013).

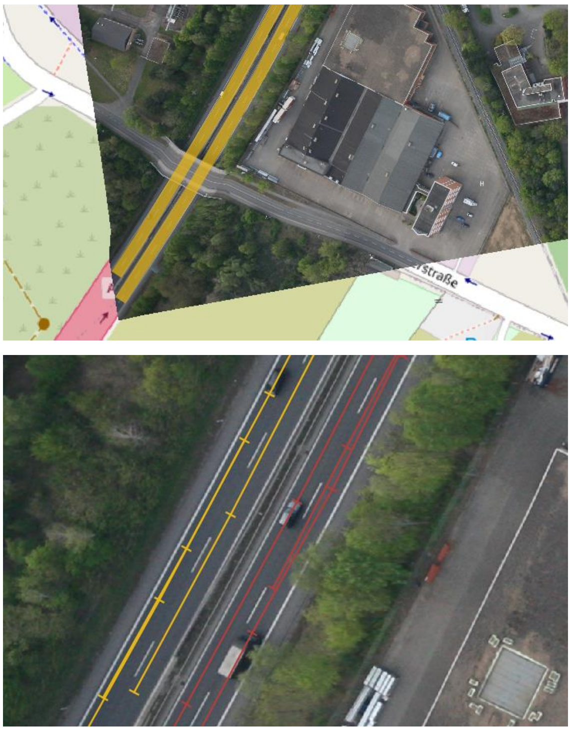


Figure 4: Aerial image crop with highlighted measurement section for vehicle counting at reference point in time t^* (top) and zoomed-in sample for vehicle trajectories, derived over image sequence (bottom)

In addition to these requirements, there is an important hypothesis for the derivation of the traffic flow Q and subsequently the extrapolation to the AADT: the traffic density ρ and the mean speed V at the time of observation are sufficiently good estimators for the analyzed time slice. For further analysis, this hypothesis is assumed to be fulfilled; the hypothesis is evaluated in the end.

3.1. Generating trajectory data from aerial image sequences

According to estimation theory, randomized sampling is required to obtain an unbiased estimator. Here, randomization refers to the selection of the specific route segment and the time of observation. However, the selection of aerial imagery is limited and, at the same time, it must be assumed that vehicles are not homogeneously distributed in the section due to the tendency for pile-ups. Therefore, the randomization is approximated by selecting as reference point in time t^* the aerial image of a sequence which shows the road for the first time with the greatest coverage. The condition regarding the coverage can be justified as follows: at high speeds and correspondingly large headways on motorways, it may otherwise be possible to evaluate not enough vehicle trajectories. However, due to implications of (1), the section length must be limited to the homogeneous part. From aerial imagery, a change in the number of lanes is clearly apparent, while variations in the speed limits are not visible. These

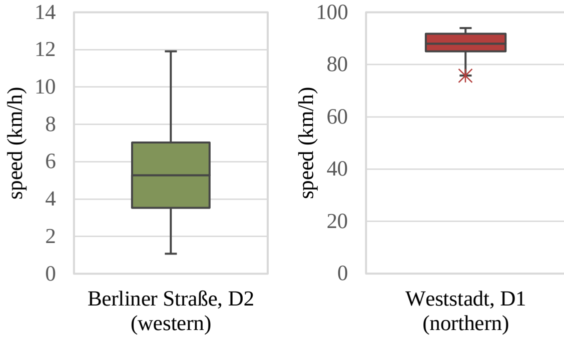
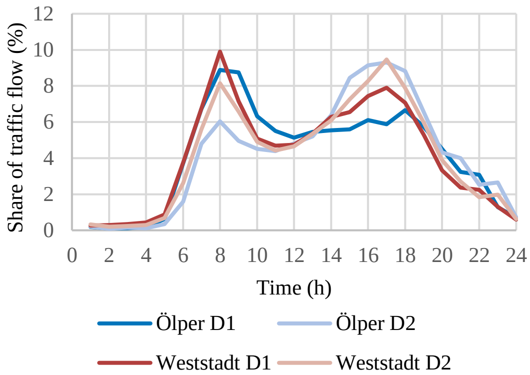


Figure 5: Boxplots of determined vehicle speeds per trajectory element, for selected direction of travel and observation, 2019-10-30



Source: traffic flows published by Federal Highway Research Institute (2020)

Figure 6: Daily time series for selected counting stations at 2019-04-25, adopted from 2018-04-25

uncertainties must be taken into account in the trajectory analysis. Figure 4 (top) shows an exemplary aerial image crop at t^* with part of the measurement section drawn in.

This reference image is also used to determine the number of vehicles for which trajectories are derived. A vehicle is considered as object of observation if the front bumper is located in the measurement section. By evaluating the image sequence around t^*, the spatial positions can be determined over time. It is also possible to include vehicles in the analysis that are occluded by bridges at t^*, for example, but are still in the measurement section. Depending on the flight direction and the road orientation, it is possible that vehicles at the edge of the section can only be seen at t^*. In this case, the velocity of the nearest vehicle was transferred to the corresponding vehicle. This issue occurred in 9 cases, which represents about 2 % of all vehicles analyzed and is considered low for this study. The resulting trajectories are exemplary shown in Figure 4 (bottom).

3.2. Deriving spatial mean speed V

Another implication of (1) concerns the measurement of the spatial mean speed V. This is a temporal instantaneous mean. This can only be inferred as an approximation from the aerial image sequences, since no telemetry data of the vehicle is transmitted. The approximation can be judged as good on theoretical grounds: the length of sequence Δt is due to the frame rate and overlap between subsequent images small. This applies in particular to a flight direction orthogonal to the course of the road section. Empirically, this is also evident: the coefficient of variation of the speeds of the individual vehicle is 0.01 at the median, with a maximum of 0.83. For the vehicle with the highest observed coefficient of variation, the standard deviation of the speed is about 3 km/h and occurs on 2019-10-30 at the section *Berliner Straße* due to congestion. Other relatively high coefficients of variation can also be attributed to the traffic jam, as well as to the inflow of the traffic light controlled *Ölper* junction (see Figures 2 and 3).

Figure 5 shows that deriving speed from trajectory data offers two advantages: On the one hand, it allows the traffic situation to be taken into account. Due to the congestion on *Berliner Straße*, vehicles drive at considerably lower speed than permitted (50 km/h). A naïve transfer of the speed limit would otherwise significantly overestimate the traffic flow. On the other hand, it can be seen from Figure 5 that at the time of image acquisition, the speed limit of 80 km/h on *Weststadt* section is exceeded by many vehicles at low traffic density. Trajectory data can be used to consider this when estimating traffic flow.

3.3. Extrapolating measured data to AADT

Using the known number of vehicles per measurement section as well as the trajectories, equation (1) can be used to calculate the traffic volume estimator for the analyzed time slice. The calculation procedure for estimating AADT per section and direction follows the approach commonly used in Germany (Schmidt et al., 2020): in a first step, the daily traffic volume of the counting day is extrapolated by means of hour/day factors. They can be derived by the daily time series exemplarily shown in Figure 6. In the second step, the extrapolation to AADT is done by day/year factors. A differentiation is made according to trip-purpose groups (working day, Sunday, holiday).

Unfortunately, there is no cross-sectional data by permanent counting stations for 2019, so that time series from 2018 (and 2015 for *Hansestraße*) must be used. For the section *Lehndorf*, a propagation takes place to the effect that the time series for *Weststadt* can be observed in a similar way there. An analogous traverse line approach is chosen for *Berliner Straße*, which presumably shows the greatest similarity to *Ölper*. If the same date of the 2015/2018 year is not part of the same trip-purpose group, the closest same day of the week was set. This represents a first approximation that can be refined by further investigation.

4. RESULTS

A total of 66 image sequences were evaluated, showing an average of 360 m of the measured sections. Overall, 452 vehicles were counted, whose trajectories could be tracked over 5 images on average. In conjunction with the described extrapolation method, the following information are extracted: AADT estimators for the first and second observation per day and from these derived estimates by averaging the observation values per day and all days, respectively. Table 1 presents an example for *Weststadt*. It can be seen that the AADT estimators vary considerably when extrapolating on the basis of a single observation. At the same time, it can be shown: by summarizing individual measurements, both within a day and across all observations, the estimate stabilizes and, in particular, approximates the presumed true AADT (see Table 2). Also, the directional reference of the vehicle flows is better represented by averaging: according to the values of the year 2018, no significant difference in AADT per direction has been observed.

There are some aspects to consider when comparing the estimated with the reference AADT: on the one hand, the time series for extrapolation and the absolute values originate from previous years. Changes are conceivable, for example, due to traffic volume growth, infrastructural changes or different weather conditions. This tends to have a greater effect when there is a large gap between the reference and observed year. On the other hand, parts of the reference values are to be questioned with regard to their accuracy: for the counting station *Hansestraße* it is stated that the time series was estimated. For the *Ölper* section, it is reported that the AADT was calculated with a limited data basis (Federal Highway Research Institute, 2020). Due to the propagation of the *Ölper* time series for *Berliner Straße*, this also has an effect here.

The deviations between the estimated and measured AADT for the *Ölper* and *Weststadt* sections are 5 and 10 %, respectively, which is considered a good agreement. This demonstrates: special cases like the *Ölper* motorway end are not problematic as such. For the section *Berliner Straße* there are substantial differences (factor 3 between estimated and measured AADT). This could have several reasons: first, there are no information on the extent to which the *Ölper* time series reflects usual traffic patterns on *Berliner Straße*. It can be assumed that the

Table 1: Estimated AADT on different aggregation levels for counting station Weststadt

Date	Estimated AADT for observation no., per direction							
	1		2		average, per day		average, all observations	
	D1	D2	D1	D2	D1	D2	D1	D2
2019-04-25	37,572	46,217	31,279	34,693	34,426	40,455	29,396	37,204
2019-10-30	31,426	29,692	17,307	38,214	24,366	33,953		

Table 2: Estimated AADT for all observations in 2019 and reference AADT for 2018 (unless otherwise noted)

Counting station	Direction, AADT type					
	D1		D2		both	
	estimated	reference	estimated	reference	estimated	reference
Berliner Straße	15,982	-	16,472	-	32,455	10,658*
Weststadt	29,396	36,644	37,204	37,066	66,000	73,710
Hansestraße	47,649	33,032*	31,299	33,173*	78,948	66,205*
Lehndorf	51,615	-	-	-	-	71,000*
Ölper	11,398	12,751	13,118	13,171	24,516	25,922

*Source: reference AADT from Federal Highway Research Institute (2020). Note: values marked with * reflect the status as of 2015*

construction site and the resulting traffic jam on 2019-10-30 caused a spatial and possibly temporal change in route choice, so that the transferability of the time series is lower. Second, even with the 2019-04-25 data (no narrowing), AADT estimates are showing to be too high, so the section characteristics could also be contributing to the discrepancies. In the urban area, traffic lights are closely spaced, resulting in pile-ups of vehicles due to acceleration and deceleration phases in quick succession. Thus, this could represent a borderline case in the applicability of equation (1).

The estimated AADT for *Hansestraße* seems to be significantly to high compared to the measured AADT. This is put into perspective when the increase in vehicle volume since the year 2015 is considered: For *Weststadt*, a 7 % AADT increase can be observed between 2015 and 2018. Transferring this to Hansestraße as a proxy for the growth rate, the gap between estimated and measured AADT is reduced to about 11 %. The AADT estimate for *Lehndorf* is not very meaningful, because – as can be seen in Figure 1 – the range of validity falls only slightly into the flight area. In addition, only one image sequence covers the section, which contains zero and three vehicles per direction, respectively. Therefore, in one direction the AADT cannot be calculated, while in the opposite direction of travel the sample size is very small.

5. DISCUSSION AND OUTLOOK

The AADT estimators obtained on the basis of single aerial image sequences are subject to large fluctuations, which can be attributed to the short observation time of less than 10 seconds on average per section. Nevertheless, the combination of up to four acquisition dates turns out to be promising given the stabilization of the calculated AADT. Considering the good agreement between the estimated and measured AADT for *Weststadt* and *Ölper* sections, and in light of numerous factors of uncertainty regarding the reference data (not available everywhere or only for earlier years, partly estimated based on short-term counts, not reflecting road works), the methodological assumption made before can be justified: traffic density ρ and the mean speed V are sufficiently good estimators for the vehicle volume in the analyzed time slice. It also shows, that traffic patterns are reasonably constant and therefore transferable to subsequent years. However, refinements of the extrapolation approach would be valuable for estimating AADT, especially to avoid the propagation of time series of nearby counting stations and instead to apply a series type based on a classification scheme.

Limitations of the method can be seen in two aspects: first, a single overflight at a time of day with low traffic volume increases the risk of not encountering the route section with any vehicle. The AADT would then be wrongly estimated to zero. Secondly, the demonstrated method can only be applied to homogeneous sections. Particularly in the inner-city area, these sections are sometimes short due to access ways and junctions, so that in combination with pile-ups due to traffic lights strongly varying AADT estimators are determined. Further investigations are needed to explore the limits of applicability in urban areas.

For a large-scale application of the approach, the manual detection of vehicles and the extraction of their trajectories is not practical. Remote sensing methods can be further developed to automate this process. Depending on the study area, other remote sensing platforms may be more suitable for collecting the data, so the processing chain can be designed for generalizability.

6. CONCLUSION

Less than 10 seconds on average – this is how long motorways and federal state roads in Braunschweig (Germany) are covered by an image sequence dataset taken by an aircraft. By counting the number of vehicles and extracting their trajectories at different times and dates, traffic flows are determined. Permanent and temporary counting station data in the study area are used to extrapolate the point-in-time observations to robust estimators for the AADT. At the same time, the detector counts serve as a reference to assess the accuracy of the estimate.

For two measurement sections, there is a deviation of 5 and 10 %, respectively, compared to the metered AADT of the previous year (data of the same year were not available). If the growth in traffic volume is considered, a similar variation (11 %) could be achieved for a third metering section, for which reference data recorded four years earlier were provided. A federal road section in the urban area shows significant deviations, which are attributed to both a short evaluation segment between two traffic light controlled junctions and a reduction in lanes due to a construction site.

According to these findings, the presented method is suitable for provide timely AADT estimators with focus on motorways. The approach may be of particular interest to states that do not have a well-developed network of permanent counting stations supplemented by temporary measurements. Third parties that do not have access to counting data may also benefit.

REFERENCES

[1] Apeltauer, J., Babinec, A., Herman, D., & Apeltauer, T. (2015). Automatic vehicle trajectory extraction for traffic analysis from aerial video data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-3/W2*, 9–15.

[2] Azevedo, C. L., Cardoso, J. L., Ben-Akiva, M., Costeira, J. P., & Marques, M. (2014). Automatic Vehicle Trajectory Extraction by Aerial Remote Sensing. *Procedia – Social and Behavioral Sciences*, *111*, 849–858.

[3] Azimi, S., Vig, E., Kurz, F., & Reinartz, P. (2018). Segment-and-Count: Vehicle counting in aerial imagery using atrous convolutional neural networks. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-1*, 19–23.

[4] Bajdek, C., Menge, C., Mazur, R. A., Pate, A., & Schroeder, J. (2015). *Recommended Best Practices for the Use of the FHWA Traffic Noise Model (TNM).* US Federal Highway Administration.

[5] Bosserhoff, D., Follmann, J., & Klotz, S. (2006). Verkehrserhebungen: Grundlagen und Grundbegriffe. In V. Bach, D. Bosserhoff, J. Follmann, S. Klotz, & W. Reuter (Eds.), *Handbuch für Verkehrssicherheit und Verkehrstechnik* (pp. 91–112). Hessisches Landesamt für Straßen- und Verkehrswesen.

[6] Cascetta, E. (2009). *Transportation Systems Analysis: Models and Applications.* Springer.

[7] Coifman, B. A., McCord, M. R., Mishalani, R. G., Iswalt, M., & Ji, Y. (2006). Roadway traffic monitoring from an unmanned aerial vehicle. *IEEE Proceedings – Intelligent Transport Systems*, *153*(1), 11–20.

[8] Federal Highway Research Institute (2020). *Automatische Zählstellen 2018.* https://www.bast.de/BASt_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/Aktuell/zaehl_aktuell_node.html?nn=1819516&cms_strTyp=A&cms_map=0

[9] International Transport Forum. (2019). *Glossary for transport statistics* (5th ed.).

[10] Jiang, Z., McCord, M. R., & Goel, P. K. (2006). Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data. *Journal of Transportation Engineering*, *123*(7), 523–530.

[11] Kaack, L. H., Chen, G. H., & Morgan, M. G. (2019). Truck Traffic Monitoring with Satellite Images. *Proceedings of the ACM SIGCAS Conference on Computing and Sustainable Societies*, 155–164.

[12] Kurz, F., Türmer, S., Meynberg, O., Rosenbaum, D., Runge, H., Reinartz, P., & Leitloff, J. (2012). Low-cost optical Camera Systems for real-time Mapping Applications. *Photogrammetrie – Fernerkundung – Geoinformation, 2012*(2), 159–176.

[13] Larsen, S. Ø., Salberg, A.-B., & Eikvil, L. (2013). Automatic system for operational traffic monitoring using very-high-resolution satellite imagery. *International Journal of Remote Sensing 34*(13), 4850–4870.

[14] Mishra, R. K. (2012). Automatic Moving Vehicle's Information Extraction from One-Pass WorldView-2 Satellite Imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XXXIX-B7*, 323–328.

[15] Ortúzar, J. d., & Willumsen, L. G. (2011). *Modelling Transport* (4th ed.). John Wiley & Sons.

[16] Schmidt, G., Frenken, T., & Mahmoudi, S. (2020). *Straßenverkehrszählung 2015: Methodik der manuellen Zählungen.* Fachverlag NW.

[17] Treiber, M., & Kesting, A. (2013). *Traffic Flow Dynamics. Data, Models and Simulation.* Springer.

[18] Vortisch, P. (2006). *Modellunterstützte Messwertpropagierung zur Verkehrslageschätzung in Stadtstraßennetzen.* Institute for Transport Studies, University of Karlsruhe.

[19] Yang, T., Wang, X., Yao, B., Li, J., Zhang, Y., He, Z., & Duan, W. (2016). Small Moving Vehicle Detection in a Satellite Video of an Urban Area. *Sensors*, *16*(9), Article 1528.

ASSESSING THE PERFORMANCE OF A COLD RECYCLED PAVEMENT WITH FOAMED BITUMEN UNDER ACCELERATED PAVEMENT TESTING (APT) WITH MOBILE LOAD SIMULATOR (MLS30)

Mehdi Kalantari, Msc.Eng.
Design and Structure of Pavements, German Federal Highway Research Institute (BASt)
Bruederstr.53, 51427 Bergisch Gladbach, Germany
Kalantari@bast.de

ABSTRACT

In recent years by increasing the demand on more sustainability in transportation sector, different solutions have been defined and applied on production process of pavement's materials and also in their construction activities. Among different methods, cold recycling has the potential of using higher rates of recycled material and at the same time, lower production temperature. During the last decade, different laboratory researches and monitoring activities were performed on this technology in Germany. Positive international experiences reported on bitumen stabilized materials (BSM) application, especially with foamed bitumen, gained interest in Germany in recent years. In 2018 a research project on topic cold recycling with foamed bitumen was defined in German Federal Highway Research Institute (BASt) in cooperation with the Wirtgen GmbH. The main goal of the project was to gather more information on the behavior of cold recycled material with foamed bitumen by monitoring its response and performance under accelerated loading. The findings of this project will be combined with previous research results for updating and further development of national guidelines and specifications. A 100-meter test section was built at the demonstration, investigation and reference areal of BASt (duraBASt) with two different pavement types: one conventional type (as reference) and one with cold recycled layer. The sections were loaded with the mobile load simulator MLS30 and the performance was monitored with different methods. This paper presents some results and findings from the non-destructive (FWD and rutting measurements) monitoring and laboratory tests on the extracted cores at different time periods.

Keywords:
cold recycling, bitumen stabilized material (BSM), foamed bitumen, cold recycled pavement performance, accelerated pavement testing (APT), duraBASt, MLS30

1. INTRODUCTION

Increasing demand of more sustainability in transport infrastructure sector has affected the construction and maintenance methods of the pavements. Nowadays, sustainability plays an important role in developing, evaluating and selecting between different construction and maintenance options. Recycling techniques are one of the well-known sustainable methods which decrease the negative environmental effects of the pavement's construction and rehabilitation. It is possible to classify asphalt recycling methods in to three big categories of hot, warm and cold methods. The main difference between these methods is in the production temperature of the mix. Cold recycling methods are performed in ambient temperatures and as there is no need to heat the aggregate mix, the energy consumption and emission production is noticeably lower than the other recycling methods. As the mixture is produced cold, there is no need to heat the RAP (reclaimed asphalt pavement) which, is the big issue in adoption of asphalt plants for hot recycling process. This makes the production process relativly simpler and also eases the possibility of using higher rates of RAP.

To be able to mix the bitumen with aggregates in ambient temperature, the viscosity of the bitumen should be decreased. At the meantime there are two methods for that. Suspending the bitumen droplets in water (bitumen emulsion) or making foamed bitumen. By injecting a little amount of water in to the hot bitumen, the water droplets will evaporate and lead to foaming of the bitumen (Wirtgen, Wirtgen Cold Recycling Technology, 2012). The produced foam will stay for some seconds and this is enough to be able to mix it with cold and wet aggregates. Beside the bitumen, normally cement or hydrate lime is added as the second binder to increase the moisture resistance of the recycled material and its early life strenght. The compacted mix gains strength and stiffness over time mainly by evaporation of water and hydration of the cement (known as curing in the technical language).

Like other countries, cold recycling has been applied in Germany, but in a limited amount. The main reason can be the lack of national-level performance data on the behavior of the recycled material which, has affected the existing national guidelines too. During the last decade, different laboratory researches and monitoring activities were performed on this technology in Germany (Wacker, Kalantari, & Diekmann, 2020; Kalantari, 2020). The results of these activities and the reported positive international experiences, led in to gaining of more interest in Germany on this type of cold recycling. In 2018 a research project on the topic of cold recycling with foamed bitumen was defined in German Federal Highway Research Institute (BASt) in cooperation with the Wirtgen GmbH aiming to gather more data on the behavior of this material. A test lane was planned to be constructed and tested in the BASt's demonstration, investigation and reference areal named as duraBASt.

duraBASt is the outdoor test facility of BASt which is located east of Colonge parallel to A3 motorway north direction intersection with A4 motorway with a total length of approx. 1,000 m and area of 25,000 m^2. The construction of the test facility took place between 2015 and 2017. There are three main areas in duraBASt as the reference area (R areal) which is for the approval of the surface characteristics measuring vehicles and then the demonstration and investigation areas (D/U areal). They consist of nine lanes, six of them are located in the central area (with different widths from 3.5 to 5.5 m and lengthes up to 100 m). For more information about the reader is referred to duraBASt website and (Wacker & Jansen, 2020)

This paper aims to present the APT [1] program phases of this project and some of the results from that.

2. APT PROGRAM AND ITS PHASES

Accelerated pavement testing (APT) is an approach for simulating the truck loads in a short and compressed time period (Steyn, 2012). BASt's APT program has been defined and standardized based on the gained experiences over the years of performing that in house (Wacker & Jansen, 2018).

The APT program of this project can be divided into four different steps:

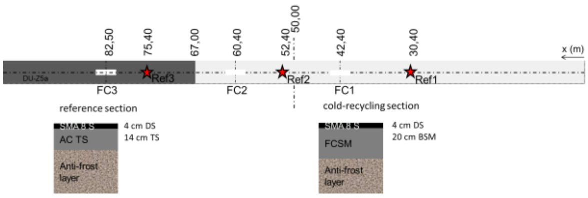
1. Designing activities: test lane layout, mix design of the materials, structural design of the test sections, justifying of the standard loading and monitoring plans.
2. Construction activites: preparing the test lane, material production, sensors installation, laying and control tests.
3. Loading and monitoring: preparations, loading and monitorings during and after the loading.
4. Analysis and evaluation of the data which, in some cases is parallel with loading.

Each of the activities will be explained in the coming texts.

2.1. Test lane layout

For this project a test lane with 100 m length and 3.5 m width was selected form the D/U part of the duraBASt. It was decided to construct a conventional pavement section type beside the cold recycled pavement, as a reference. Two third of the test lane was planned for the cold recycled pavement and the rest one third for the conventional pavement.

The section with the foamed bitumen cold recycled base layer will be referred as cold recycled (CR) section and the section with the hot mix asphalt base will be referred as the reference section (RF) in the rest of this paper. For the CR section, two loading areas (one as reserve) and for the RF section, one loading area were planned. Each loading area has a reference point far enough out of that, which is instrumented with embedded sensors and measured with nondestructive methods each time when a measurement is performed on the loading areas. The aim of these reference points is to be able to check the difference between the loaded and not loaded pavement and also to control if the devices are measuring correct.



Source: BASt, Project data

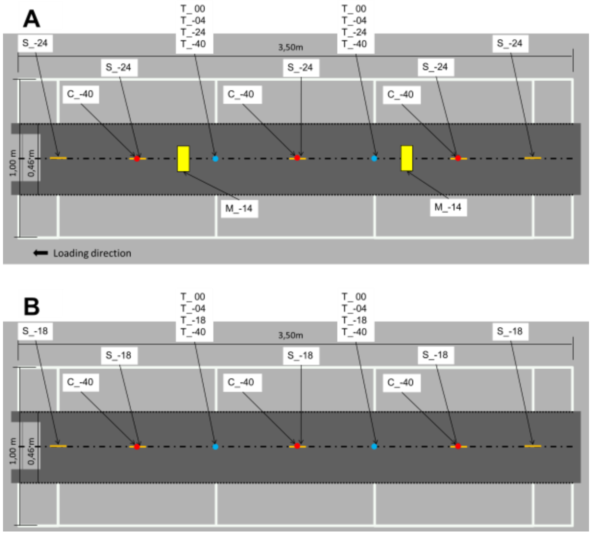
Figure 1: Sketch of the test lane, its pavement sections and different loading areas

Figure 1 shows the sketch of the test lane with different sections, its loading areas and reference points. The loading areas in the cold recycled section are named as FC1 and FC2 with Ref.1 and Ref.2 as the reference points for each. The loading area of the

[1] Accelerated Pavement Testing

reference section is named as FC3 with its reference point named as Ref.3.

Different types of sensors including pressure cells, strain gauges, temperature and moisture sensors were installed at different depths of the pavement in loading areas and reference points of different sections. Figure 2, shows the type and position of different sensors in the loading areas.



Source: BASt, Project data

Figure 2: Schematic picture of loading areas and different sensors installed at different depth of tham. A: for FC1 and FC2 and B: for the FC3

The letters show the sensor's type and the numbers shows its depth in cm. T stands for temperature sensors, S stands for strain gauges, C stands for pressure cells and M stands for moisture sensors.

2.2. Materials and thicknesses of the pavements

The reference section was selected from the existing national design code for pavements in Germany RStO 12 (FGSV, RStO 12, 2012) for 1 million ESAL (of 10 ton standard axle) bearing capacity. It consists of 4 cm of wearing coarse and 14 cm of hot mix asphalt base coarse. Stone mastic asphalt with nominal aggregate size of 8 mm was selected for the wearing coarse (SMA 8 S) and asphalt concrete with nominal aggregate size of 22 mm was selected for the base coarse (AC 22 T N 50/70 BAS RA) which contains around 30% of RAP.

For the mix design of the foamed bitumen mixture, the proposed method for foamed bitumen stabilized mixes (BSMs) from Wirtgen was applied (Wirtgen, 2017). Nyfoam80® from Nynas company was used to produce the foamed bitumen. Foaming tests were performed at different bitumen temperatures and foaming water contents with the WLB 10 S foaming equipment from Wirtgen. The bitumen temperature of 160° C with 2% foaming water satisfied the desired foaming parameters (expantion ratio of 13 times and half life of 14 seconds). To increase the moisture resistance of the mix, cement (I-425 N) was seclected as the secondary binding agent.

The reclaimed asphalt pavement (RAP) was selected and stockpiled in an asphalt plant. During the preprocessing the RAP was sieved in to two fractions of 0-8 mm and 8-22 mm sizes. The activity tests (Wirtgen, 2017) showed that the bitumen in the RAP is still active. To nutralise its effect on the resulted mix, addition of crushed aggregates or crushed dust is recommended (Wirtgen, 2012). The tests with the addition of 15% of crushed sand (0-5 mm) or crushed granular (0-20 mm) led to satisfactory results in case of neutralizing the bitumen activity of the RAP. The grading curve results also showed lack of fines which is necessary for foamed bitumen mixes. Different mix combinations of the two RAP fractions and 0-2 mm sand were produced with different amounts of foamed bitumen and 1% cement.

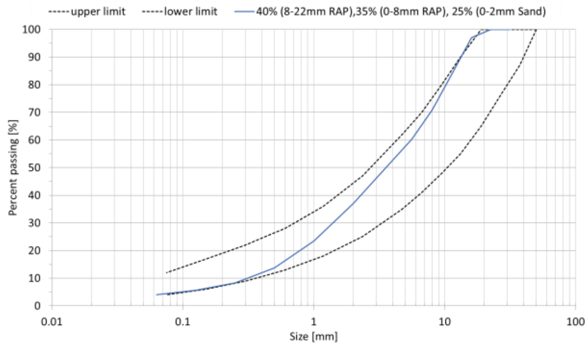


Figure 3: The grading curve of parent material mix of the selected mix design. Upper and lower limits are from Wirtgen (Wirtgen, 2012)

Based on the indirect tensile strength tests results (ITS wet and dry), the final mix combination of 75% RAP (40% 8-22 mm and 30% 0-8 mm) and 25% sand (0-2 mm) was selected as the parent material (Figure 3) with 2.2% foamed bitumen, 1% cement and 4.5% water. The amounts are based on dry weight of the parent material mix.

Based on the existing national guidlines for cold recycled materials (FGSV, 2005; FGSV, 2007), for the same bearing capacity as the reference section, the design should consist of 4 cm of wearing coarse, 6 cm of hot mix asphaltic base and 18 cm of cold recycled layer. According to international literature and structural designs from the author (Kalantari, 2020), it was figured out that the recommended design is kind of conservative for the bearing capacity of 1 million ESAL (of 10 ton standard axle) and therefore, it was decided to adjust it to 4 cm of wearing course and 20 cm of cold recycled layer with foamed bitumen (CRF) as the base layer.

2.3. Production of the recycled mix

Recycled mix was produced in the asphalt plant where the RAP was already stockpiled for this project at the time of mix design tests. To produce the mix, a Wirtgen KMA220 mobile plant was utilized. On day before the production, the 3 fractions of parent material were dry mixed together by the KMA220 based on the mix design formula. At the production day, the prepared mix was fed in to KMA220 and the recycled mix was produced with 2.2% foamed bitumen and 1% cement (Figure 4). The amount of additional water was adjusted based on the measured in situ moisture of the mixed fractions. The produced mix was loaded directly on trucks and transported to the construction site.

Source: BASt, Project data

Figure 4: Production of the cold recycled mix with mobile plant (KMA 220)

2.4. Construction of the test sections

Construction was started with preparing the existing lane by milling the asphalt and excavation up to -40 cm, compacting, installing the pressure cells, then laying the granular anti frost layer back and compacting to reach the desired depth level for each of the sections. Plate load tests were perfomed at reference points and also at some points in the loading area for quality control.

The construction started from the cold recycled section (11.09.2019) by laying the recycled layer with asphalt paver and compact it with steel drum and rubber wheel rollers. Before laying the layer, the strain gauges were installed at planned positions. Moisture sensors were installed in the fresh laid layer before the compaction.

Source: BASt, Project data

Figure 5: Strain gauges installation, laying and compaction of the cold recycled layer

To control the construction quality, fresh samples of recycled mix were taken from both sides of the test section during the laying process at different positions (30, 40, 50 and 60 m) and were compacted in the laboratory at the same day to produce specimens for ITS tests. After curing (72 hours in 40° C) they were tested in dry (cured state) and wet (24 hours in 25° C) states at 25° C. Beside the ITS controls, the Medium Falling Weight Deflectometer (MFWD) was applied on middle points of the 2 loading areas (FC1 and FC2) and their reference points (Ref.1 and Ref.2) from the day of construction and up to 2 days after that to monitor the evolution of the stiffness at early life of the material. More information on the device and its specifications can be found in (Jansen, 2013). Figure 6 shows the increase of the Ev_{dyn} over time at different measured points. It can be seen that the stiffness development on FC1 point (at 42.40 m) is different from the others.

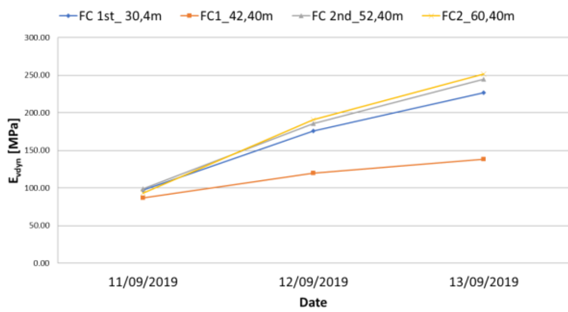


Figure 6: The results of the MFWD measurements

Table 1 shows the ITS and the TSR (Tensile Strength Ratio) for the positions at which the samples were taken (each one is the average of 3 replicates).

Table 1: The results of the ITS tests on fresh samples compacted and tested in laboratory

Parameters	Positions of the samples [m]			
	30	40	50	60
ITS_{dry} [kPa]	339	224	252	341
ITS_{wet} [kPa]	303	133	207	276
TSR [%]	89	59	82	81

It can be seen that the results of the 40 m point are weaker than the other points which confirms the results from the MFWD measurements. A relatively lower TSR and lower rate of early stiffness gain, can be the evidence of different amount of cement in this section. It can be also the result of a sudden change in aggregate gradation (kind of segregation). These observations opened the topic of homogeneity in production and construction.

To investigate the extend of this issue and the effect of curing on that, extra FWD measurements were planned which are explained later under the monitoring plan.

The in-situ density was determined in 10 m intervals with the balloon method. Table 2 shows the positions and the amounts. The dry density in mix design stage was 2.2 gr/cm^3.

Table 2: The results of in-situ density tests (in gr/cm^3)

Positions [m]	20	30	40	50	61
Density	2.26	2.24	2.25	2.28	2.21

Source: BASt, Project data

Figure 7: Controlling the in-situ density of the recycled layer

The asphaltic base layer for the reference section was laid a week later (17.09.2019). A polymer modified fluxed bitumen was sprayed on the surface of the cold recycled layer at the same day (Viaflex® C60 BP-4-S). The whole surface of the test lane was then overlaid with the wearing course the day after. During the laying of the asphaltic layers, fresh material was samples for the laboratory mix controls.

2.5. Loading of the test sections

BASt uses a Mobile Load Simulator (MLS30) since 2013 for its indoor (test halls) and outdoor (duraBASt) APT programs. MLS30 is a device which has been designed to simulate the wheel passes of the heavy trucks in a compressed time duration.

Source: BASt, Project data

Figure 8: Mobile load simulator of BASt (MLS30)

It has four test bogies with one wheel on each. They rotate in aclosed frame on guid rails in a vertical-oval direction. Every time one of them rolles on a linear strip over the loading area. It is possible to apply variable wheel loads between 45 to 75 kN and also different types of wheels (single, super single or dual). Variable loading speeds of 6.5 up to 22 km/hr (6.1 m/s) are achievable. At its maximum speed, it is possible to get around 6000 passes per hour which, means every 600 ms one pass on the same position (Wacker, Scherkenbach, & Jansen, 2018). It has the ability of 50 cm lateral load wander on each side. For loading of each area, the MLS30 is positioned on the desired place and will rest on its 4 side colums (one each side).

MLS30 setup for this project was to use the super single tyre with 50 kN load and the speed of 6000 passes per hour. Based on the gained experiences from the last projects, it was planned to load each section to 3 million passes. To minimize the effect of weather changes, it was decided to load both sections (reference and cold recycling loading) together by changing the loading in weekly periods between them. From the cold recycling section, the FC2 loading area was selected for this purpose. The loading started on Feb. 2020 and till the end of Nov. 2020, each loading area was loaded to 2.2 million passes. The loading started again in Feb. 2021 and will continue till reaching to 3 million passes for each. It is also planned to load the FC1 area (the 2nd one from the cold recycling section) after that up to 1 million passes too.

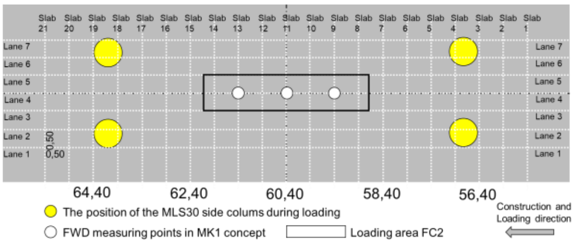
2.6. Investigation plan of the project

Investigation plan of the project is consisted of two main parts: field monitoring and tests plus laboratory tests on the extracted cores. Field monitoring plan consisted of bearing capacity measurements with Falling Weight Deflectometer (FWD), subsurface scan with a 3D Ground Penetration Radar (GPR), transverse profile measurements along the loading area and monitoring with the installed sensors.

FWD measurements

The standard bearing capacity measurement plan aims to assess the bearing capacity of the pavement and to monitor if it changes with the evolution of loading. It consists of 3 main concepts. For each measuring point, 3 drops of 50 kN are applied. All the 3 reference points (named as Ref. 1 to Ref. 3) are measured in all of the FWD measurements. In concept 1 (named as MK1), measurements are performed on three points along the center line of the loading strip in each loading area. Concept 2 (named as MK2) contains the same points from the first concept plus 2 more points each in left and right sides of the centerline points. These points are on the line parallel to the center line and with 25 cm apart to each other.

In measuring concept 3 (named as grid measurement), a grid of points are measured around the loading area. Grid lines will have a distance of 50 cm from each other and cover an area of 5 m from the center of the loading strip to each side (Figure 9). Each concept has different measuring intervals based on the load cycles.



Source: BASt, Project data

Figure 9: FWD grid measuring concept points of the FC2 loading area

Beside the standard FWD plan, two extra concepts were defined for this project one after the construction and the other one after the progress of loading. The first one is to measure the whole test lane on its center line at 50 cm intervals (named as whole lane concept). The aim was to assess the construction homogeneity considering the ITS and MFWD results after construction. After the first measurement round, it was decided to continue this concept at different time intervals to monitor the process of stiffness gaining of the recycled material over the time through the whole test lane. The second additional concept aimed to look closer to the effect of temperature changes on FWD measurements over the cold recycled section. As the cold recycled material has lower bitumen content plus cement as the secondary binding agent, therefore its response to the load is less temperature dependent than the asphalt section. In this concept (named as FWD-Temp.), different points were selected through the centerline of the CR section and in different days FWD measurements were performed on them through the day at different temperatures (of the pavement) to be able to assess its effect.

All the 5 above mentioned measuring concepts were applied for this project.

Transverse profile measurements (evenness)

A profilometer is used for this measurement which has a small wheel rolls over the surface. The measurement is performed along the loading area at 5 different cross lines (Figure 10).

Source: BASt, Project data

Figure 10: Measuring the transverce profile on a cross lane in FC2 (cold recycled section) loading area

Three measurements are done at each cross line before the start of the loading to define the average zero measurement profile. Measurements are performed at different loading intervals starting from 50,000 intervals and later to 200,000 loading intervals after the first 200,000 cycles. The measured data are used to calculate the rutting depth and its change over the loading cycles.

The rutting depth is determined based on the difference between the minimum (Min.) and maximum (Max.) amounts calculated for each profile measurments. The Min. amount is determined by averaging the loading strip results over a cord length and the Max. amount is the average of two side points. Figure 11 shows the method concept. The cord length is 20 cm.

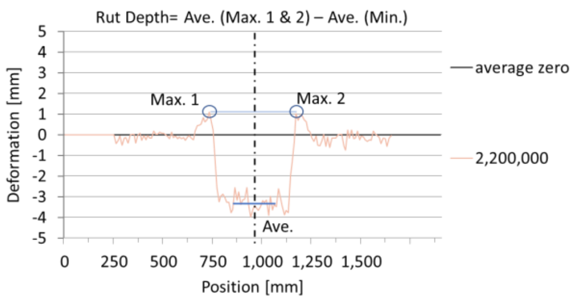


Figure 11: Rut depth calculation method

Ground Penetration Radar measurements

This type of measurement is to get a better view of the thickness of different constructed layers and to be able to evaluate the construction homogeneity. Knowing the as built thickness of the layers is a crucial parameter for interpreting the observed performance of the pavement sections. It is also needed for computational models of the pavement sections to calculate the remaining life based on the material as built parameters determined from the tests on the extracted cores.

Sensors monitoring

As described before in the layout section, different types of sensors were planned for this project and were installed at different depth of the tests sections during the construction process. Strain gauges and pressure cells capture data every 15 minutes for a duration of 30 seconds in each loading period but the temperature and moisture sensors gather continuously every 15 minutes.

Laboratory research plan and tests

A comprehensive laboratory investigation was planned parallel to the field monitoring. It has two main goals: to monitor the evolution of stiffness/strength in the cold recycled material over the time (curing process) and to assess material's main characteristics and performance parameters in both test sections. A coring plan has been prepared for this purpose to extract cores at different intervals after the construction and to perform stiffness, indirect tensile strength and fatigue tests on them. Coring is performed to obtain 150 mm diameter samples.

The cores taken from the reference section, will be cut to separate the wearing and the base coarses. The specimens will be used for cyclic indirect tensile stiffness tests at different temperatures and frequencies according to German method (named TP Asphalt-StB Teil 26-2018) which is very similar to EN12697-26: 2012-06 (annex F). The results will be used to obtain the stiffness master curve of these two types. The asphaltic base coarse specimens will later be used for indirect tensile fatigue tests at 20° C and 10 Hz (also based on the German method (named TP Asphalt-StB Teil 24-2018) which is similar to EN 12697-24: 2012-08 (annex E). The results are used to determine their fatigue properties.

The cores taken from the cold recycled section aiming to monitor the curing, are cut to get specimens with 60 mm height. The loading and the sensor places around the speciemens are evened with gypth and then they are conditioned in the oven at 40° C for 72 hours to accelerate the loss of their moisture because of coring and cutting. After that they are tested under cyclic indirect tensile test (at 20° C and 10 Hz) at different stress levels (up to 0.1‰ horizontal strain level) to obtain the stiffness modulus at different horizontal strain levels. This test is named as multi-step stiffness test; the method and the way to analyze the results developed by the author during his PhD (Kalantari, 2020). It captures the stress/strain dependency of the stiffness which is one of the characteristics of the foamed bitumen recycled/stabilized mixes beside their temperature dependency of the stiffness. Minimum of four specimens are used for this test. Based on the results it is possible to determine the stiffness at any desired strain level and probability level. The samples are then used for ITS tests at 25° C as the stiffness tests in the mentioned strain range doesn't affect the ITS amount (Kalantari, 2020). These cores are planned to be extracted at different time intervals after the construction and it was tried to perform a round of the whole lane FWD measurement simuletaneously.

The second cores are planned for about a year (or a little more) after the construction (considering that the curing is mostly completed). The aim is to obtain the stiffness master curve and the fatigue behavior of the recycled mix. The tests are the same as the ones for the asphaltic base coarse samples. The only different is that to be sure the specimens are not damaged, the stiffness tests are performed at lower horizontal strain levels than the normal amounts for HMA (around 0.03 to 0.04‰). In his PhD, the author has developed a method to produce the general stiffness model of the material by combininig the results of the multi-step stiffness tests and the stiffness master curve tests. This general model captures both the strain dependency of the stiffness (which comes from its granular part behavior) and also its temperature dependency (which comes from the asphaltic part behavior).

Considering the loading areas and the requirements of from FWD measurement concepts and the construction homogenity of the test sections, the position of the cores were planned. Extra cores will be extracted from the loading areas (in both test sections) after the end of APT loading with MLS30 to investigate the possible changes in the material parameters.

3. RESULTS AND DISCUSSIONS

As the program is still undergoing all the results are not analyzed and assessed for reporting at this moment therefore only some of the results from monitoring and the laboratory plans are presented here in this section. The MK concepts of the FWD measurements and the results of the second core series are not reported here.

3.1. Transverse profile measurements results

The amount of rutting was determined for each loading area at different loading cycles based on the results of transverse profile measurments. Figure 12 and Figure 13 show the total rutting amounts at five different measurement lanes along the loading areas of FC2 (CR section) and FC3 (RF section).

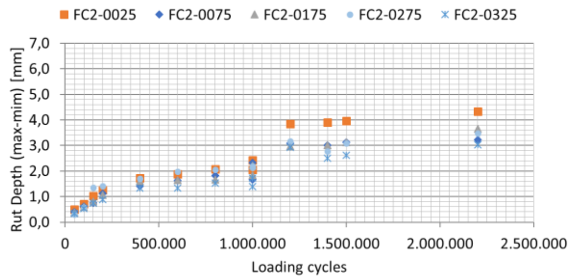


Figure 12: The rutting amount versus loading cycles for CR section

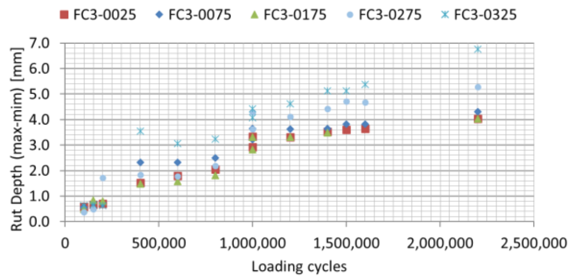


Figure 13: The rutting amount versus loading cycles for RF section

Comparing the results of the two construction types together shows a slightly higher amount of total rutting in FC3 section with HMA base layer. Both sections have no cracks sofar. The maximum amount of rutting in the cold recycled section (FC3) is 4.4 mm after 2.2 million cycles of loading. Considering the allowable design limit of 10 mm (5% of the layer thickness), the section behavior is assessed as very good. It can be concluded that the performance of the recycled section against the rutting, is equal to the reference section.

For a better comparison between the two construction types and also to be able to define a rutting model, it is necessary to consider its variations between the two sections during the loading periods, As the temperature is an important factor which affects the rutting rate of the bituminous layers. To get a closer look to this issue the data from the temperature sensors have been used.

Table 3 and Table 4 show the surface temperatures up to 1 million cycles of loading for each of the loading areas. The sensors were successful to record the surface temperature for about 75% of the 1 million loading cycles. The temperature in each cumulative loading range (loading ranges in the first colum of the tables) was calculated based on a weighted averaging on the temperatures during each of the loading ranges.

Table 3: Average weighted temperature (AWT) on the surface of FC2 loading area during the first 1 million loading cycles

Cold recycled section (FC2)		
loading range [x1000 rolling]	% of the total loading	AWT [°C]
All	74.8	21.7
0-200	20.0	20.1
200-253	5.3	27.5
355-600	24.5	17.8
750-800	5.0	24.8
800-1000	20.0	25.9

Table 4: Average weighted temperature (AWT) on the surface of FC3 loading area during the first 1 million loading cycles

Reference section (FC3)		
loading range [x1000 rolling]	% of the total loading	AWT [°C]
All	75.0	23.5
0-200	20.0	16.1
200-400	20.0	22.7
650-800	15.0	25.0
800-1000	20.0	30.5

Looking to the tables shows clearly that each jump in the rutting amounts of the reference section (FC3) is related to a temperature increase. It was decided to plan the future loadings on the section by considering the temperature forecast of the coming week. The data of the temperature sensors will be analysed for the whole loading cycles.

3.2. FWD measurements results

Here the results of three different concepts of the FWD measurements are presented.

Whole lane measurements results

The first measurement proved the findings from the quality control tests. Figure 14 shows the SCI300 parameter (it is the difference between the surface deformations measured at the center of the loading plate and 300 mm apart from that) in the centerline of the whole test lane at different measurement intervals (50 and 490 days after construction).

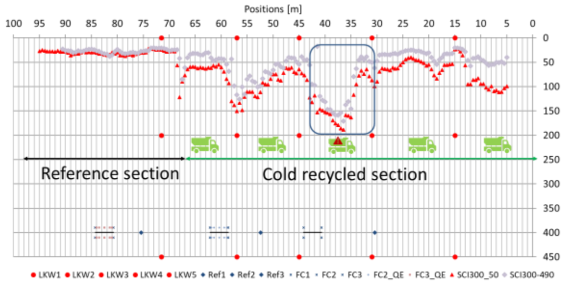


Figure 14: The results of the whole lane FWD measurements (SCI300 parameter)

Looking to the cold recycled section, by comparing the results with the distance covered with each of the mixture transport tracks, shows which truck had the suspicious mix. It is also possible to see that with the changes of the hauling truck, a change has happened in the pavement too which can be because of that the operator closed the paver hoper and the mix gradation was changed because of the segregation. The findings opened the question of the effect of laying method on the homogeneity of the pavements with CR layer. Looking to the reference section results, it can be seen that the construction is more homogenous.

The results also were used to plan the position of the cores. The measurements were continued at different intervals to monitor the changes of the material over the time. Figure 15 shows the SCI300$_{@\ 20°\ C}$ parameter from different intervals of whole lane FWD measurements on a section of the cold recycled section, where the cores are extracted to check the stiffness evolution over the time. The cores were extracted between 27 to 29 m position. It can be seen that the SCI300$_{@\ 20°\ C}$ parameter decreases over the time which is a result of the increase of the stiffness in cold recycled layer.

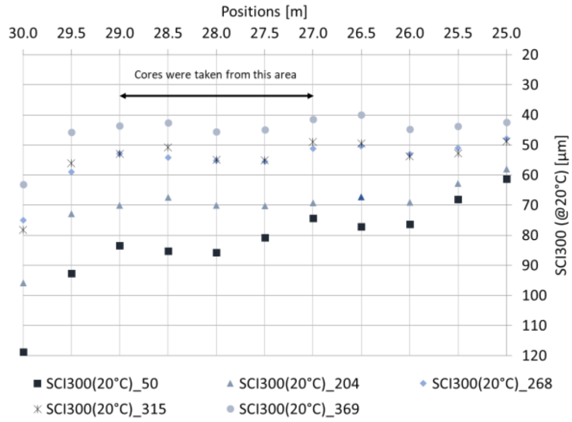


Figure 15: The evolution of the bearing capacity (SCI300@ 20° C) over the time in cold recycled material (the numbers in legends, show the days after construction)

FWD-Temp. concept results

The aim of this concept was to assess the effect of temperature changes on the FWD results. 6 different points (containing the weakest, strongest, 2 reference points and the middle of each loading area) were selected and were measured at different days. One of the analysis was to look at the changes of the SCI300 at different temperatures during each daily measurement. After the calculation of SCI300 for each measuring point and each temperature, a linear equation was fitted to the results of each point and used to determine the SCI300 at 20° C for that point at that measurement day (which can be related to days after construction).

Putting all the calculated SCI300$_{@\ 20°\ C}$ together in a time order after construction, showed that even by having the same temperature for them, their values change over the life of the pavement. To investigate that, the weather data (temperature and precipitation) were combined with the results (Figure 16).

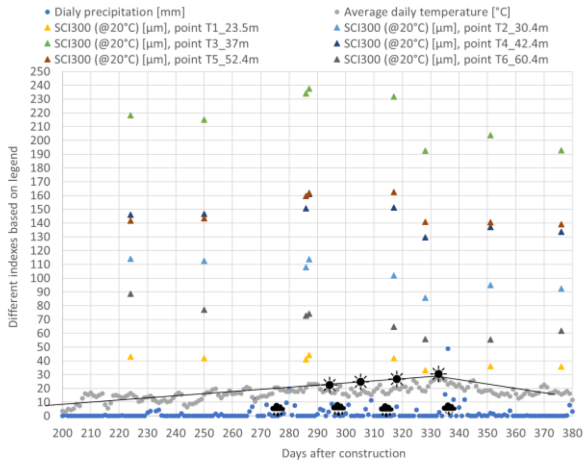


Figure 16: Changes of the SCI300@ 20° C over the time at different points in combination with weather parameters

The results showed that the response of the material is not only temperature dependent but also depending to the wet-dry cycles (or seasonal variations) too. It is then crucial to consider beside the temperature corrections of the FWD measurements, this effect in to account too.

Grid measurement cocepet results

As mentioned before this concept is used to get a more detailed look on the changes in the sections because of the loading cycles. Till now two sets of measurements were performed on each of the loading areas. First round before the start of the loading and the second round after 1.5 million cycles of loading on each. Figure 17, shows the difference of d_0 (temperature corrected) between the two measurements for the FC3 loading area and around that.

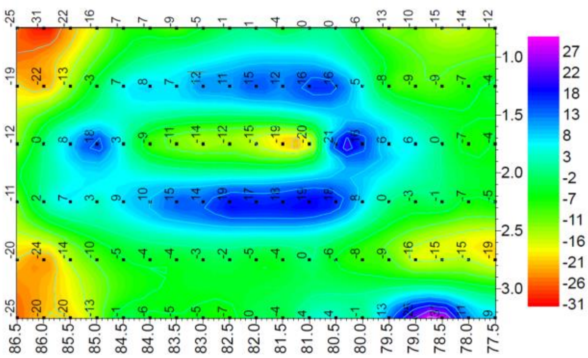


Figure 17: The difference of the parameter d_0 (temp. corrected) [µm], between 0 and after 1.5 million cycles for FC3

It is clear that the loading strip has got some damage as the difference is with negative values. The side amounts with positive values show the decrease of the d_0 because of the loading. This is because of the side effect of loading on consolidation of the granular anti-frost layer and also the subgrade which, leads to deacrease in the whole deformation mould response. The upper and lower left sides show the damaging effect of the heavy corner colums of the MLS30 but as can be seen they didn't affected the loading strip.

The same calculation was done for the results on FC2 (cold recycling) area. Figure 18, shows the results.

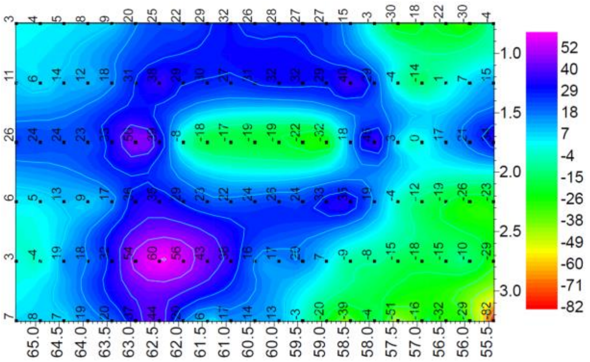


Figure 18: The difference of the parameter d_0 (temp. corrected) [µm], between 0 and after 1.5 million cycles for FC2

The same as the other section, the loading strip shows damage and the around points show the decrease in the d_0 which in this case is a combination of not only the consolidation of the granular and subgrade but also the cold recycled layer too. As the material is semibound it has partly responses like granular material. The second point is the increase of its stiffness because of curing over the time which is more obvious on not loaded areas. Looking to the sides it is possible to see the effect of the corner side columns of the MLS30. As the section was not homogenous before the loading, the resulted damage from them is also different in each side but they didn't affect the loading strip.

3.3. Laboratory tests results

In this section only the results of the stiffness tests on the cold recycled specimens are presented as the other results are not completely analyzed at the time of writing this paper. As explained in the laboratory plan, the cores were taken at different intervals after the construction from a section which had a homogen layer of cold recycled material. Figure 19Figure shows an example of the cores and the prepared specimens from them for the tests. Normally from each core, two specimens from upper and middle part of the recycled layer are prepared for the tests.

Source: BASt, Project data

Figure 19: An example of the sawed cores and the prepared specimens from them for the tests

Figure 20 shows the evolution of the stiffness over the time which was calculated from the results on upper part specimens. They are at 20° C and 10 Hz and determined for 0.05‰ horizontal strain level with 95% level of confidancy. Looking to the trend, clearly shows the curing process of the material over the time which continues even after one year from construction. The trend is also in agreement with the FWD results (Fgure 15).

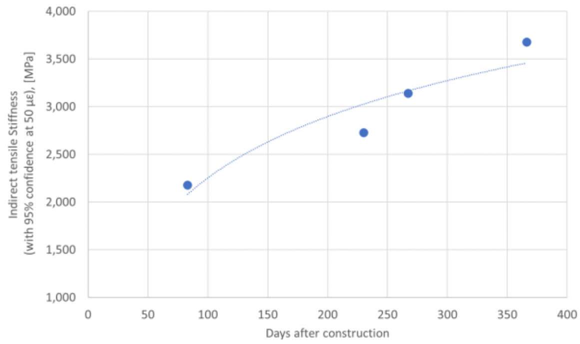


Figure 20: The stiffness of the cold recycled material from the cores at different days after construction

4. SUMMARY AND CONCLUSIONS

Cold recycling is a method which got a high amount of interest during the recent years in Germany. It has the benefits of lower energy consumption and up to 100% use of the high amount of stockpiled RAP in asphalt plants. Considering the positive international reports on the behavior of cold recycled mixes with low amounts of foamed bitumen and cement (classified as BSM) and relatively low amounts of experience and performance data on these type of material, encouraged BASt to define a research project in cooperation with Wirtgen GmbH. The project applies the APT concept to evaluate the performance of a pavement type containing foamed bitumen recycled layer. The project started end of 2018 and continued with the preparation tests and activities till the construction of the test lane in Sep. 2019. The loading with MLS30 started in Feb. 2020 and is still ongoing. This paper aimed to present the APT program of this project, its different phases from planning to execution, monitoring, analysis methods and a part of the results.

Different points and lessons learned from this research in fields of preparation and production of the mix, construction and quality control, field monitoring and nondestructive tests and laboratory tests to assess the material behavior. Some of them can be mentioned as below items:

1- Parent material characteristics have a great impact on behavior and performance of the resulted recycled mixture. It is recommended to crush the RAP to get a better size distribution and internal angle of friction in the compacted aggregate skeleton. This also increase the possible rate of RAP usage in the mix as the need for virgin aggregate to adjust the gradation curve will decrease. It is also important to consider the level of the activity of the old bitumen in the RAP especially when different sources of RAP are mixed together in a stockpile.

2- In case of the specimen compaction in laboratory, using the vibratory hammer showed good experiences in case of ease of compaction and the production rate of the specimens. It is still needed to gain more experience with the method. Using the optimum moisture contents (OMC) amounts from modified proctor compaction tests as a base for this compaction method seems not be appropriate. It is recommended to use the vibratory hammer compaction also for the determination of the OMC too.

3- In case of using the mobile plant machine to produce the cold recycled mix, the experience with this project showed that the method needs a very short setup time (half a day) and a small compact working place. This advantage can be considered in topics like resiliency in pavement construction. It has the possibility of mixing different fractions/types of the parent materials together and of more control on the production process. Beside the possibility of higher production quality in case of utilizing mobile plants, it is important to notice that the construction and laying of the mix should be performed with care to avoid segregation and reach a homogenous layer.

4- The experience in this project showed that Medium Falling Weight Deflectometer (MFWD) tests are fast and reliable method to check the daily work, to evaluate the construction quality and the evolution of material strength at its very early stages of the life before having the construction tests results. Still more data should be collected to be able to define acceptance thresholds for the measurement results.

5- The results of FWD-Temp. concept measurements proved the lower temperature dependency of the cold recycled material compared to HMA. Beside that they showed that the material's field response is affected by the seasonal weather changes (wet-dry cycles) which, should be considered in correction method for FWD measurements.

6- Monitoring the development of rutting with the load cycles, showed a very good performance of the cold recycled pavement section with a considerably lower than expected rut amounts. The results may suggest the need of revising the existing national design guide for the cold recycled pavements. Comparing the results from the reference section, shows the importance of temperature on the rate of permanent deformation in pavements with thick HMA packages.

7- The results of stiffness tests on the specimens from the extracted cores at different lifes of the recycled layer showed the increase of the stiffness over time even after one year of construction. They agree with the different FWD results together.

The program will continue with the loading of the test areas of each sections till 3 million cycles with the monitoring activities. The stiffness and fatigue tests results on the specimens from the extracted cores will be applied in to a mechanistic-empirical software to assess the life of the section with their field behavior. The next phase of the research will deal more in detail with the failure mechanism of this group of material and to define failure transfer function for structural design purposes.

Acknowledgments

The author emphasizes his great appreciation to the big project team which is from different sections of BASt institute. Their high flexibility in adoption of different plans, their special care on variable tasks of this complicated project and their trust was the key to the success.

REFERENCES

[1] FGSV. (2005). Merkblat für Kaltrecycling in situ im Straßenoberbau, MKRC. Cologne, Germany: FGSV.

[2] FGSV. (2007). Merkblatt für die Verwendung von pechhaltigen Straßenausbaustoffen und von Asphaltgranulat in bitumengebundenen Tragschichten durch kaltaufbereitung in Mischanlagen, M VB-K. Cologne: FGSV.

[3] FGSV. (2012). Richtlinien für die Standardisierung des Oberbaus von Verkehrsflächen, RStO 12. *2012*. Cologne, Germany: Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV).

[4] Jansen, D. (2013). Dynamic load plate tests: calibration procedure for test equipments in Germany. *Ninth International Conference on the Bearing Capacity of Roads, Railways and Airfields*, (pp. 229-238). Trondheim, Norway.

[5] Kalantari, M. (2020). Assessing the effect of binding agents on resulted mechanical and fatigue characteristics of foamed bitumen and cement stabilized mixes,Draft manuscript for PhD, unpublished. Siegen, Germany: University of Siegen.

[6] Steyn, W. (2012). NCHRP Synthesis 433: Significant Findings from Full-Scale Accelerated Pavement Testing. Washington, D.C.: Transportation Research Board.

[7] Wacker, B., & Jansen, D. (2018). Toolbox for APT program monitoring at BASt and optimized data handling on new test facility duraBASt. *TRB Annual Meeting.*

[8] Wacker, B., & Jansen, D. (2020). Two Years of APT Program on the New Test Site duraBASt. Chabot A., Hornych P., Harvey J., Loria-Salazar L. (eds) Accelerated Pavement Testing to Transport Infrastructure Innovation. Lecture Notes in Civil Engineering, vol 96 (pp. 32-39). Springer, Cham. doi:https://doi.org/10.1007/978-3-030-55236-7_4

[9] Wacker, B., Kalantari, M., & Diekmann, M. (2020). Cold Recycling in Germany—Current Experiences and Future Projects. Raab C. (eds) Proceedings of the 9th International Conference on Maintenance and Rehabilitation of Pavements—Mairepav9. Lecture Notes in Civil Engineering, vol 76 (pp. 813-823). Springer, Cham. doi:https://doi.org/10.1007/978-3-030-48679-2_76

[10] Wacker, B., Scherkenbach, M., & Jansen, D. (2018). *Temperaturuntersuchungen am und im Mobile Load Simulator MLS30.* Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).

[11] Wirtgen. (2012). *Wirtgen Cold Recycling Technology* (1st ed.). Windhagen, Germany: Wirtgen.

[12] Wirtgen. (2017). *BSM Cold Recycling, Laboratory Handbook.* Windhagen, Germany: Wirtgen GmbH.

HAZARD PERCEPTION, TRAFFIC SITUATION OVERVIEW AND REACTION TIME IN DRIVERS WITH EARLY ALZHEIMER'S DISEASE

Kas Kamphuis, Msc
SWOV
Bezuidenhoutseweg 62, The Hague, Netherlands
kas.kamphuis@swov.nl

ABSTRACT

In the Netherlands patients with early Alzheimer's disease (AD) are allowed to drive if their fitness to drive is positively evaluated. The assessment consists of a clinical interview and an on-road examination. Previous research concludes that on-road performance by AD patients could be accurately predicted (92.7 percent) within a clinical setting on the bases of a combination of three assessments: 1) a clinical interview, 2) neuropsychological tests and 3) driving in a driving simulator (Piersma et al., 2016). In the current research at SWOV, the Dutch Institute for Road Safety Research, we examine if this method leads to the same prediction with a new group of AD patients. The neuropsychological tests include a series of reaction time tests, a hazard perception test, the Adaptive Tachistoscopic Traffic Perception Test and a driving simulator test drive. It is hypothesized that early AD patients react slower, have more difficulty overseeing traffic situations and perceiving possible hazards while driving. Results confirm the hypothesis in the sense that, compared to healthy controls, AD patients have a slower cognitive/motoric RT to visual/auditory/combined stimuli, need more time to react to traffic situations, have a harder time getting an overview of traffic situations and brake later in response to a driver that suddenly swerves onto the road in a simulated drive. In future research the reduction of hazard perception ability, the ability to get an overview of traffic situations and cognitive/motoric RT at which an AD driver can still safely participate in traffic and also the effectiveness of coping strategies could be investigated.

Keywords:
Alzheimer's Disease, Traffic safety, Hazard perception, reaction time, Fitness to drive Introduction

1. INTRODUCTION

As people get older age related diseases become more frequent. Dementia is a well-known example of such an age related disease. The most common type of dementia is Alzheimer's Disease (AD) (50 – 75%), followed by vascular dementia (20%), dementia with Lewy bodies (5%) and frontotemporal lobar dementia (5%) [1]. AD has developed from a rare disease at the beginning of the 20th century to the sixth leading cause of death in the world [2,3]. AD is a progressive disease in which multiple cognitive domains are affected. In addition to the memory domain, also the domains of attention, visuospatial abilities, executive functioning, language and praxis are frequently impaired [4].

Alzheimer's disease includes a pre-dementia and dementia phase. The pre-dementia phase is associated with a loss of episodic memory. During this phase patients often forget recent memories or experiences but are not yet impaired in performing their daily activities [5]. The pre-dementia phase is more commonly called mild cognitive impairment (MCI). As the MCI progresses to actual dementia the memory symptoms worsen (now also affecting older memories) and linguistic and spatial orientation deficits become apparent [5]. Reduced ability to plan, judge and organize start to interfere with ability to complete complex tasks [6].

An example of a complex task commonly performed by older adults is driving. Driving is the preferred mode of travel for older adults and is a meaningful instrumental activity of their daily living as it enables them to visit family, friends and make use of services they require [4,7].

As AD affects domains that are highly relevant to driving skill, fitness to drive may be severely impaired [8]. Research shows that demented drivers forget to fasten their seatbelt, have difficulties with the identification of landmarks and traffic signs and get lost more often [8].

The degree to which a patient is impaired by AD is often rated using the clinical dementia rating scale (CDR). The CDR is a 5-point scale that is used to categorize six cognitive and functional domains: memory, orientation, judgment, problem solving and community affairs. The resulting scores range from 0 (no impairment) to 3 (severe impairment).

In the Netherlands drivers with very mild or mild AD (a CDR-score of 0.5 or 1.0) are allowed to drive a car if their fitness to drive is positively evaluated. Fitness to drive is evaluated with a series of neuropsychological tests and an on road evaluation. Previous research concludes that on-road performance by AD patients could be accurately predicted (92.7 percent) within a clinical setting on the bases of a combination of three assessments: 1) a clinical interview, 2) neuropsychological tests and 3) driving in a driving simulator [4].

Currently at SWOV, the Dutch Institute for Road Safety Research, a follow-up study is performed to examine if this method leads to the same prediction with a new group of AD patients. The neuropsychological tests include a series of reaction time tests, a hazard perception test, the Adaptive Tachistoscopic Traffic Perception Test (ATAVT, which gives insight into the ability to get an overview in traffic situations) and a driving simulator test drive in which another driver suddenly swerves onto the road from a parking spot. These four tests are of particular importance to road safety as these indicate how quickly a dangerous traffic situation is recognized and subsequently is acted upon.

The aim of this study was to determine if the reaction time and hazard perception ability of drivers differed between healthy drivers and drivers with early AD. It is hypothesized that early AD patients react slower, have more difficulty getting an overview of traffic situations and have more difficulty perceiving possible hazards while driving.

2. MATERIALS AND METHODS

This section describes the materials and method used in the current research.

2.1. Participants with AD

Patients with AD were assessed at hospitals and nursing homes. Inclusion criteria were that the patient: was 50 years of age or older, had a CDR score of 0.5 (very mild AD) or 1 (mild AD), had not been diagnosed with a different neurological or psychiatric condition or used medication that was detrimental to driving ability and wished to keep driving. If the patient met the inclusion criteria a folder was handed to them by their geriatrician, neurologist, psychiatrist or general practitioner after they were informed about their diagnosis. Alternatively, participants with AD could also find information about the research online and directly contact SWOV to apply for the research. Participation was on a voluntary basis and after assessment positive results could be used in an official relicensing procedure. After license renewal the license lists a code referring to the driver's condition. In case of a negative assessment the research concluded and the participant was strongly advised to stop driving. The participant could then choose to cease driving immediately or to let their driving ability be assessed a second time directly by the Dutch driving test association (CBR). Failing the CBR assessment would lead to a revoked license immediately. Participants also received a monetary compensation for their time and effort.

In total 42 patients with AD were included in the research. Patients were aged 53 to 90 years (mean = 75.4 years; SD = 7.4 years) and 29 (70.7%) of the patients were men. All participants held a valid driver's license for between 35 to 68 years (mean = 54,7 years; SD = 7.4 years). On average participants drove between 10.000 – 20.000 kilometers per year, due to a missing response one participant was excluded from this average (n =41).

2.2. Control-group participants

In total 27 healthy individual participated in the study. The criteria for inclusion were: 65 years of age or older, no diagnose of a different neurological or psychiatric condition that is detrimental to driving ability or usage of medication that has a negative effect on driving ability. The minimal age for inclusion was higher for control participants than for participants with AD to keep the groups comparable. As AD is an age related disease the chance is higher for the control group participants to be younger, this is somewhat balanced by the higher age limit for inclusion. The majority of this group is very healthy and has no age-related driving problems. Participants were recruited through means of a call on the SWOV site and through word of mouth, meaning participants were asked if they knew someone who would also be interested in participating in our research. Participants then contacted SWOV by e-mail or telephone and were sent more detailed information about the research by post. Participation was on a voluntary basis and also the control-participants received a monetary compensation for their time and effort. The healthy participants were aged 65 to 86 years (mean = 71.6 years; SD = 5.4 years) and 20 (74.1%) of the participants were male. All participants held a valid driver's license for between 40 to 62 years (mean = 51.6 years; SD = 4.6 years). On average participants drove between 10.000 – 20.000 kilometers per year.

2.3. Measuring tools

The current research at SWOV is a longitudinal validation study of the research done by Piersma et al. (2016). In both studies the participants take part in a series of pen and paper tasks, followed by a neuropsychological assessment which consists of tasks that measure cognitive functions that are important to driving and lastly four test drives in a driving simulator.

The measuring tools that are relevant to the current paper consist of a driving questionnaire, the clinical dementia rating-scale (CDR-scale), a series of reaction time tasks, the Adaptive Tachistoscopic Traffic Perception Test (ATAVT), the Hazard perception task developed at SWOV and a driving simulator ride. I recommend consulting the paper of Piersma et al. (2016) for an overview of all measuring tools, including those outside the scope of the current paper.

2.4. Driving questionnaire

All participants were requested to fill in a driving questionnaire. The driving questionnaire was adapted from the Safe Driving Behavior Measure [9] and consists of three parts: a demographical profile of the participant (7 questions), a driving profile with questions concerning driver experience, mileage, the use of medication and knowledge of traffic laws (23 questions) and questions relating to the driving skills of the older driver (54 questions). These 54 questions were translated from the Fitness-to-drive-screening measure or FTDS by Classen et al. (2013).

2.5. Clinical Dementia Rating (CDR)

The CDR rating is derived from a semi-structured interview with the patient and their informant, they are questioned separately from one another. The input of informants is of particular importance because they can provide valuable background information about the condition of the patient and their answers also serve as a means to 'fact-check' the answers of the patient. The aim of the interview is to assess the degree to which AD impairs the functioning of the patient in six domains: memory, orientation, judgment, problem solving and community affairs. Examples of questions are: 'Could you tell me what you did last weekend' (memory), 'What day is it today? / Where are we right now?' (orientation) to 'How many 5-cent coins are there in a euro?' (problem solving). Each domain is scored on a 5-point scale: 0 (not impaired), 0.5 (very mildly impaired), 1 (mildy impaired), 2 (moderately impaired) and 3 (severely impaired). Using the scores on the six domains an overall CDR score can be calculated [10]. The memory-domain weighs more heavily compared to the secondary domains. For example if memory is scored as 1 or higher the overall score is automatically 0.5 or higher.

2.6. Vienna Test System (VTS)

The RT-tasks [11] and the ATAVT [12] are administered to the participant using the Vienna Test System [13]. The VTS is a test system used for computerized psychological assignments to measure abilities of the participant that are relevant to road safety. It is a helpful tool to assess fitness to drive in a participant.

2.6.1. Reaction time (RT) tasks Visual, Auditory & Combined.

Three RT-tasks are administered: a visual task, an auditory task and an inhibition task in which visual and auditory cues are combined. In the visual task the participant looks at a black circle and has to press a button the moment the circle changes to yellow. In the auditory task the participant has to press a the moment an auditory cue is heard. In the combined task the participant should only press the button when the black circle turns yellow and at the same time an auditory cue is heard. The button should not be pressed if: 1. the circle turns red 2. the circle turns yellow without an auditory cue or 3. only an auditory cue is heard. Between cues the participant rest his/her finger on a 'rest button'. The cognitive RT is measured from the moment the cue is presented until the moment the index finger moves from the rest button. The motoric RT indicates how fast the participant is able to press the button after the correct stimuli is recognized. It is measured from the time the finger of the participant leaves the rest button till the moment the answer button is pressed. The VTS provides the mean and standard deviation of the cognitive /motoric RT.

2.6.2. ATAVT

The ATAVT (Schuhfried, 2009) is used to measure the ability of the participant to get an overview in traffic situations. The participant is presented with a series of photographs that only stay visible for roughly one second. After each photograph the participant has to choose which elements were present in the shown photograph. The list of options consists of: pedestrians, cars, (motor)cyclists, traffic signs and traffic lights. The ATAVT is an adaptive test, meaning the number of elements shown in the photographs increase or decrease based on the performance of the participant. A psychometric model [14] is used to estimate a general performance parameter, that takes into account the difficulty of the items that participants saw and their performance on them. The parameter indicates the ability of the participant to get an overview in traffic situations given the age of the participant. A score above 0 reflects a better ability, below 0 reflects a worse ability in comparison to the participants age group.

2.7. Hazard perception test

In the hazard perception test participants were presented with 25 photographs of traffic situations. The photographs were taken from the driver's point of view and the current travelling speed was displayed on the mock displayed (see *Figure 1*). Participants had 8 seconds to decide whether they would do nothing, slow down by releasing the gas pedal or brake in the given traffic situation. The participants took the test either with or without time pressure. This decision was made by the person that administered the test to prevent the participant from getting upset/stressed as in some cases the participant was unable to respond within the time limit. The number of correct answers and reaction times were measured. Separate statistical outcomes are presented in the results section for participants in the timed condition and all participants together (timed/non-timed).

Figure 1: Example of a trial used in the Hazard Perception test.

2.8. Driving simulator rides

At SWOV a fixed base driving simulator was used to assess various aspects of driving behavior. The simulator consisted of an open cabin mock up with a steering wheel, gear box, gas pedal, brake pedal, clutch and simulated driving sound. Three screens provided the participant with a 200⁰ view on the road and surroundings. The dashboard and rear view mirror were visualized on the middle screen, the car windows and side mirrors were visualized on the screens to the left and right of the participant. The graphical interface was designed and scenario's were programmed in previous research [4].

After familiarizing themselves with the driving simulator in a practice drive the participants took part in a lane tracking ride, an intersection ride and a merging ride. In this paper we will only consider an event that takes place during the intersection ride, see [4] for a detailed description. Near the end of the intersections ride the participant is confronted with a car turning onto the road from a parking bay. It was programmed in such a way that the participant had to brake strongly to avoid collision. The outcome measure is whether or not the participant braked for the car.

2.9. On Road assessment

Participants took part in an on-road assessment during daylight hours and in their own car. The assessment was carried out by approved experts from the CBR. These experts are trained to evaluate whether an impaired driver is still fit to drive. The expert did not know if the participant was part of the healthy or the AD group. The expert makes use of a Test Ride Investigating Practical Fitness to drive (TRIP) form consisting of 60 items to evaluate fitness to drive in the participant. Each item is scored sufficient, doubtful or insufficient. The 60 items concern 11 subject: place on the road, following distance, speed, looking behavior, traffic signs, taking over and passing, anticipation, communication with other road users/pedestrians, judgement of traffic situations, operation of the vehicle and overall impression. The expert also gives an overall judgement of pass, doubtful or fail.

3. PROCEDURE

SWOV collaborated with several hospitals to recruit participants with AD, patients took part on a voluntary basis. After the patient was informed about their diagnosis and the possible consequences for their driving ability, the research at SWOV was discussed. Patients that met the inclusion criteria and wished to keep driving received a folder by their healthcare professional detailing the specifics of the research. The patient or informant subsequently contacted SWOV to apply for the research. In other cases the patient gave their healthcare professional permission to share their contact information with SWOV. The patient was then contacted by a SWOV researcher shortly after to provide more details about the research and schedule an appointment. In other cases participants or their family looked up information on the SWOV website and contacted the research team directly. Healthy participants were recruited from the general community through means of a call on the SWOV website detailing the specifics of the research and through word of mouth. All participants provided their written consent to take part in the research.

As in the original study by Piersma et al. (2016) the assessment of each participant was divided into two parts. In the first part the participant is invited to SWOV to conduct clinical interviews, a neuropsychological assessment and driving simulator rides. The second part consists of an on road assessment. The day of testing and the on road assessment make up one cycle of the research. Each participant goes through a maximum of three cycles. The results on the day of testing and the on road assessment together lead to an evaluation. The on road assessment weighs more heavily in this evaluation compared to the (neuro)psychological assessment. If participants are positively evaluated the research continues and they are again invited after a few months to start the subsequent cycle. A participant is only allowed to continue the research if the on road assessment is positively evaluated. All participants received a monetary compensation for participating in the study. This compensation was either 25, 50 or 100 euro depending on the number of cycles the participant took part in during the study. Patients with AD that were positively assessed after a cycle could use this outcome in an official relicensing procedure.

4. STATISTICAL ANALYSIS AND RESULTS

Analysis were performed using IBM SPSS Statistics 25. Patients with AD and healthy participants were compared using independent sample t-tests for the ATAVT and RT-tasks. Significance level alpha was set at .05. In case the data was non non-normally distributed a Mann-Whitney U test was used instead of the independent sample t-test. Furthermore, a Chi-square test was used to analyze whether the groups differed on braking during the event in the simulator drive.

4.1. Missing values and outliers

For two tests results were not distributed normally. In both cases this was related to outliers due to participants who were not able to understand or execute the task, which were removed from further analysis. This was done to prevent putting the AD participants under unnecessary stress.

After removal of these outliers the data was normally distributed. There was a similar situation with regards to the ATAVT test analysis. One participant could not complete the test and was therefore excluded from the analysis. Furthermore, several participants were excluded from the simulator rides analysis due to simulator sickness. Lastly, in the brake RT analysis participants that failed to brake were excluded from that analysis.

4.2. Results

In this section results from the different measuring tools are presented for drivers with AB and healthy controls. In Table 1 the groups are compared according to general characteristics. In Table 2 the groups are compared according to outcome measures.

Table 1: Comparison of patients with AD and healthy participants according to general characteristics.

Characteristics	AD (n = 42)	Healthy (n = 27)
Age, mean	75 (7.3)	71.1 (5.4)
Male sex, No. (%)	29 (69%)	20 (74.1%)
Education, mean of 7 stages (SD)	3.38 (1.1)	4.41 (.9)
CDR-score, No. (%)		
0	1 (2.4%)	27 (100%)
0.5	38 (90.5%)	0 (0%)
1	3 (7.1%)	0 (0%)
	AD (n = 41)	Healthy (n = 26)
Driving experience mean (SD), y	54.7 (7.4)	51.6 (4.6)

Table 2: Comparison of participants with AD and healthy participants according to outcome measures.

Reaction time task, time in ms					
	AD	N	Healthy	N	P-value (df)
Visual cognitive	347.2 (74.7)	40	302.2 (52.8)	27	**.009** (65)
Visual motoric	271.5 (84.3)	42	219 (62.9)	27	.007 (67)
Auditory cognitive	299.1 (71.6)	42	264.4 (65.7)	27	**.047** (67)
Auditory motoric	235.6 (72.3)	42	193.7 (62.4)	27	**.014**[61]
Combined cognitive	602.3 (124.4)	42	503.4 (106.9)	27	**.001** (67)
Combined motoric	284.7 (85.4)	42	229.9 (74.8)	27	.008 (67)

Hazard perception task					
Answers correct	14.5 (4.11)	42	18.22 (2.81)	27	.000[6]
Reaction time, sec (timed)	5.39 (0.76)	30	4.39 (0.73)	27	**.000** (55)
Reaction time, sec (all participants)	9.84 (9.31)	42	4.39 (0.73)	27	.003 (67)

ATAVT					
Performance parameter	-1.145 (1.179)	41	0.487 (0.915)	27	**.000** (66)

Simulator rides					
Braking, yes/no	22/5	27	21/0	21	.059 (1)[2]
Brake reaction time, sec	1.82 (.42)	22	1.27 (.26)	21	**.000** (41)

Results indicate that the AD participants had a significantly shorter (motoric) reaction time to visual cues compared to the reaction time of the participants in the control group. Similar results were found on the auditory cue RT-task for both cognitive- and motoric RT. In the combined task, when presented with combined stimuli and certain

[1] As the values were not normally distributed a Mann-Witney U test was used. This test does not provide degrees of freedom

[2] A chi-square test was used. One of the cells had a count of less than five which violated an assumption of the Chi-square test. To correct this the Fisher's Exact Test was performed.

responses were to be inhibited again a similar result was apparent for both cognitive and motoric RT.

On the hazard perception task the results show that AD participants answered fewer questions correctly compared to the control group. AD participants also needed significantly more time to provide an answer compared to healthy participants.

The two groups also differed on the ability to get an overview in traffic situations. Again results showed a significantly worse performance for the AD participants.

During the simulator ride an event took place where a driver suddenly turned into the road from a parking bay, forcing the participant to brake heavily to avoid collision. In general AD participants failed to brake five times while the healthy participants failed 0 times. This effect however only approached significance. After excluding the participants who had not braked, the driver braking reaction times of those who did brake were compared among the groups. Results again show that AD participants reacted slower to the car turning into the road than the control participants did.

5. DISCUSSION

The present study aimed to determine if reaction time, hazard perception and the ability to get an overview in traffic situations differ between healthy drivers and drivers with very mild (CDR = 0.5) or mild (CDR = 1) AD. It was hypothesized that drivers with AD would react slower, have more difficulty getting an overview in traffic situations and have more difficulty perceiving possible hazards while driving.

The results are relatively straightforward. Patients with AD disease reacted significantly slower cognitively than healthy controls whether it is to visual, auditory or combined choice stimuli. These findings are in line with previous research [4,15,16]. Motoric RT, the time to press the button after recognizing the correct stimuli, was also slower for AD participants in all RT tests compared to healthy controls. This finding is also in line with previous research [4]. It was also apparent from the results that AD participants had a slower response to hazards compared to healthy participants. The results on the hazard perception task also showed that AD participants made more incorrect decisions and for the correct ones, they had a longer reaction time. It should however be taken into account that the participants responded to static images of traffic situations and therefore could not see the traffic situation develop like you normally would be able to while driving.

Patients with AD also had more trouble to quickly oversee traffic situations, this is in line with previous research as well [4]. The results on the driving simulator ride showed that there were 5 AD participants out of 27 who did not manage to brake for the car entering the road while this did not happen for the control group participants. However, it has to be noted that this result only approached significance.

Reaction time, hazard perception ability and the ability to get an overview quickly are instrumental to take part in traffic safely as the driver needs to be able to handle complex traffic situations. It is vital that the driver is able to quickly recognize what elements make up the traffic situation, how to prioritize them and what the appropriate action to take is (e.g. giving or taking priority). AD drivers need more time to get an overview in traffic situations and respond to hazards later compared to healthy controls. Once the hazard is recognized the AD driver also has a slower motoric RT which further delays the intended action (e.g. braking or evading). The AD driver is therefore at an increased risk of being involved in a traffic accident.

AD drivers use coping strategies to mitigate their impairment, for example by driving more slowly to allow for more time to react or by avoiding complex traffic situations altogether (e.g. only drive outside of rush hour or only on non-busy roads). Research shows mixed results in the effectiveness of these coping strategies. It was shown that avoidance strategies employed by older adults are related to reduced driving skills [17]. This implicates that coping strategies might actually worsen ones ability to take part in traffic safely. On the other hand research shows that older adults can use tactical compensation strategies to reduce crash risk [18]. More research is needed to: 1) determine at what reduction of hazard perception ability, the ability to get an overview and cognitive/motoric RT an AD driver can still safely participate in traffic and 2) determine the role that coping strategies play. A naturalistic study of AD drivers in which (motoric) RT, traffic overview and hazard perception outcomes are related to crash or near-crash involvement could be performed to this end. This could aid in deciding when it is the right time for the AD driver to stop driving. This is of vital importance not only with traffic safety in mind but also the wellbeing of AD patients as driving is a meaningful instrumental activity of daily living that enables them to stay socially connected to the community.

REFERENCES

[1] Cunningham, E.L., McGuinness, B., Herron, B. & Passmore, A.P. (2015). Dementia. *Ulster Med J, 84*(2), 79 – 87.

[2] Cornutiu, G. (2015). The epidemiological scale of Alzheimer's disease. *Journal of clinical medicine research, 7*(9), 657.

[3] Haque, R. U., & Levey, A. I. (2019). Alzheimer's disease: A clinical perspective and future nonhuman primate research opportunities. *Proceedings of the National Academy of Sciences, 116*(52), 26224-26229.

[4] Piersma, D., de Waard, D., Fuermaier, A., Davidse, R., Bredewoud, R., Brouwer, W., & Tucha, O. (2016). How to investigate fitness to drive of patients with Alzheimer's dementia. In International Conference on Traffic and Transport Psychology 2016.

[5] Guarino, A., Favieri, F., Boncompagni, I., Agostini, F., Cantone, M., & Casagrande, M. (2019). Executive functions in Alzheimer disease: a systematic review. *Frontiers in Aging Neuroscience*, *10*, 437.

[6] Förstl, H., & Kurz, A. (1999). Clinical features of Alzheimer's disease. *European archives of psychiatry and clinical neuroscience*, *249*(6), 288-290.

[7] Breen, D. A., Breen, D. P., Moore, J. W., Breen, P. A., & O'Neill, D. (2007). *Driving and dementia*. *Bmj*, *334*(7608), 1365-1369.

[8] Toepper, M., & Falkenstein, M. (2019). Driving fitness in different forms of dementia: an update. *Journal of the American Geriatrics Society*, *67*(10), 2186-2192.

[9] Classen, S., Wang, Y., Winter, S. M., Velozo, C. A., Lanford, D. N., & Bédard, M. (2013). Concurrent criterion validity of the safe driving behavior measure: A predictor of on-road driving outcomes. *American journal of occupational therapy*, *67*(1), 108-116.

[10] Hughes, C.P., n.d.. *Clinical Dementia Rating (CDR) Scale*. Retrieved May 12, 2021, from: https://www.mdcalc.com/clinical-dementia-rating-cdr-scale#evidence

[11] Schuhfried, G. & Prieler, J. (2008). RT. SCHUHFRIED gmbh, Mödling.

[12] Schuhfried, G. (2009). ATAVT. SCHUHFRIED gmbh, Mödling.

[13] Schufried. (n.d.). *The Vienna Test System*. Retrieved March 23, 2021, from: https://www.schuhfried.com/en/vienna-test-system/

[14] Rasch, G. (1993). Probabilistic models for some intelligence and attainment tests. MESA Press, Chicago, Illinois.

[15] Hutchison, K. A., Balota, D. A., & Ducheck, J. M. (2010). The utility of Stroop task switching as a marker for early-stage Alzheimer's disease. *Psychology and aging*, *25*(3), 545.

[16] Tse, C. S., Balota, D. A., Yap, M. J., Duchek, J. M., & McCabe, D. P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology*, *24*(3), 300.

[17] Schulz, P., Beblo, T., Spannhorst, S., Labudda, K., Wagner, T., Bertke, V., Boedeker, S., Driessen, M., Kreisel, S.H. & Toepper, M. (2019). Avoidance behavior is an independent indicator of poorer on-road driving skills in older adults. *The Journals of Gerontology: Series B*, *75*(10), 2152-2161.

[18] De Raedt, R. & Ponjaert-Kristoffersen, I. (2000). Can strategic and tactical compensation reduce crash risk in older drivers?. *Age and ageing*, *29*(6), 517-521.

DEVELOPING AN INTELLIGENT SERVICES SYSTEM IN TRANSPORT SECTOR

Aristi Karagkouni, PhD Candidate
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece
arkaragk@econ.duth.gr

Maria Sartzetaki, Senior researcher/Adjusted lecturer
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece
University Neapolis, Pafos, Cyprus
msartze@econ.duth.gr

Dimitrios Dimitriou, Professor Associate
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece
ddimitri@econ.duth.gr

ABSTRACT

In a competitive business environment, series of decisions should be connected to data leverage of strategic importance – the data-driven management could mitigate risks and increase awareness, while re-engineering the out-performance production process will be much easier to be implemented. This paper deals with the presentation of the key components for developing a data driven dynamic corporate management to assess uncertainty and unpredicted risks. The business sectors of the analysis focused on transport and supply chain infrastructure operators where the capital-intensive business environment may dramatically affect by out of time decisions. Adopting a System of System (SoS) approach the utility of data driven service in the real business is addressed. By a dedicated literature review the key areas for action are highlighted and the spots for innovation are depicted. The analysis includes the results from a questionary survey, where the needs of transport infrastructure operates are evaluated and the key priorities are presented. Key objective is to present the transport market needs for a data driven management system (platform) upon the development of intelligent services to support decisions for operators and large enterprises managing transportation systems and critical transport infrastructure. Conventional wisdom is to present the key components for an intelligent data-driven risk assessment system for the different group of potential users' in transport sector, which addressing the stakeholder's expectations, shareholder's prospects and management abilities to response. The research outputs will provide essential messages for managers and decision makers towards business intelligence and corporate management performance optimization, especially in sector of transport and supply chain.

Keywords:
Intelligent services, Research Infrastructure, Prospects and Transportation

1. INTRODUCTION

The research needs are demanding both at disciplinary level – requiring more and more often the acquisition of diverse complementary data from different methods and instruments – and also across disciplines – as the study of complex phenomena demands to jointly analyse data obtained from Research Infrastructures belonging to different scientific domains [1]. Transport infrastructure, which includes physical networks, terminals and intermodal nodes, information systems, as well as refuelling and electrical supply networks, is necessary for the safe, secure operation of road, rail, civil aviation, inland waterways and shipping and is crucial to the European Union's (EU) economic growth and social development. Unlike other transport operations, transport infrastructure is owned by public sector organisations at country or regional level [2]. An issue of concern arises from the growing pressure on existing transport systems, combined with regularly underfunded maintenance activity, the growing digital connectivity, infrastructure vulnerabilities to man-made or natural disasters and the carbon efficiency of infrastructure reuse. Life-cycle optimisation and integrated efficient operation can lead to improvements in the existing and future infrastructure stock.

Intelligent infrastructure is the deep embedding of sensing, computing, and communications capabilities into traditional urban and rural physical infrastructures such as roads, buildings, and bridges for the purpose of increasing efficiency, resiliency, and safety. According to [3] "Intelligent Infrastructure" is defined as "the integrated sensing and data analytics with municipal capabilities and services that enable evidence-based operations and decision making" which is:

- Descriptive: Provides an accurate and timely characterization of current state, e.g., water level in a storm drain or traffic congestion.
- Prescriptive: Recommends immediate and near-term actions, e.g., re-routing traffic or dispatching onsite service personnel.
- Predictive: Anticipate future challenges and opportunities, based on assessment of the current state, patterns of past activity and available resources and capabilities, e.g., street-level flooding by incorporating water sensors, weather patterns and runoff capabilities.
- Plan-Full: Guides complex decision making and scenario planning, incorporating economic data, to inform future investment.

This paper deals with the analysis of the real requirements for digital services provided by 'Intelligent Research Infrastructure for Shipping, Supply Chain, Transport and Logistics' (EN.I.R.I.S.S.T.) project in order to support policy making, strategic and business planning and decision making in transportation sector [4]. EN.I.R.I.S.S.T. is a unique and pioneering Research Infrastructure that aims to fill a significant existing research gap in the fields of Shipping, Supply Chain and Transport in Greece. The research outputs provide results about the real needs for data analytics, event observation, cost-benefit analysis, market trends and forecasting, for a variety of potential users in supply chain business ecosystem.

2. BACKGROUND RESEARCH AND LITERATURE

2.1. Research Infrastructure (RI) objectives

Since the business ecosystem in which organizations act, operate and regulate is changed predictions, forecasts and assessment of scenarios have always been necessity. Especially, recent years where technological innovation and external events such as climate change and epidemic are drivers for functional changes in market, management and level of service. Therefore, in today's management the wording of the term "business resiliency" is wider than the typical evaluation of the financial performance and it also deals with the internalization of the risks caused of the external factors [5]. The evaluation of the impact of these factors in business ecosystem is know the baseline of the risk assessment promoting new areas of strategic and business planning in the frame of operational contingency, product line continuity and financial productivity.

In transport sector, organization that cannot react quickly to changing conditions and that cannot have plans to mitigate future risks, in other words cannot foreseen the future with a degree of accuracy are boomed to extinction and are characterized as un-resilient [6]. In this outline, key driver of creating Intellignet Transport Systems bargains with the advancement of precision for estimations around long-term in medium- and long- term time skyline. The complexities of nowadays worldwide commerce environment and globalization of transport trade have boosting the transport division in tall request levels pulling in expansive sum of capitals to oblige the growing demand; but on the other create huge sums of information and an overpowering got to extricate exact, important and valuable data from these information [7].

Research Infrastructure are essential tools for transport sector ecosystem to support planning, management and decision making to meet the new conditions of the future for which the knowledge is imperfect. In addition, the need for business analytics cuts across all functional lines as well as across all types of organizations [5]. Scenario planning based on accurate data are needed in master and business planning, finance, marketing, personnel and production areas covering cross the board of transport sector from small business units up to large multinational companies and national organizations.

RIs must be recognized as long-term strategic investments at all levels, deeply rooted in society, and indispensable both for enabling and developing excellence in their respective scientific domains, and also as key players contributing to competitiveness with a very large perimeter. The long-term benefits of Research Infrastructures to society at large are unquestionable irrespective of the size or scientific focus of the RI concerned [8].

2.2. Intelligent Infrastructure contribution to Research and Innovation

Intelligent infrastructure is the profound implanting of detecting, computing, and communications capabilities into conventional urban and country physical frameworks such as roads, buildings, and bridges for the reason of expanding effectiveness, resiliency, and security. For example, inserting controllers, crossing point schedulers, and sensors along streets makes modern capabilities to control activity signals and optimize activity stream. Creating intelligent infrastructure gives the implies to supply modern efficiencies in control era and transmission, moved forward versatility to common and human-originated disturbances, and streamlined integration of unused sources vitality sources such as wind [3].

Intelligent Infrastructure can incrementally increment operational execution and the capacity to direct changes through decision support. These "loops" or cycles of learning span computerization and choice back to the possible generation of generalized information. For case, progressed transportation frameworks might incrementally learn to oversee diverse designs of activity, at that point give choice back for proactively overseeing extraordinary cases (e.g., disaster reaction), at that point bolster arranging and prioritization for unused road/control adjustments, and at long last report generalized information that can be connected over diverse cities with changing transportation capabilities.

One of the chief difficulties in developing integrated intelligent tools in transport sector is an unexpected and significant shift of the economic outputs caused on unpredicted events. Among such factors are changes in oil and energy prices, inflation surges, currency rates, mobility restriction measures and border/visa policy affected demand in small or large regions and most of the times impact even whole economic systems such is Euro-zone in time of COVID-19 pandemic. Therefore, the cornerstone of the term "intelligence" deals with the use of large size information (megadata) to (a) predict future level of demand or economic output (revenues, prices, etc); (b) assess risks and business resiliency; and (c) provide accurate data to support decisions. Intelligent Infrastructure has the potential to transform daily life and civic services across many configurations of municipal systems and services.

2.3. EN.I.R.I.S.S.T. Project Overview

EN.I.R.I.S.S.T. is a unique and pioneering Research Infrastructure that aims to fill a significant existing research gap in the fields of Shipping, Supply Chain and Transport in Greece. It combines the collection and processing of data (with the aim of protecting privacy and copyright), the development of innovative models and programming techniques, the development of useful applications, secure and user-friendly, and finally the development of digital observatories aimed at support for public and private stakeholders (businesses, public bodies, research organizations, etc.) [4]. In this way, the vision of EN.I.R.I.S.S.T. is to become a center of excellence that will promote and support research in its scientific fields. The objectives of EN.I.R.I.S.S.T. are:

- Develop an intelligent research and business platform to support key economic activities and small and medium-sized enterprises active in the areas of research infrastructure interest.
- Collect process and provide researchers and users with information & tools on national & international passenger & freight transport including sea, air, inland and intermodal transport.
- Support stakeholders in original research, investment plans and policies (academic community, researchers, infrastructure operators, private & public companies, policy makers)
- To create a multi-dimensional institution of economic & research development for Greece, by creating new & enhancing existing networks that will ensure the flow of knowledge & information

The Intelligent Research Infrastructure EN.I.R.I.S.S.T. is composed of 11 Partners which include 8 Educational Institutions and 3 Research Centers in Greece. Researchers specialized in shipping, supply chain and transportation, as well as software developers, come together and unite their expertise. This specialization promotes research excellence, ensures an integrated approach to the challenges of the above sectors, leading to successful implementation of the infrastructure. In addition to the major partners, the EN.I.R.I.S.S.T. infrastructure ecosystem is made up of Industry Representatives, Private Companies, Public Enterprises, Policy Makers, Independent Researchers and Citizens.

3. CONCEPTUAL ARCHITECTURE

During the architectural design process, critical issues related to the development and operation of the infrastructure were addressed. It was found that many of the requirements of the users were conflicting and that the available resources are limited. A service-oriented approach is proposed for infrastructure development. According to it, each functional unit is implemented as a service, which is integrated into a platform that will contain a set of different services and communicates in a predetermined way with the rest of the components.

A mechanism is required to maintain flexibility in the design of the architecture, ie to be able to add and remove services without significantly increasing the computational cost. It is noted that the development of the services of the various platforms will take place in parallel. In order for EN.I.R.I.S.S.T to provide new opportunities in its areas of activity, in which complex processes with economic, social and technological elements are performed, it is required in the design of its architecture to be taken seriously:

- The size of the data that its platforms have to manage.
- The number of calculations of its services prosedures
- The complexity of the procedures performed by its services.
- All modern technological and business trends.

In general, the research infrastructure will use multiple data sources and a set of new methods that process and analyze data from these data sources. It therefore makes sense for its applications to have high data requirements. This has a decisive influence on EN.I.R.I.S.S.T's development and conservation strategy. Based on the above, it appears that EN.I.R.I.S.S.T will integrate a large number of applications, which will provide useful services to research centers and companies. A wide variety of technologies will be used to implement them, which testifies the inherent complexity of the systems and services involved in data analysis. In fact, their complexity increases as they develop and expand.

The presentation of the architecture begins with the use cases supported by the infrastructure system and then the consolidation of the necessary elements at the functional level. Based on the usage scenarios, the cases of using the system are determined. The sequences of actions that implement the necessary processes to serve the usage scenarios are identified while at the same time the independent functions that are combined to form a use case are provided in an abstract way. These cases of use of the system reveal the degree of complexity and the need for communication and orchestration of the various elements that will be integrated into the EN.I.R.I.S.S.T infrastructure. The proposed architectural approach aims to cover all complex scenarios of use of the RI while providing great flexibility in processes of expansion and change of the system.

The EN.I.R.I.S.S.T research infrastructure will implement and provide a set of platforms for Shipping, Supply Chain and Transport; (a) the ecomarine platform, (b) the maritime heritage platform, (c) the passenger platform, (d) the infrastructure platform, (e) the inland and intermodal freight platform and (f) a decision support tool for shipping and financial markets. Each platform aims to make available a set of data and web services (with as user-friendly interfaces as possible), and data analysis tools/processes. These services aim to provide access to data and through their analysis to help both researchers and companies to devise new strategies to become more competitive. In order to be able to do this, it is necessary to somehow record the needs of the RI's potential users and then to form some user groups.

The presentation of the research infrastructure architecture follows a logical path that starts from the supported use cases and proceeds to its functional parts at a conceptual level. The architecture must be scalable and flexible to modifications. In order to be successful, the architecture design must cover the following aspects:

- Easy to use
- Efficiency
- Security
- Scalability
- Reliability
- Maintenance

3.1. Modeling Framework Development

The overriding consideration in choosing an analysis methodological framework is that the results must facilitate the decision making process of the organizational managers. The essential requirement is that the analysis outputs should be understandable and reasonable for the users. Therefore, the results outputs should be accurate, timely and understood by the users so that to use the analysis outputs help them to produce better decisions.

The recognition that any analysis techniques operate on the data generated by historical events and situations leads to the identification of the following four step in Infrastructure architectural developing process:

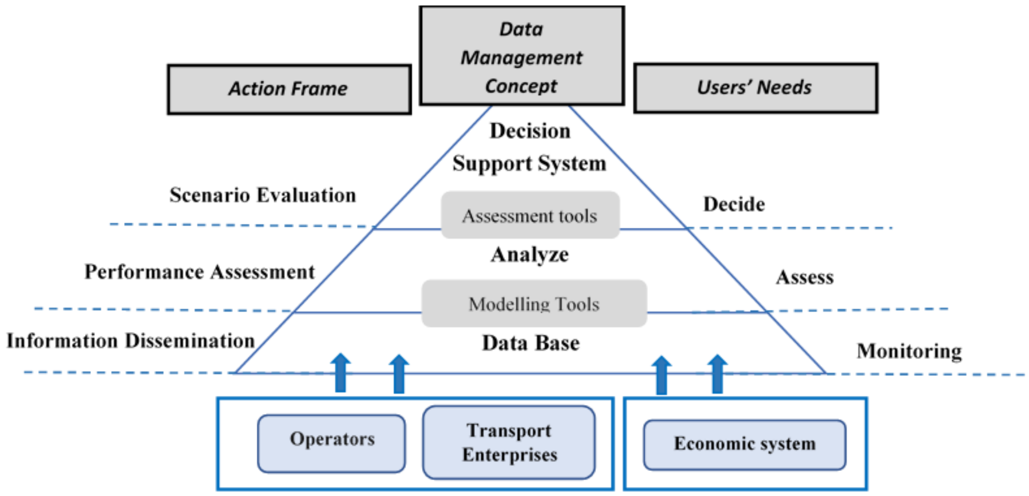


Figure 1: Data inflow and analysis levels for developing RI services

1. Data collection – (Development of data base)

Consists of data gathering from operators, transport enterprises and the economic system related to demand (traffic), socioeconomic characteristics, operational characteristics (capacity) etc. Suggest the importance of getting the proper data and making sure that are those are needed. This is often the most challenging step of the entire RI development and the most difficult to monitor since subsequent step can be performed on data whether relevant to the problem at hand or not. Collection and quality control problems usually abound whether it comes necessary to obtain pertinent data in an organization.

2. Model building – (Creation of metadata warehouse)

Consists of the analysis of the incoming data, using quantitative methods and modelling analysis tools, in order to create KPIs and forecasting. Modell building involved fitting the collected data into different quantitative models delivering a variety of outputs. Often a balance must be struck between a sophisticated modelling framework that offers slightly more accuracy and a simple approach that is easily understood and gains the support of users. Obviously, judgment is involved in this selection process.

3. Model extrapolation – (Service outputs and interface)

Consist of the actual modeling extrapolation that occurs once the appropriate data have been collected, storage and tested. Often the accuracy of the process is checked for recent periods in which actual values are known. It is essentially the creation of an intelligent decision support system (DSS), as shown in Figure 1. Monitoring the RI's potential users' needs and expectations is a key pillar of this process, as the developed services are mainly aimed at professionals and researchers in the transport sector. RI's data bases are powered by data providers from the transport sector, such as transport infrastructure operators, transport enterprises and authorities. Data inflows are analyzed by using appropriate modeling tools in order to rate the developed services, according to users' needs and expectations, in terms of support decision making in the transport sector.In order to develop the EN.I.R.I.S.S.T. Research Infrastructure, it is necessary to consult with potential project users in order to gather their requirements and needs regarding the research infrastructure's services related to shipping, transport and supply chain activities. This process is performed in 3 steps, as shown in Figure 2.

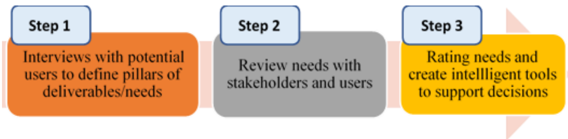


Figure 2: Depiction of process to define the RI outputs

In the first step, interviews are conducted with executive professionals and researchers from a variety of organizations related to the above sectors, in order to define the main pillars of the project's deliverables. Then, a questionnaire survey is applied in order to gather and review the potential users' needs and prospects in shiping, transport and supply chain sectors. The third and final stage concerns the rating of the potential users' needs and their relevance to the services provided by EN.I.R.I.S.S.T. research infrastructure. The main purpose of this infrastructure is to create an intelligent tool, which will be able to receive and analyze data in real time in order to support decisions in transport, shipping and supply chain sectors. For this reason, it is particularly important that the services offered by the infrastructure and the platforms that will be created meet the needs of its potential users.

The ability to manage the development of a such process is crucial for the value of the outcome. The usefulness and utility of the service can be improved if management adopts a more realistic attitude. The analysis and services outputs should not be viewed as a substitute for prophecy but rather as the best way of identifying and extrapolating establish patterns of relationships to support planning, risk assessment and decision making. Given that, the key questions raised if the outputs is to be properly conducted, where such questions are:

Regarding the RI offered services content:

- Why the analysis outputs needed?
- Who will use the offered services and what are their special requirements?
- What level of detail or aggregation is required and what is the proper time horizon?
- What data are available and will the data be sufficient to generate the needed analysis outputs (offered services)?

Regarding the RI modelling framework for providing metadata:

- How accurate and in what time frame or measurement type to be?
- Will the modelling outputs be made in time to support decision making?
- Does the outputs clearly understand how it will be used in the organization?
- Is the feedback process available to evaluate outputs after it is made and to adjust the calculations accordingly?

4. CONCLUSION

Transport infrastructure, which includes physical networks, terminals and intermodal nodes, information systems, as well as refueling and electrical supply networks, is necessary for the safe, secure operation of road, rail, civil aviation, inland waterways and shipping and is crucial to the European Union's (EU) economic growth and social development [9]. In order to address current socio-economic challenges within an ever-changing complex and competitive environment, the transport sector requires new technological developments. This will be achieved through research and innovation (R&I), which allows new quality standards in relation to the mobility of people and goods.

Moreover, Research Infrastructures play a key role in the advancement of knowledge and technology and provide an important link in the innovation chain. The purpose of a RI is to provide metadata to support transport sector business ecosystem (operators, authorities and organizations) to observe and monitoring up to day performance, assess internal and external risks and support decisions in planning, management and financing. The analysis presented in this paper pointed out that the development of a RI in transport sector must adhere two primary rules:

- The metadata must be technically correctly and produce outputs accurate enough to meet the transport system ecosystem;
- The procedure and its results must be effectively presented (interface) according to the needs of managers and administrators so that the modelling outputs are utilized in decision making process providing advantage to transport business resiliency.

Acknowledgement

The paper research outputs is supported by the research project "ENIRISST – Intelligent Research Infrastructure for Shipping, Supply Chain, Transport and Logistics" which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

REFERENCES

[1] European Strategy Forum on Research Infrastructures (ESFRI) (2018) Roadmap & strategy report on research infrastructures, Part 1, pp. 8-14.

[2] European Commission, Commission staff working document — Towards clean, competitive and connected mobility: the contribution of transport research and innovation to the mobility package, 2017, SWD(2017) 223, Brussels.

[3] Mynatt E., Clark J., Hager G., Lopresti D., Morrisett G., Narhstedt K., Pappas G., Patel S., Rexford J., Wright H., & Zorn B. (2017). A National Research Agenda for Intelligent Infrastructure, http://cra.org/ccc/resources/ccc-led-whitepapers

[4] Intelligent Research Infrastructure for Shipping, Supply Chain, Transport and Logistics (EN.I.R.I.S.S.T.) (2021), Available at: https://www.enirisst.gr

[5] Dimitriou D., Sartzetaki M. (2020) Assessment framework to develop and manage regional intermodal transport network, Int. Journal: Research in Transportation Business & Management, DOI: https://doi.org/10.1016/j.rtbm.2020.100455

[6] Dimitriou D., Karlaftis, M., Stathopoulos, A. (2006). Public Transport Sampling for Special Events: A venue based approach, Annual Transportation Research Board (TRB), Washington D.C., USA, Proceedings of Transportation Research Board

[7] Sartzetaki M., Dimitriou.D., Karagkouni A. (2019), Optimum allocation of time resources for transport operation enterprises, International Journal of Business Research and Management (IJBRM), Volume 9, Issue 1, pp.29-33.

[8] European Strategy Forum on Research Infrastructures (ESFRI), Long-Term Sustainability of Research Infrastructures, 2017, Vol. 2, pp. 10-12.

[9] Gkoumas, K., van Balen, M., Ortega Hortelano, A., Tsakalidis, A., Grosso, M., Haq, G., Pekár, F. (2019) Research and innovation in transport infrastructure - An assessment based on the Transport Research and Innovation Monitoring and Information System (TRIMIS), EUR 29829 EN, Publications Office of the European Union, ISBN 978-92-76-09539-2

PRIORITIZATION OF THE INVESTMENTS IN TRANSPORT INFRASTRUCTURE TOWARDS BUSINESS RESILIENCY AND SUSTAINABILITY

Aristi Karagkouni, PhD Candidate
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece
arkaragk@econ.duth.gr

Maria Sartzetaki, Senior researcher/Adjusted lecturer
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece
University Neapolis, Pafos, Cyprus
msartze@econ.duth.gr

Dimitrios Dimitriou, Professor Associate
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece
ddimitri@econ.duth.gr

ABSTRACT

Investing in transport infrastructure projects is a major concern in terms of the planning process and strategic analysis in order to achieve business resilience and sustainable development. Decision making implies making choices, specifically in the case of transportation infrastructures related to policy making for budget allocations and choices between alternatives for a new transport infrastructure. Governments and decision makers promote investments in transportation infrastructure projects in order to achieve socioeconomic and sustainability goals. Arguments for significantly boosting investment, in capital intensive transport infrastructures, in order to achieve sustained growth rest on high returns on investment in capital-scarce environments and the pressing deficiencies in these areas. One of the most critical issues for decision makers is to select which public investment projects will be funded. The proposed methodology provides an integrated decision support framework based on a combination of ex-ante methodologies for evaluating the impact of the investments on business ecosystem resilience and sustainability. The main content of this paper deals with a dedicated literature review on the subject, promoting the key challenges for planners, managers, and decision makers in the sector of transport. By a system-of-system approach, the role of sustainability in air transport sector ecosystem is depicted, providing evidence that its linkage with business performance is a major challenge for planners, managers and decision makers towards business resiliency and competition.

Keywords:
transport infrastructure investment, transport development, climate change, business resiliency, sustainability

1. INTRODUCTION

Transport is a major contributor to economic growth that requires operational productive and efficient infrastructures and services. The necessary condition, that is, ensuring economic, social and environmental sustainability, also affects the challenges and opportunities facing the built environment and, especially, the functioning of critical organizations and businesses [1]. Global sustainability challenges are shaping the way business operates in the 21st century.

Businesses are under increasing pressure from multiple stakeholders (e.g. shareholders, customers, employees, society) to manage their positive and negative impacts with clear responsibility and strategic intent. Hence, the transformation to sustainable transport requires a redirection, rather than any substantial increase, in infrastructure expenditure [2]. The need for new approaches in transport planning, management and policy development, new technology, and the global resolve toward achieving the 2030 Agenda for Sustainable Development and the Paris Climate Agreement is crucial for the operation of modern businesses [3,4]. Moreover, the analysis of the environmental implications impacts is essential for the transport industry operation and economic environment. Sustainable transport variables are defined as a development process that incorporates the basic principles of sustainable development. These processes should comply with the objectives of environmental, social impact and economic growth [5]. In this context, the adaptation to Climate Change implications in critical infrastructures is crucial for ensuring business resiliency and sustainability.

The main aim of this paper is to provide a decision support framework towards investments on business ecosystem resilience and sustainability. The methodology is focused on an extended break down of the effects of climate change to critical transport infrastructure based on the estimation of conditional probabilities and the severity in terms of cost to adapt or mitigate climate change impact at airports and the level of criticality in terms of time to implement policies and measures to maintain growing air transport demand. The numerical application provide results for Athens International Airport, the largest and busiest airport in Greece.

2. TRANSPORT CONTRIBUTION TO ECONOMIC DEVELOPMENT

The transport sector is an important component of the economy and a common tool used for development because of its intensive use of infrastructures [6,7]. This is even more so in a global economy where economic opportunities have been increasingly related to the mobility of people and freight, including information and communication technologies. Governments in different nations have different attitudes and priorities regarding how much attention they pay to these issues. In a mixed economy, one of the ways of measuring the success of government's influence on the economy is by comparing how it has affected logistics and transport growth. The external dimension of transport leads to opportunities and benefits of economic and social influence throughout the economy [1].

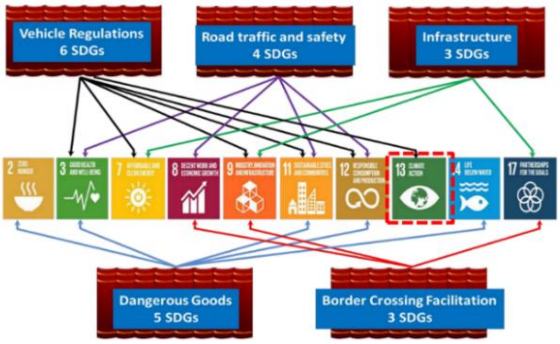
Transport is also a strategic sector of the EU economy, with transport services accounting for about 5% of the EU's gross value added and 5.2% (or around 11 million persons) of all jobs in 2016. It directly affects the everyday lives of all EU citizens and ensures the flow of goods to consumers from more than 11 million EU producers and manufacturers [8]. This emphasizes the importance of having a well performed transport system for economic european integration. Well thought-out, sustainable and fully interconnected transport networks are a necessary condition for the completion and correct functioning of the European single market. Developing the EU's transport infrastructure requires a considerable financial outlay. The Commission estimates that the total investment needs in this area are about €130 billion per year, with further significant investment needed for maintenance. The successful coordinated deployment of intelligent transport management systems, which is currently ongoing, is vital to the achievement of a pan-European, co-modal and truly integrated transport system. Moreover, automation, digitalisation and shared mobility are rapidly expanding trends that have the potential to make transport systems more efficient. However, new technologies and mobility patterns also bring challenges relating to the suitability of the legislative framework, privacy protection, safety, liability and data security.

Passenger and freight transport volumes in the EU have been steadily rising in recent decades, from 5,335 billion passenger-kilometres in 1995 to 6,802 billion in 2016 for passenger transport and from 2,846 billion tonne-kilometres in 1995 to 3,661 billion in 2016 for freight [8]. The Commission estimates that they will continue to grow, albeit at a slower pace than in the past. It projects an increase of 42% for passenger transport activity and 60% for inland freight between 2010 and 2050. The projected rise, for international maritime transport is still greater, at 71% over the same period.

Efficient transport services and infrastructure are necessary to utilize the economic strengths of all EU regions, to support the internal market and growth and to promote economic, territorial and social cohesion. Given its central role, transport also has close ties to policy areas such as the environment, jobs and growth, competition, social policies and digitalisation. The adaptation of infrastructure to new mobility patterns, as well as the deployment of new infrastructure for clean, alternative fuels, pose additional challenges that necessitate new investment and a rethinking of network and business model design. Charging infrastructure is essential to facilitate the development of electro- mobility, in particular the uptake of electric vehicles. In 2013, the EU launched its clean fuels strategy, which is intended to ensure a network of alternative fuel stations (including electric charging points) with standardised design and use

3. SUSTAINABLE TRANSPORT

Transport drives development, linking people, connecting local communities to the world, building markets and facilitating trade [2]. In turn, sustainable transport can drive sustainable development. Sustainable transport is the provision of services and infrastructure for the mobility of people and goods - advancing economic and social development to benefit today's and, mostly, future generations which will be able to live and produce sustainably - in a manner that is safe, affordable, accessible, efficient, and resilient, while minimizing carbon and other emissions and environmental impacts [2]. However, all goals can not be reached simultaneously. Critical is the role of priorities given by the industry regarding the achievement of sustainability. For istance, it is necessary to maintain a balance between the measures taken to reduce emissions from air transport and to serve the ever-increasing demand of the sector. The challenges are great, but so are the opportunities, in developed and developing countries alike, for visionary decisions now and in the coming years in the realm of transport that will set cities and nations on a sustainable development path. This path will be shaped by ambitious goals, targets and indicators. Progress will need to be monitored and evaluated, with course corrections where necessary. The 2030 Agenda for Sustainable Development charts this kind of path to sustainable development more generally, and the guideposts are the 17 Sustainable Development Goals (SDGs). Accomplishing the SDGs will rely on advances in sustainable transport. Some SDGs are directly and indirectly connected to sustainable transport through targets and indicators, as illustrated in the figure below.



Source: United Nations Economic Commission for Europe, 2017 [3]

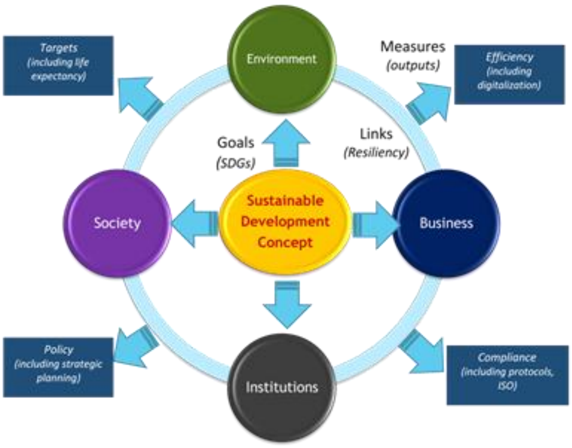
Figure 1: Linkage between transport sector and the United Nations' Sustainable Development Goals

Effective transport planning combines the need for short-term deliverables with a long-term strategic view, incorporating the social, economic and environmental aspects of transport and of development more broadly [8]. While consideration of both short- and long-term needs is common sense for all planning, it is particularly relevant for transport because of its multi-faceted nature and the resource-intensive, locked-in quality of many transport infrastructure and systems decisions. The business case for sustainable development is related to effective and long-term planning. For example, investing in green energy, or resilient infrastructure, for instance could be proved efficient and save money for companies over the longterm [9]. For private sector companies, and even, as noted above, for local and national governments, competition is a powerful driving force, and competing to become more sustainable will pay dividends over the long-term for the companies, the governments and the world at large [7]. Resilience planning is an important feature of sustainable transport development, ensuring that passenger and freight transport networks including infrastructure, service and operations are able to adapt well to climate change related events and other chronic stresses, such as high unemployment and endemic violence and acute shocks, such as earthquakes and terrorist attacks. Moreover, resilience planning could lead to more effective adaptation of the transport sector to unexpected events which can cause llong-term movement disruptions, like the Covid-19 pandemic which has a dramatic impact, especially on the aviation sector [10].

4. SUSTAINABLE TRANSPORT AND CLIMATE CHANGE

Transport accounts for approximately a quarter of all greenhouse gases (GHG) emissions in the EU. As transport emissions started increasing again since 2014, the sector is becoming one of the main challenges to the EU's overall decarbonisation goals. In October 2014, the EU leaders adopted the 2030 climate and energy framework, including a target of at least a 40% reduction in greenhouse gas (GHG) emissions by 2030 (relative to 1990). In 2015, the EU and all 28 Member States signed the Paris Agreement. Under this agreement, they were requested to submit long-term plans by 2020 showing the efforts by each country to reduce national emissions and adapt to the impacts of climate change. According to the European Environment Agency, transport emissions rose between 1990 and 2007 and then fell until 2014. In 2015 and 2016 they rose again. This means that the sector has become one of the main challenges to the EU's overall decarbonisation goals. Under current policies, account being taken of the expected growth in freight and passenger transport, by 2050 GHG emissions from transport are projected to decrease by 15% relative to 2005. However, emissions would still be 10% higher in 2050 than in 1990, owing to the fast rise in transport emissions during the 1990s. Achievement of the emissions reduction targets will require a fundamental shift towards using less energy, and cleaner energy, as well as the more efficient use of transport infrastructure.

Mega infrastructure projects are typically designed for a long-term horizon, requiring sufficient infrastructure provisions to be included at the beginning of the project to ensure its sustainability in the long run [12]. With these preventive measures to protect infrastructures, airports are expected to remain resilient against future climate changes and adverse weather conditions well into the future. The linkage between the key pillars which ensures the resilient and sustainable business performance is shown in Figure 2. The linkage between environmental awareness and the business ecosystem is crucial and the appropriate measures to ensure the balance between these two pillars could contribute to the efficient business function. In this context, the adaptation to climate change implications in critical infrastructures is crucial for ensuring business resiliency and sustainability.



Source:Dimitriou et al., 2020 [13]

Figure 2: Depiction of resilient and sustainable business performance

4.1. Effect of climate change on Air Transport Infrastructures

Climate change related disasters and extreme weather events are expected to significantly increase the risk of damages on networks, systems and human assets. In view of these anticipated adverse effects, growing attention is placed on adaptation measures, in the form of preventive actions aiming to minimize induced hazards' negative impacts and to enhance cross-sectorial resilience. Transportation, as a key economic sector of today's society, is no stranger to this regime. The expected climate-related changes will impact both transportation infrastructures and networks of operations, independent of transportation means, [14]. The daily functioning of transport systems is sensitive to fluctuations in precipitation, temperature, winds, and visibility (and for coastal cities, rising sea levels with the associated risks of flooding and damages).

Aviation growth provides significant social and economic benefits. Many city regions across Europe, especially those in geographically remote locations, are highly reliant upon the air transport sector. Aviation does, however, have significant environmental costs, as is accountable for 3.6% of the total EU28 greenhouse gas emissions and for 13.4% of the emissions from transport [15]. The adverse environmental impacts associated with the growth and operations of the industry have the poten-tial to 'feed back' and restrict aviation's ability to respond to demand and contribute to sustainable development. At an airport level, the disturbance caused by aircraft noise, local air quality and even the location of sensitive habitats in surrounding areas can give rise to operational and infrastructure constraints.

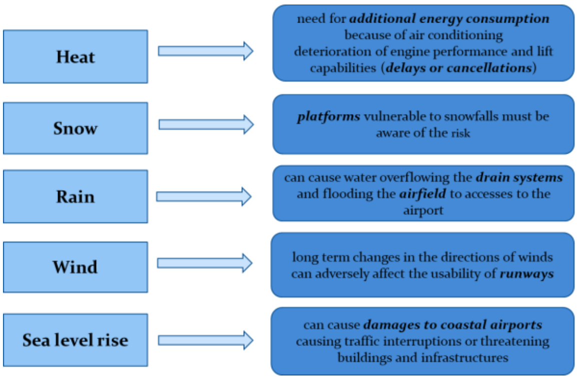
Over the past decade, the issue of climate change has emerged as another, perhaps even more significant environment threat to the growth of the industry, at both a local and a global level. This has far reaching implications for the future development of aviation in Europe. While the consequences of aircraft emissions for climate change have received much attention in the past, there has until very recently, been little focus on the effects of climate change itself or regulatory and institutional responses to it, both of which present a challenge to the future growth and development of aviation.

Furthermore, regarding air transport, and as temperatures are expected to increase in many territories around the globe, airports will need to increase the length of their runways to allow airplanes to take-off without incidents. Due to the increased temperatures, the air becomes less thick and as a result, there is not enough thrust for the airplanes to leave the ground, so a larger runway is required. Increased temperatures affect northern countries as well, in which many airports are constructed on ground that consists of ice or snow. These airports are in many cases just for military purposes or only for cargo airplanes. Nonetheless, these airports will have to be relocated, because thawing will undoubtedly have an effect on the runway materials, affecting safety on landing and takeoffs. However, it must be pointed out that possible relocation of airport infrastructure due to thawing inflicted materials depends on the correct identification of the endangered areas and acts as prerequisite.

As climate changes, precipitations will become more and more frequent and storms will intensify, leading to oftener flooded runways. Though, more advanced draining systems can be applied, another approach can be adopted. Porous asphalt can reduce flooding effects, as draining becomes easier. However, the mixture will have to be reinforced, because of the high loads with which airplanes strain the landing surface. Additionally, hurricanes will also become stronger, resulting in takeoffs that are more difficult and landing procedures due to strong winds. As a result, windbreakers are inevitable if safety is to remain in a high level. In cases where windbreakers are not an option, relocation of the runways must be undertaken. After a complete research on the direction and intensity of the occurring winds, runways must be placed in such a way as to avoid side-winds that may affect the direction of approaching aircrafts [16].

Airports are often classed as nationally critical infrastructure as they facilitate both mobility and economic growth. However, due to their fixed infrastructure and vulnerability to disruptive weather, they are particularly at risk from the potential consequences of climate change, with impacts such as sea level rise, higher temperatures and greater weather extremes creating both an operational and business risk. Therefore, to protect vital infrastructure and ensure future service continuity for airport operations, it is necessary to develop resilience to such risks [17].

There has been broad scientific consensus for several years that climate change will cause impacts such as higher temperatures, sea-level rise and greater weather extremes [18]. However, there is now growing realization that this will require all sectors of society to take action to adapt and develop resilience to such impacts [19]. Due to their fixed infrastructure and vulnerability to disruptive weather, airports are particularly at risk from the potential consequences of climate change, with impacts such as sea level rise, higher temperatures and greater weather extremes creating both an operational and business risk [20,21]. Overall, the impacts of the climate key factors on air transport infrastructure are presented in Figure 3.



Source: Dimitriou et al., 2019 [5]

Figure 3: Impacts of climate key factors on airports

4.2. Adaptation plan of climate change impacts on airports

Airports are often classed as nationally critical infrastructure as they facilitate both mobility and economic growth. Therefore, to protect vital infrastructure and ensure future service continuity for airport operations, it is necessary to develop resilience to such risks [5]. More extreme weather- and climate-related events are expected as the climate continues to change. The frequency, intensity, spatial extent, duration and timing of events are expected to increase while slow-onset incremental changes may lead to fundamental transformation of the socio-economic system. Many airports may remain vulnerable to these events as the risks of flooding, flight disruptions and cancellations become more likely. Airports need to understand the risks and initiate adaptation measures for both existing and new infrastructure, as well as managing critical operations to become more resilient to the changing climate.

An airport is a business, multimodal transport interchange, employment node, and essential piece of regional and national infrastructure for the communities it serves. As an essential service provider to a wide range of stakeholders and users, the airport infrastructure and operations must have high levels of availability, reliability and resilience [17]. Vulnerabilities to ongoing services from short- and long-term projected climate changes must be identified as part of a responsible business continuity plan.

Compiling potential impacts and consequences of extreme weather events on all aspects of the airport business and operation can enable airports to prioritize and better respond to these risks. This may be based on the airport's exposure to changing climate conditions, its sensitivity to adverse impact, or the adaptive capacity when faced with such challenges. Those impacts ranked the highest priority or of prime concern should be addressed first through detailed investigation of mitigation options and assessment of cost-benefits as part of an airport's resiliency plan.

The work on climate-change adaptation and resilience should include operational considerations on safety and security, and legal, environmental, financial and business effects on airport operations. Only comprehensive climate-change risk-management strategies will ensure the continuity of operation, profitability and asset value. Some airports are already witnessing gradual change in the investment environment, whereby investors are keen to evaluate climate-change-related risks and opportunities in accordance with the framework recommended by the Task Force on Climate-related Financial Disclosures (TFCFD) [21].

Another important element of an adaptation plan is the coordination with broader airport stakeholders and surrounding communities. An inclusive, systematic approach to collect intelligence, assess risks, and interact proactively with these stakeholders will help mitigate long-term financial, economic and operational impacts. Furthermore, as a networked infrastructure, disruptions in one airport may have a cascading impact on other airports, the wider economy, and even national resilience.

5. METHODOLOGY FRAMEWORK

5.1. Criticality Analysis

The analysis of the environmental implications impacts the key parameters of transport industry, analyzed analytically in previous section. The transport industry parameters can be affected directly, indirectly or through induced effects. Climate variations in frost, snow, fog, wind, rain, sea water level and heat are key climate factors impact each transport system business as usual scenario [22]. Therefore, the assessment of the sensibility of each transport option to these climate factors is essential to evaluate the importance for mitigation actions and adaptation changes on one hand; and define strategy towards sustainable development of transport infrastructure [5].

In this research, the criticality number used to rank the identified deviations in sensibility. It cannot be used as a risk measure and it is a product of three rough estimates. Before a criticality analysis can be performed guidelines have to be developed on how to determine P, B and S. There are no generally accepted criteria for criticality applicable to a system.

Table 1: Criticality probability factors and severity

Criticality probability factors and severity					
Probability P		Conditional Probability B		Severity S	
Frequent	A	Frequent	A	Very low	I
Probable	B	Probable	B	Low	II
Occasional	C	Occasional	C	Significant	III
Remote	D	Remote	D	high	IV
Improbable	E	Improbable	E	Very high	V

Table 2: Criticality Probability level adjusted by level of risk

Qualititave definition	Level	Specific Individual Item
Frequent	A	Likely to occur often during the year, with a probability of occurrence during year less than 1.0 but greater than 0.5.
Probable	B	Will occur several times during year, with a probability of occurrence less than 0.5 but greater than 0.2
Occasional	C	Likely to occur sometime during the year with a probability of occurrence less than 0.2 but greater than 0.1
Remote	D	Unlikely but possible to occur during the year, with a probability of occurrence less than 0.10 but greater than 0.05
Improbable	E	So unlikely, it can be assumed occurrence may not be experienced, with a probability of occurrence less than 0.05 in that life.

The criticality number is calculated as follows:

$$C_r = P * B * S \qquad (1)$$

Where: C_r = the criticality number used to rank the identified deviations in sensibility

Subsequently, a sensibility scale for variable C_r is adopted, selected from 1.0 to 5.0, where 1.0 is the impede and 5.0 is the impossible level of business as usual scenario. The aim is to determine the average severity rate for each weather factor [5]. The risk levels are coloured classified into four different levels, as shown in Table 4 [23].

Table 3: Mishap Index/ Criticality Matrix

Criticality				
Hazard/ Mishap Category Frequency	(I) High 3,5<Cr<5	(II) Critical 2,5<Cr<3,5	(III) Marginal 1,5<Cr<2,5	(IV) Negligible 1<Cr<1,5
(A) Frequent				
(B) Probable				
(C) Occasional				
(D) Remote				
(E) Improbable				

Table 4: Mishap Risk Category

Mishap Risk Category	**Mishap Risk actions**
High	Should be mitigated as soon as possible
Critical	Should be mitigated within a reasonable time period unless costs demonstrably outweigh benefits
Marginal	Should be mitigated within a reasonable time period unless costs demonstrably outweigh benefits
Negligible	Acceptable No action required

6. CASE STUDY

Air transport's contribution to the Greek economy is significant. The principal benefits are created for the customer, the passenger or shipper using the air transport service. In addition, the connections created between cities and markets represent an important infrastructure asset that generates benefits through enabling foreign direct investment, business clusters, specialization and other spill-over impacts on an economy's productive capacity. The industry supports 457,000 jobs and contributes EUR 17.8

billion to Greek GDP, accounting for 10.2% of the total [24].

Based on the above, a comprehensive review of the situation of the Greek environment in last decade is presented analytically as well as the most important and urgent environmental issues Greece faces recently. Environment protection policies and the competitive environment in transport are expected to be a significant threat for the further development of transport-associated industries in Greece, such as transport. By mitigating or adapting to the environmental weaknesses and threats, the decision to promote transport industry growth is strengthened the Greek economy recovery. Among the subsidiary environmental impacts, which the transport infrastructure might cause, will be discussed in detail in connection with the environmental impact matrix [5].

Greece is strongly committed to the implementation of the 2030 Agenda for Sustainable Development and its 17 SDGs, as they provide an ambitious and transformative framework for a new, fair and sustainable development path, which ensures a balance between economic growth, social cohesion and justice as well as protection of the environment and of the country's unique ecological wealth. On the environmental pillar, progress has been achieved in all related SDGs. Key national priorities include the shift towards a low carbon circular economy and improvement in waste reduction, reuse and recycle for creating new jobs and increasing resource efficiency. Firmly committed to the Paris Agreement objectives, Greece is already in a good place to meet its national GHG reduction targets earlier than 2030. Successes also include the full application of Integrated Water Resources Management principles considering both social aspects and ecosystem needs, significant increase in RES penetration and progress towards the full digitalization of land-uses (in land and sea), ensuring a high protection status of the county's ecological wealth [25].

The Greek economy is heavily dependent on aviation for tourism which accounts for 20% of Greek GDP. Approximately 73% of tourists arrive in Greece by air while 80% of international arrivals are tourists. Climate change not only has the potential to have significant impacts on the pattern of demand, it also has consequences for the ability of some tourism locations to meet demand when and where it arises (for example as a result of water shortages). Emissions from transport contribute the largest share of total Greek tourism industry emissions. Concerns about the environmental impact of the national transportation system in Greece are an important issue in the planning process of sustainable development. It is, also, clear that climate change has the potential to have significant impacts on the pattern of air transport demand [26].

6.1. Case study airport key features

Athens International Airport S.A. was established in 1996 as a public private partnership with a 30-year concession agreement, the Airport Development Agreement (ADA). Ratified by Greek Law 2338/95, the concession agreement grants the Company the exclusive right and privilege of the 'Design, Financing, Construction, Maintenance, Operation, Management and Development' of the new Athens International Airport. AIA is a privately managed company with the Greek State holding 55% of shares (25% held by the Greek State and 30% by the Hellenic Republic Asset Development Fund-HRADF), while private shareholders collectively hold 45%.

Athens International Airport is the largest and busiest international airport in Greece, a ground breaking international public-private partnership, the first major greenfield airport with private sector involvement. The annual passenger traffic of Athens International Airport in the last 10 years (2010 – 2019) is shown in Figure 4. Airport's passenger traffic has been increasing in recent years, following a slight decrease from 2010 to 2013. Overall, Athens International Airport reported an all-time high output of 25.57 million passengers in 2019, surpassing the previous year's traffic by 1.4 million (+6 %). This result was motivated primarily by the robust growth of the foreign market (+1.4 million or +8.6 %) while the domestic market stayed at the previous year's pace (+0.3 %) [27].

Source: Athens International Airport, 2020 [27]

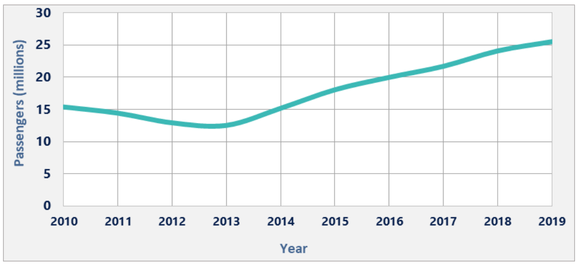


Figure 4: Annual passenger traffic at Athens International Airport, (2010 – 2019)

7. RESULTS

In this section, the results of the criticality analysis are presented in detail. Table 5 presents the criticallity results for Athens International Airport for the next 10, 30 and 50 years, based on data regarding probability, conditional probability and severity of the main climate factors according to the observation of the climatre related events in the area. The results show that heat stress is the climate factor with the highest criticality index regarding the extent to which it may affect operation of the airport. The heat stress is expected to impact essential demand patterns (especially, leisure travelers) increasing the business risks for operators and carriers.

After heat stress that seem to affect mostly Athens International Airport, wind and rain will respond to the highest score of the criticality index, while the lowest responds to snow. Water rise, fog and frost do not seem to have much of an impact on the airport over the years. Furthermore, average criticality index is 1.43 over the next 10 years and 2.00 and 2.71 over the next 30 and 50 years respectively. Based on the above it is concluded that the cost for mitigate the severity of climate factors in the infrastructure will increase significantly in the next 50 years, as the average criticality index varies between 2.5 and 3.5 (critical).

Table 5: Criticality index for Athens International Airport per climate factor and over 50 years

Criticality Index Results					
Category	Climate Factors	Next 10 years	Next 30 years	Next 50 years	**Average per factor (max 5)**
1	Frost	1	2	2	1.66
2	Snow	1	1	2	1.33
3	Fog	2	2	2	2.00
4	Wind	2	2	3	2.33
5	Rain	2	2	3	2.33
6	Water rise	1	2	3	2.00
7	Heat	1	3	4	2.66
Average criticality index (max 5)		1.43	2.00	2.71	2.05

8. CONCLUSION

This paper deals with the estimation of the Climate Change implications in the transport enterprises that are capital intensive, like airports, presenting a risk assessment methodology towards determination and rating of the operational and business risks in enterprises, caused by the Climate Change implications. The key findings contribute to the prioritization of the infrastructure investments needed to ensure adaptation to CC implications as well as business resiliency and sustainability.

The results in the numerical application highlighted that heat stress, wind and rain are very crucial for Greece, because of its 1.6 million kms coastline and the numerous spatial located island in Mediterranean. Therefore, is expected to impact essential demand patterns (especially, leisure travelers) increasing the business risks for operators and carriers [20]. The sea level rise and rain flood will affect infrastructures located close to coast is expected to very sensitive to sea lever rise and rain flood. Finally, the frost and snow are climate factors with the lowest severity for Greece due to geographical location in the Mediterranean region.

Among the many effects of CC, the existing unforeseen strain on extend capacity by new infrastructure revaluated. Findings of this research provide important messages for decision and policy makers for Greece, which is a key pathway for freights to Europe, in the context of new infrastructure investments. Findings promote key issues for effective sustainable development for regions with similar characteristic. Environment implications for transport require not only for demand accommodation activities but also national strategies to sustain these activities in long run. The costs of adaptation policies must be weighed against the economic benefits, which are very substantial in Greece.

REFERENCES

[1] Dimitriou D. and Sartzetaki M., (2019). Social dimension of aviation on sustainable development, Sustainable aviation, greening the flight path, published by Palgrave Macmillan, pp. 173-191. ISBN 978-3-030-28660-6

[2] United Nations (UN) (2016). Mobilizing Sustainable Transport for Development, Analysis and Policy Recommendations, Outlook Report, High-level Advisory Group on Sustainable Transport, available at https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=2375&menu=35

[3] United Nations Economic Commission for Europe (2017). Inland Transport Committee. SDGs and the UN Transport Conventions, available at: https://unece.org/DAM/trans/conventn/UN_Transport_Agreements_and_Conventions.pdf

[4] United Nations (UN), 2019, The Sustainable Development Goals Report, ISBN: 978-92-1-101403-7.

[5] Dimitriou D., Sartzetaki M. and Karagkouni A. (2019). Quantify the climate change implication to air transport development for tourist destinations, 9th International conference on transport research, 24-25 October 2019, Athens, Greece, conference proceedings

[6] Rodrigue, J-P. (2020). The Geography of Transport Systems. Fifth Edition. New York: Routledge, 456 pages. ISBN 978-0-367-36463-2

[7] Nistor, F. and Catalin, P. (2014). The Role of Transportation in Economic Development. “Mircea cel Batran” Naval Academy Scientific Bulletin. XVII. 10-12.

[8] Dimitriou D., Karlaftis M., Kepaptsoglou K., Stathopoulos A., (2006). Public Transportation during the Athens 2004 Olympics: From Planning to Performance, Public Transport Sampling for Special Events: A venue based approach, Annual Transportation Research Board (TRB), Washington D.C., USA, Proceedings of Transportation Research Board 2006.

[9] Dimitriou D. and Voskaki J.A., (2010). Regional airports environmental management: key messages from the evaluation of ten European airports, Int. Journal of Sustainable Development Planning, 5(2), p. 149-161

[10] European Commission (2020). Communication from The Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Sustainable and Smart Mobility Strategy – putting European transport on track for the future. Brussels, 9.12.2020. COM (2020) 789 final.

[11] European Court of Auditors (2018). Landscape Review. Towards a successful transport sector in the EU: challenges to be addressed. European Union, available at: https://www.eca.europa.eu/Lists/ECADocuments/LR_TRANSPORT/LR_TRANSPORT_EN.pdf

[12] Airport Council International Policy Brief (2018), Airport's Resilience and Adaptation to a changing climate.

[13] Dimitriou D., Karagkouni A. (2020). Assessment methodology and outputs towards sustainability and resiliency in transportation, Proceedings of the XIV Balkan Conference on Operational Research (BALCOR) 2020, 30 September - 3 October 2020, Thessaloniki, Greece. ISBN – 978-618-85079-0-6. pp. 308-312. Available at: http://balcor2020.uom.gr/BALCOR_2020_Book_of_Proceedings.pdf

[14] Dimitriou D., Sartzetaki M., (2018). Social Dimension of Air Transport Sustainable Development, International Journal of Industrial and Systems Engineering, Vol:12, No:4, 2018, pp.568-571.

[15] European Environment Agency (EEA), European Union Aviation Safety Agency (EASA), Eurocontrol. (2019). European Aviation Environmental Report 2019. ISBN: 978-92-9210-214-2. Available at: https://ec.europa.eu/transport/sites/transport/files/2019-aviation-environmental-report.pdf

[16] Stamos I., Mitsakis E., (2014). A review on climate change adaptation policies for the transportation sector, Munich Personal RePEc Archive (MPRA), Paper No. 61535, posted 25 January 2015.

[17] Dimitriou D., Sartzetaki M., (2016). Sustainable Development Variables to Assess Transport Infrastructure in Remote Destinations, International Journal of Civil and Environmental Engineering, Vol.10(10), pp 1314-1321.

[18] IPCC (2013) 'Summary for policymakers'. In: T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

[19] IPCC (2014) 'Climate Change 2014: Impacts, Adaptation, and Vulnerability. A Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change', World Meteorological Organization, Geneva.

[20] Burbidge, R. (2014) Airports and Climate Change: identifying the risks and building resilience, Airports in the Urban Network Conference 2014.

[21] Airport Cooperative Research Programme Synthesis (ACRP) (2012) ACRP 33: Airport Climate Adaptation and Resilience, Transportation Research Board of the National Acadamies, Washington.

[22] Sartzetaki M. (2019). Assessment of Climate Change Implications for Managing Transport Enterprises, Conference: ICMBEF 2019 : Management, Business, Economics and Finance 2019 Athens.

[23] Saunders, W. S. A. (2012). Innovative land use planning for natural hazard risk reduction in New Zealand. People, Environment and Planning. Palmerston North, Massey. Doctor of Philosophy: 229.

[24] Oxford Economics. (2011). Economic Benefits from Air Transport in Greece. Available at: https://sete.gr/_fileuploads/entries/Online%20library/GR/120326_Economic%20Benefits%20from%20Air%20Transport%20in%20Greece.pdf

[25] Dimitriou D., (2016). Climate Change Implications in Transport and Tourism Market Equilibrium, Climate Change Management, Springer, pp 409-424.

[26] Dimitriou D., (2018). Evaluation of tourist airports productivity towards tourism development, International Journal of Cogent Business and Management, Vol. 5 (1464378), pp. 1-15.

[27] Athens International Airport (2020). Facts & Figures. Available at: https://www.aia.gr/company-and-business/the-company/facts-and-figures

TESTING OF DYNAMIC CHARACTERISTICS OF POROELASTIC ROAD SURFACE MATERIAL

Barbara Likar, BSc
Darko Kokot, MSc
Stanislav Lenart, PhD
Slovenian National Building and Civil Engineering Institute
Dimičeva ulica 12, 1000 Ljubljana, Slovenia
barbara.likar@zag.si

ABSTRACT

Among other environmental effects of road traffic in the cities one can observe also annoying noise pollution. In the case of urban areas or also on the roads where the landscape will be loaded with different noise barrier structures, the alternative solution could be noise absorption surface layer. In the presented research various poroelastic road surface materials were developed to be used in different applications. The reduction of CO2 and other emissions from traffic is reached in case of decreasing of rolling resistance for the vehicles rolling on such a pavement, due to poroelastic rubber compound.

The poroelastic material has good characteristic properties, such as low density, high porosity, high elasticity and low stiffness as well as high thermal insulation capacity. The last one is particularly important for Slovenian roads, as freeze resistance of road construction is important due to strong winters. During the process of developing poroelastic material, firstly basic laboratory tests were performed such as determination of mechanical and acoustic properties, adhesion, workability, porosity and different types of resistance. With optimizing those properties, the optimal mixture of poroelastic material was chosen.

The presenting work will focus on the determination of a dynamic characteristic of poroelastic material that was installed in the test field in Slovenia. The tests were performed in the hollow cylindrical apparatus. The main focus of the tests was the determination of elastic and shear modulus due to different vertical and torsional loads, as well as the determination of damping during the cyclic loads.

Keywords:
surface layer, alternative material, poroelastic material, noise absorption material, dynamic torsional and axial tests

1. INTRODUCTION

The disposal of used tires is a worldwide problem, as used tires have already become a major type of waste in recent years. For example, at the end of 2003, approximately 290 million scrap tires were generated in the United States [1], while in 2009, over three million tons of scrap tires were to be landfilled or incinerated in the twenty-seven Member States, with the corresponding consequences for soil and air pollution [2].

The core of the management strategies for waste tires, in order to assess the suitability of possible treatment methods, is the waste hierarchy, which proposes the following practices for the disposal of tire stocks in decreasing order: Reduction - Reuse - Recycling - Energy recover - Disposal [3].

Reduction as the most desirable option cannot help to reduce the amount of waste tires already in landfills. The process of retreading tires, as an option of reuse, has a limit as tires cannot be treated in this way indefinitely.

Tyre shreds are produced primarily to reduce the volume of waste tires transported after collection. Within the European Union there is a ban on the landfilling of tire material to reduce the overall volume of landfill and to encourage recycling [Directives 1999/31/ EC & 2003/33/ EC] [4, 5]. Until recently, the main disposal option was energy recovery in industrial processes. However, in the European Union, legislative measures have been taken to promote the recycling and recovery of scrap tires and the reuse of tire materials in construction works is listed as a disposal option.

Recycling includes all uses of scrap tires in civil engineering and ground rubber applications that require little or minimal processing of the tires. Examples of applications for ground rubber include [6]:

- Athletic/recreational surfaces: Use in artificial sports turf, natural (grass) turf, and playground cushioning to protect children when they fall.
- Molded and extruded products: Mats, bumpers and other creative products.
- Rubber-modified asphalt and sealants: addition of ground rubber to asphalt binder to improve highway performance characteristics, including road durability.
- Tires/Automotive: Use of ground rubber from scrap tires in the manufacture of new tires, in rubber compounds for retreading worn tires, and in molded automotive parts.
- Other (mainly coarse rubber): Used as colored mulch in landscaping.

Tire shreds have interesting technical properties that can be used to advantage in civil engineering applications. Some characteristic properties of tyre shreds are low density, high elasticity, low stiffness, high drainage capacity, high thermal insulation capacity and intrinsic wear resistance. These properties open up opportunities to use the material in innovative ways, as was the case in the PERSUADE project.

1.1. PoroElastic Road SUrface: an innovation to Avoid Damages to the Environment

Low-noise road surfaces are now widely accepted as a cost-effective means of combating traffic noise. Noise reduction can be achieved by optimising the wearing course surface texture and porosity. In this way, a lower limit of 3 dB noise reduction on average has been achieved compared to ordinary dense asphalt concrete. Another significant advance has to resort to another noise-relevant property, namely elasticity, by which the noise due to tyre vibrations can be largely suppressed. The solution consists of a fully rubberized porous compound: a so-called "Poroelastic Road Surface" (PERS).

This was the main research topic of the EU FP7 project PERSUADE, where the application of poroelastic materials as pavement on different types of roads or local roads was of interest. The question was where to draw the line between porous and non-porous and between elastic and non-elastic in this case, and the following definition is proposed: "A poroelastic road surface (PERS) is a wearing course for roads with a very high content of interconnecting voids so as to facilitate the passage of air and water through it, while at the same time the surface is elastic due to the use of rubber (or other elastic products) as a main aggregate. The design air void content is at least 20% by volume and the design rubber content is at least 20% by weight." [7]

Prior to PERSUADE, there were several attempts to develop a poroelastic surface. Some trials in Japan and Sweden have shown dramatic reductions in vehicle noise of up to 10 dB. However, due to a number of unsolved problems, this new technology is not yet ready for application. Important aspects could not be investigated extensively enough to draw firm conclusions. The following real as well as potential problems need to be solved and PERSUADE has focused on them: Durability, wear and tear resistance, adhesion to the base, winter maintenance (salting, snow ploughing), mechanical behaviour (rutting, fatigue, bearing capacity) and the following aspects, which are among the most important, need to be clarified: Rolling resistance (to find solutions with positive impact on CO2 emissions), skid resistance, frost behaviour, fire risk, workability and production/laying processes including safety aspects for workers.

1.2. Poroelastic mixture optimization

An important task of the project was to optimise the composition and internal structure of poroelastic road surface material and to find suitable adhesive systems. Since poroelastic pavement is a relatively new and unexplored material, there is not yet a consensus on what should be considered functional laboratory testing methods for this type of material. Therefore, the laboratory test plan included a mix of standard basic property tests and accelerated tests that simulate loads and climatic conditions in a trafficked condition.

The target properties subjected to optimization were:

- Durability of the material in the field.
- Friction properties during service life and under different climatic conditions.
- Noise absorption properties and noise generating properties in contact with tyres.
- Drainage properties.
- Rolling resistance.
- Emissions of particles and hazardous substances into the environment.

A wide collection of different poroelastic road materials was prepared for further testing. These materials differed in composition and structure. Based on the results from easy to perform laboratory tests, the collection was narrowed down. More time-consuming and comprehensive laboratory tests were conducted to narrow the selection to a few candidates with good laboratory performance in terms of durability, noise reduction, friction properties, etc. For the best performing materials, dynamic tests as well as accelerated tests were conducted in full-scale pavement testers.

2. DESCRIPTION OF THE DYNAMIC SHEAR TESTS

The torsional triaxial cyclic tests were performed on the selected poroelastic material from rubber granules (recycled car tyers), stone aggregates, synthetic resin and chemical additives. From the chosen mixture the special hollow cylindrical samples were prepared as shown in the Figure 1. The dimensions of the samples are 100 mm outer diameter, 60 mm inner diameter and 200 mm height. Each of the samples was exposed to different combination of radial and axial stresses, which were constant during the tests. The variables between the performed tests were different values of radial and axial stresses and freaquency. All the tests that were presented in this paper were performed like strain controlled tests under isotropic stress condition. In the Table 1 the basic parameters for all four performed tests were presented.

Figure 1: Hollow cylindrical sample of the chosen mixture.

Table 1: Parameters of the tests

Sample	Axial stress, σ_A [kPa]	Radial stress, σ_R [kPa]	Frequency, ν [Hz]
I0_0.5_0	0	0	0.5
I0_0.5_1	0	0	0.5
I0_1.0_0	0	0	1.0
I245_0.5_0	245	245	0.5

2.1. Test procedure

To reach the target stresses the samples should be covered with rubber membrane in the hole and outside the sample. The sample was then installed in the triaxial cell and the whole cell was installed in the loading machine (Figure 2). The radial stress was created by water pressure in the triaxial cell and the vertical stress was generated by hydraulic system of the load machine. The difference between the water pressure in the triaxial cell and the presuure inside sample is the radial stress.

Figure 2: Hollow cylindrical apparatus.

After the sample was exposed to the intial stress state the loading procedure was started. For each sample the shear strain were raised between 0.001 % and 3 %, the samples I0_0.5_0 and I0_1.0_0 to maximum shear strain 8 % (Table 2). At each step of the apllied shear strain (γ) ten cycles were performed and dynamic shear parameters were determained in 5th and 10th cycle.

Table 2: Loading stages of the test

Stage	1	2	3	4	5
γ [%]	0.001	0.002	0.005	0.01	0.02
Stage	6	7	8	9	10
γ [%]	0.04	0.085	0.17	0.35	0.70
Stage	11	12	13	14	15
γ [%]	1.39	2.80	4.10	5.58	6.11
Stage	16	17			
γ [%]	6.79	7.89			

3. EXPERIMENTAL RESULTS

The main purpose of the experimental tests was determination of the shear properties of the poroelastic material to compare it with asphalt pavement shear properties. For the poroelastic material also hysteretic damping ratio and damping energy are important.

In this paper the results of shear modulus, hysteretic damping ratio and damping energy against shear strain are presented at different isotropic loading conditions corresponding two extreme points during car pass: minimum load (0 kPa of vertical stress) and maximum load (245 kPa of vertical stress).

In the Figure 3 the results of shear modulus against shear strain are presented. It can be observed that the frequency don't have any important influence on the value of shear modulus as well as the cycle number. It is true that higher frequency showed a little bit smaller values of shear modulus but the average difference is about 0.5 MPa, which is not high difference. The same difference also appears between the tests at same frequencies and same load conditions (I0_0.5_0 and I0_0.5_1). The main difference appears between the tests with different vertical load (I0_0.5_1 and I245_0.5_0) and it is more or less constant (about 4 MPa) from 0.01% of shear strain. From these results we can conclude that the poroelastic material is more or less elastic material in the zone of the predicted shear strains (up to 1% of shear strain).

The main influence on the material has vertical load, the higher is the higher is the shear modulus. The results under the 0.01% of the shear strain are on the limit of the accuracy of the machine, however the initial shear modulus are between 15 MPa for vertical stress 0 kPa and 20 MPa for vertical stress 245 kPa. The reduction in values of shear modulus at 0.01% shear strain and 1.0% shear strain is between 3 and 4 times. The Figure 4 shows hysteresis of all cycles at one stage and the Figure 5 shows for all stages from where is also visiable that material is elastic and that the number of cycles don't have any important influence.

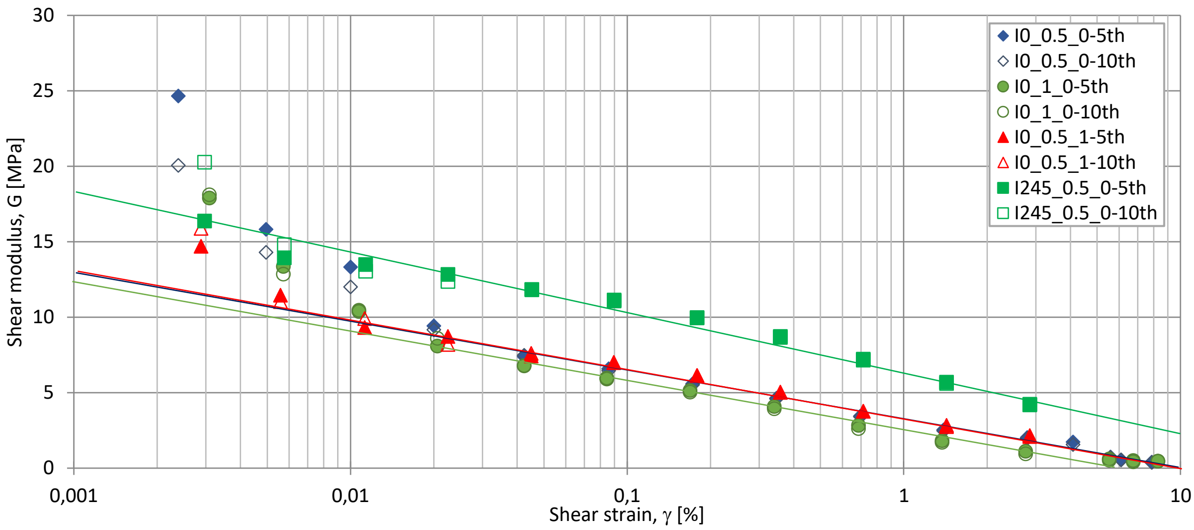


Figure 3: Shear modulus against shear strain.

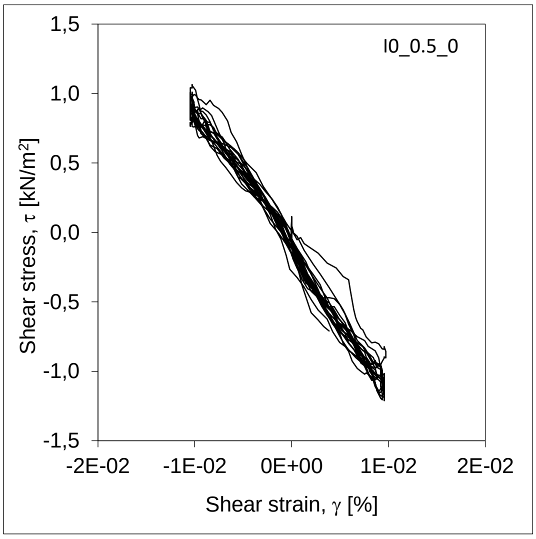


Figure 4: Hysteresis loop of one stage.

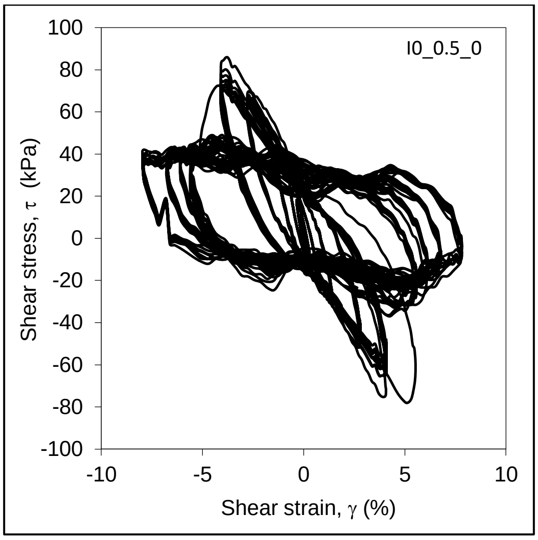


Figure 5: Hysteresis loop of all stages.

On the value of hysteretic damping ratio the vertical load doesn't have influence, which can be seen in the Figure 6. It is possible to see that frequence has small influence on hysteretic damping ratio, the values of hysteretic damping ratio are a little bit higher in case of 1 Hz frequency than in the case of 0.5 Hz frequency. It can be seen that the tests I0_0.5_0 and I0_1_0 show quite strange results but from the results of tests I0_0.5_1 and I245_0.5_0 the main observation can be explain.

The initial hysteretic damping ratio at shear strain of 0.001% is around 2% and lineary increase to value of 8.5% at 1.0% of shear strain. Increasing of the hysteretic damping ratio can be observed also from the shape of hysteresis loops at different stages (Figure 5), while with increasing shear strain loops become wider. This is also connected with damping energy that is shown in the Figure 7. While frequency on the damping energy doesn't have important influence, the influence can be observed in case of different vertical load. Higher vertical load means higher values of damping energy at the same shear strains. The difference starts at shear strain around 0.7% and increases with increasing values of shear strain. While shear modulus and hysteretic damping ration change linearly with increasing shear strains, damping energy increasies exponentially with increasing shear strain.

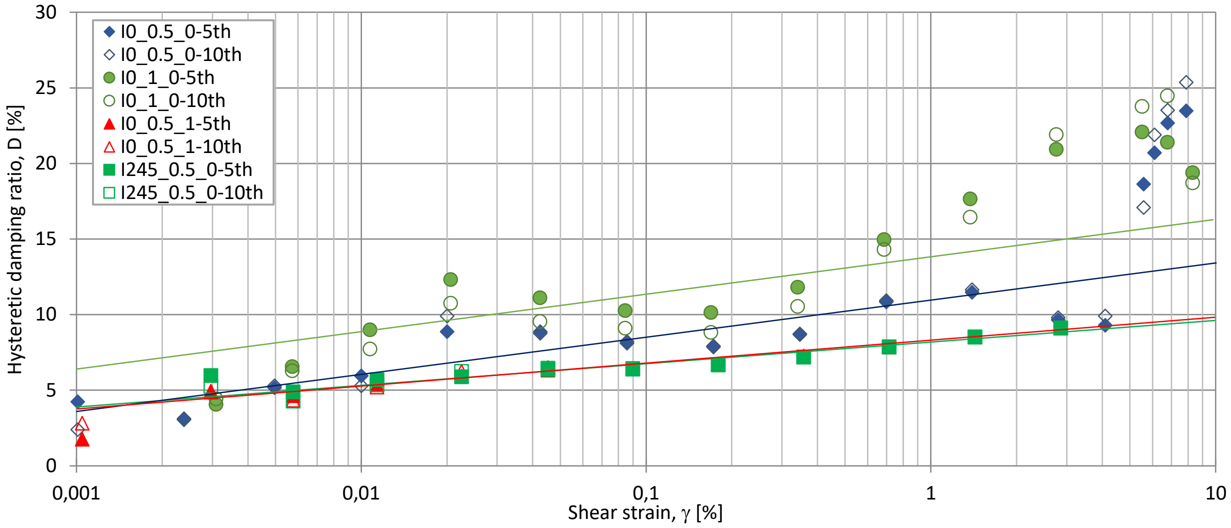


Figure 6: Hysteretic damping ratio against shear strain.

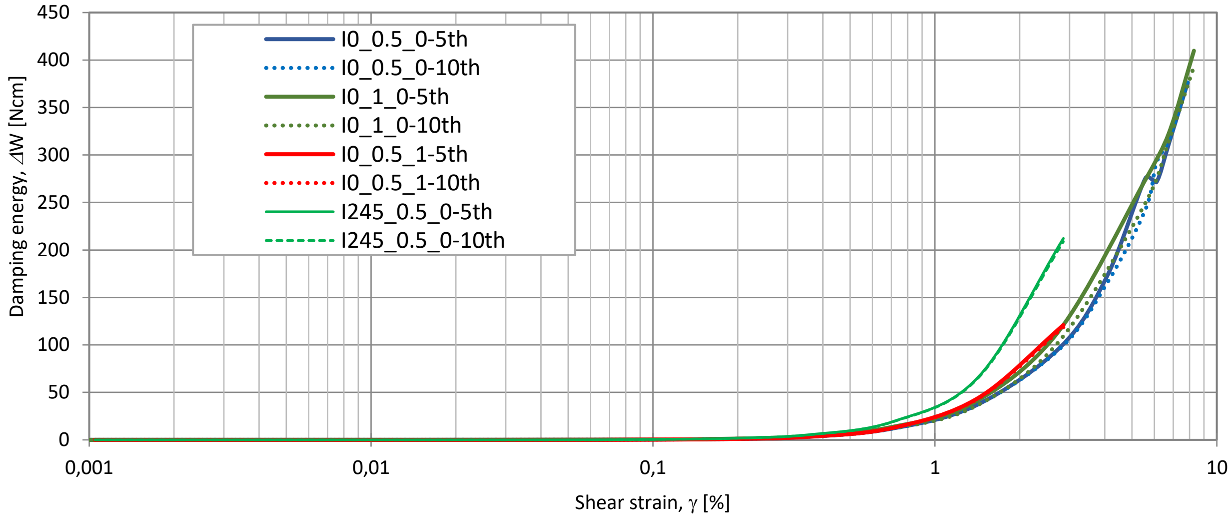


Figure 7: Damping energy against shear strain.

4. CONCLUSIONS

This paper describes results of cyclic torsional laboratory tests performed on poroelastic material developed during PERSUADE project. With performed tests the influence of frequency and vertical load was investigated at the stages of different shear strains. At each stage ten cycles at the same shear strain were performed. From all results that were valued from the measurments, changes in shear modulus values, hysteretic damping ratio and damping energy against shear strains are presented. The poroelastic material that was tested showed that material is elastic, even at shear strains higher then 1.0%. With increasing values of shear strains, values of shear modulus increase lineary, values of hysteretic damping ratio decrease linearly, while values of damping energy increase exponentaly. The initial shear modulus of tested poroelastic material is between 15 MPa and 20 MPa and is depending on the vertical load, increase of vertical load cause higher values of shear modulus. On the other hand, the damping energy is also higher with higher vertical load, but changes are observed from shear strains of 0.7%. Vertical load doesn't have influence on the hysteretic damping ratio. The values of hysteretic damping ratio are 2% at shear strain of 0.001% and increase up to 10% at shear strains of 1% and more.

The results of the tests carried out showed potential for further investigations simulating traffic loading and studying the changes in shear modulus, hysteretic damping ratio and damping energy as a function of cycle number at the same shear stress amplitude. The same test is to be performed on different mixtures of poroelastic material in order to develop poroelastic material with good mechanical and cyclic parameters as well as good noise damping properties.

REFERENCES

[1] Scrap Tires – Basic Information (2011). US Environmental Protection Agency.

[2] End of life tyres - A valuable resource with growing potential - 2010 edition. European Tyre & Rubber Manufacturers' Association.

[3] Tsang, H.-H. (2010). Uses of Scrap Rubber Tires, Ch. 12 in Rubber: Types, Properties and Uses.

[4] Council Directive 1999/31/EC of 26 April 1999 on the landfill of waste

[5] Council Decision of 19 December 2002 establishing criteria and procedures for the acceptance of waste at landfills pursuant to Article 16 of and Annex II to Directive 1999/31/EC

[6] Scrap Tires: Handbook on Recycling Applications and Management for the U.S. and Mexico, United States Environmental Protection Agency (EPA), 2010.

[7] Sandberg. U., Goubert. L., Biligiri K.P., Kalman B. (2010). State-of-the-Art regarding poroelastic road surfaces. Deliverable D8.1, Persude.

THE CLIMATIC IMPACT ON GERMANY´S TRANSPORT SYSTEM - INCREASING ROAD INFRASTRUCTURE RESILIENCE TO NATURAL HAZARDS

Anne-Farina Lohrengel
Lennart Meine
Federal Highway Research Institute (BASt), Department S1 – Adaptation to Climate Change
Brüderstraße 53, 51427 Bergisch Gladbach, Germany
lohrengel@bast.de, meine@bast.de

ABSTRACT

Climate change and the associated increase in extreme weather events present the road infrastructure with various challenges. Natural hazards such as storms, floods, gravitational mass movements and heat periods can cause damage to infrastructure and result in traffic restrictions and at worst, accidents can occur.

The Federal Highway Research Institute has already been dealing with the effects of climate change on the road transport infrastructure from 2011 to 2014. The aim of the so-called AdSVIS research program (adaptation of road transport infrastructure to climate change) was to identify relevant effects of climate change, to assess the vulnerability of individual objects of road transport infrastructure and to develop adaptation measures.

In 2016 the Federal Ministry of Transport and Digital Infrastructure formed the BMVI Network of Experts. The goal of topic 1 in this network is to develop the basics for increasing the resilience of the federal transport system to climate change and extreme weather events. Based on the analysis of climate effects on the transport system, targeted adaptation measures are taken developed or assessed, thus minimizing the negative impacts of climate change on the transport system.

When analysing the climate impacts on the transport system, the BMVI Network of Experts first funding phase (2016-2019) focused on the potential impact of floods, storms and gravitational mass movements as well as waterway-specific aspects of navigability (e.g. low water) and water quality.

Keywords:
climate change, infrastructure, climate impact assessment, natural hazards, resilience

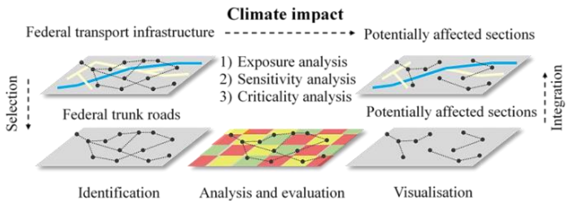
1. METHODOLOGICAL FRAMEWORK OF THE CLIMATE IMPACT ASSESMENT

In order to prepare the German transport infrastructure for climatic changes and, if necessary, adapt it accordingly, a climate impact assessment was developed by the BMVI Network of Experts.

In an effort to create a common structure and to generate comparable results on the effects of climate change on the different modes of the transport systems road, railway and waterway, a methodological framework was developed. This is divided into three sub-steps (Figure 1).

During 1) the exposure analysis, those areas of the federal trunk road network are localised which could potentially be exposed to the mentioned natural hazards. Subsequently, in 2) the sensitivity analysis, those sections of the network are identified which are particularly sensitive to the natural hazards due to their properties or equipment. Finally, 3) the criticality analysis helps to assess the relevance of the affected trunk road sections based on traffic, economic and ecological factors.

The three sub-steps of the climate impact assessment are in turn composed of many individual analyses per hazard. Part of the research is, among other things, to combine the individual analyses in a uniform and comparable framework in order to develop a model that is as real and applicable as possible. For this purpose, questions relevant to the federal trunk road network concerning climate change and extreme weather events are worked out and answered using the latest climate data from Germany's National Meteorological Service.



Source: adapted from Hänsel et al., 2020

Figure 1: Methodological framework of a climate impact assessment.

1.1. Exposure analysis

The exposure analysis examines the extent to which a system is exposed to certain climatic influences due to its spatial location and thus to the expected climate changes (Hänsel et al., 2020).

For the exposure analysis in particular, initial results have already been obtained on gravitational mass movements, river floods, flooding due to heavy rainfall events and potential impacts due to storm throw.

With the help of geographic information systems (GIS), models and map products were developed that allow the identification of exposed sections of the federal highway network. In addition to the models and maps mentioned above, first analysis tools for heavy rainfall events (a so-called "Blue Spot analysis") and the detection of vegetation occurrences along the road were developed. These allow variable application to different areas in the federal territory also with regard to the continuously updated input data.

1.2. Sensitivity analysis

The effects of climate change are not equally significant everywhere. There are regional differences in the degree to which transport or transport infrastructure in Germany is sensitive to changing climatic influences or resulting climate impacts. The crucial factors here are region-, route-, structure- or use-related parameters (e.g. structure design, construction methods, vehicle types etc.) and the resulting regional differences, for example in construction methods or environmental properties (e.g. geology, soils, topography, land use etc.) (Hänsel et al., 2020).

For the sensitivity analysis, initial possible vulnerabilities of the road infrastructure to the natural hazards mentioned have been considered so far. These include, for example, asphalt construction, which has an increased sensitivity to sustained heat effects.

1.3. Criticality analysis

Finally, the third component of the climate impact analysis represents the relevance of a road section. Important indicators for the relevance of the infrastructure segment can be traffic aspects such as freight volumes and values, traffic volumes or alternative route options.

As part of the criticality analysis, the traffic significance of sections of the federal trunk road network is assessed using different indicators based on transport and spatial planning parameters.

A main indicator for the criticality analysis is traffic volume, measured by average weekday traffic. Further indicators such as the share of heavy traffic over 3.5 tonnes as well as average travel distances can be derived from it and are also included in the analysis.

In addition, the levels of linkage function of transport axes between regionally and nationally significant centres are considered as an important indicator of relevance from a spatial planning perspective.

2. RESEARCH PROJECTS ON EXPOSURE ANALYSIS

Transport infrastructure is already affected by meteorological and hydrological influences today. While no robust climate change signals are available for assessing future exposure to storms, the future projections of the RCP8.5 scenario show an increase in potential exposure to flooding and gravitational mass movements (Hänsel et al., 2020).

In order to investigate these potential impacts on the federal trunk road network, various natural hazard-specific research projects have been carried out.

2.1. Floodings – Heavy rainfall

Future expected climatic changes could increase the vulnerability of roads to rainfall-induced floodings. In particular, the combination of prolonged drought and sudden, heavy rainfall in the summer months can lead to extreme runoff formation causing damage to roads and infrastructure. To estimate and assess the exposure of the federal trunk road network to potential flooding due to local heavy rainfall events, a Blue Spot analysis was carried out for the federal state of North-Rhine Westphalia. Thereby, sections of road where the probability of flooding is relatively high and the consequences of which are relevant to the transport system, the so-called Blue Spots, were calculated (Larsen et al., 2010a, Larsen et al., 2010b).

In the first step, hydrological modelling was carried out to calculate the runoff coefficients of three different precipitation scenarios for the project area, based on land use, relief and soil information using the Lutz method (Lutz, 1984). The runoff coefficient describes the infiltration or runoff potential with a dimensionless value of 0-1 and is an important input parameter for the subsequent hydraulic modelling. For each one hour precipitation scenario twelve runoff coefficient grids were calculated (one grid per

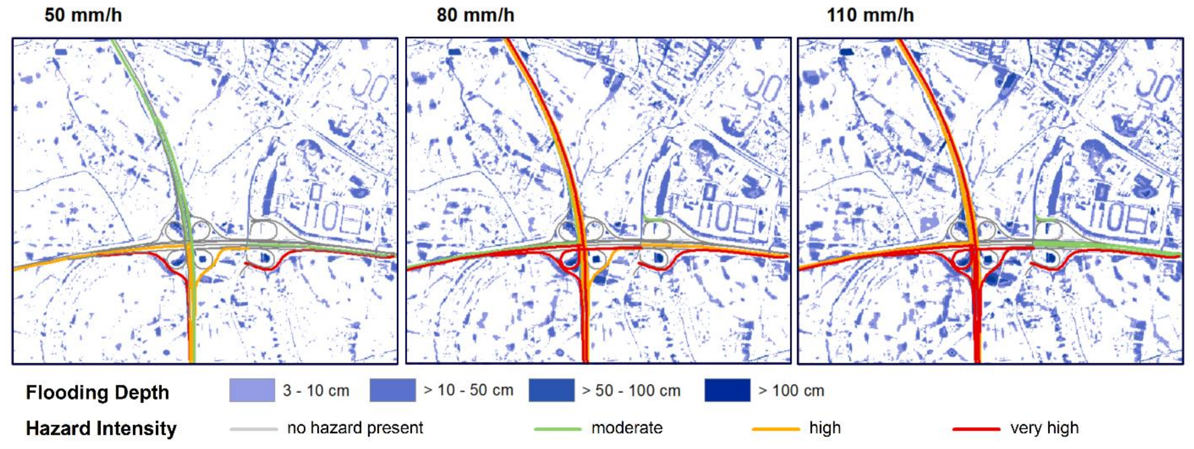
5-minute interval). The effective surface runoff per grid cell can thus be determined from the runoff coefficient and the precipitation quantity (Büche et al., 2020).

The hydraulic modelling enables the detailed derivation of surface runoff flow paths using a raster-based hydrodynamic 2D model, as well as the calculation of local flow velocities and flooding depths for different precipitation scenarios (50 mm/h, 80 mm/h, 110 mm/h).

The input variables are: the calculated runoff coefficient grids, the surface roughness derived from land use information according to the Gauckler-Manning-Strickler flow formula, information on the location and capacities of culverts or road drainage facilities and a high-resolution digital 3D laser scanning terrain model with a cell size of one meter (DTM1).

In order to save computing capacity, a hydroresampling is performed, which reduces the resolution of the DTM to five meters without losing information about hydraulically and hydrologically significant terrain properties. Information regarding the water networks, subordinate traffic routes and the federal trunk road network is transferred to the aggregated DTM5. Culverts, bridges, underground and surface water bodies are integrated into the DTM by means of pumping lines, that simulate a defined flow rate in order to be able to represent the flow paths as realistically as possible.

Finally, based on surface runoff flow paths, those flooding areas that overlap and influence parts of the federal trunk road network can be identified. These Blue Spots are provided as geo-referenced area features to which the sections of the federal trunk road network in their area of influence are assigned.



Source:adapted from Büche et al., 2020

Figure 2: Result of the Blue Spot analysis showing the potential hazard intensities for the trunk roads in the three scenarios: 50 mm/h, 80 mm/h and 110 mm/h. Blue areas show the maximum flood depths during the 3 hours of simulation.

The final risk analysis allows the evaluation of the potential impact that the floodings can have on the federal trunk road network based on different indicators. These indicators include the intensity resulting from the combination of flooding depth and flow velocity as well as the hypothetical occurrence frequency, depending on the necessary amount of precipitation causing the corresponding flooding (Figure 2). As a third indicator, the relevance of the affected road section is rated on the basis of the average weekday traffic and the proportion of heavy traffic.

The combination of all necessary methodological processing steps of the Blue Spot analysis in an ArcGIS toolbox enables the investigations to be efficiently applied to further study areas or other data sets. In addition, the implementation in a GIS enables the output of maps adapted to the needs of different users.

2.2. Gravitational mass movements

Gravitational mass movements can affect the traffic and the road infrastructure in various ways. For example obstacles on roads accidents can occur, which in turn lead to disabilities and delays in the flow of traffic result. Damage to the transport infrastructure is also possible and other impairments caused by repairs also have consequences.

The slope stability depends largely on the composition or nature of the subsoil and the slope. Different weather conditions and extreme weather events such as continuous precipitation, heavy precipitation or freeze-thaw cycles can represent triggering factors for gravitational mass movements (Krauter et al., 2012).

The exposure analysis for gravitational mass movements is mainly done by a landslide susceptibility model (Schipek & Kallmeier, 2019).

This was developed using a geotechnical knowledge based approach. It is initially carried out for a general approach that does not make any distinction between different mass movement processes. The geo-hazard potential is determined in the first step by combining slope angle classification and geo-technical rock classification in a geographic information system (Table 1).

Table 1: Knowledge-based classification scheme (decision tree) of the geo-hazard potential by combining rock classes and slope classes.

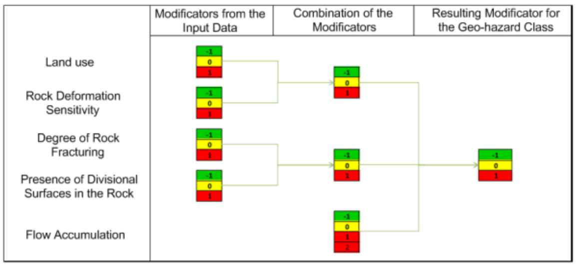
Geo-hazard Potential Class	Slope Angle Classification	Unconsolidated Rocks			Hard Rocks
		GK = 3; 7; 5 mixed-grained	GK = 4; 8 non-cohesive	GK = 2; 6 cohesive	GK = 1; 0
15					
14	DGM = 5	> 36°		> 30°	> 60°
13					
12					
11	DGM = 4	> 30 - 36°	> 36°	> 25 - 30°	> 50 - 60°
10					
9					
8	DGM = 3	> 25 - 30°	> 30 - 36°	> 10 - 25°	> 30 - 50°
7					
6					
5	DGM = 2	> 10 - 25°	> 25 - 30°	0 - 10°	0 - 30°
4					
3					
2	DGM = 1	0 - 10°	0 - 25°		
1					

Source: Schipek & Kallmeier, 2019

The initial classification result was further refined by including additional parameters like land use, rock deformation sensitivity, the degree of rock fracturing, the presence of divisional surfaces in the rock and flow accumulation (Figure 3). These modifiers were combined using an algorithm so that ultimately the main hazard classes either change by one class down (-1), remain neutral (0) or move upwards (+1).

In the final analysis step, a spatial link to the federal trunk road network was established for the potential hazard areas. For evaluation, the geo-hazard potential class ≥10 (high to very high hazard potential) were considered. In order to obtain an overview of the sections of the route, which are located next to and within a radius of 50 m, 100 m or 200 m around areas with a geo-hazard potential class ≥10, corresponding buffer zones have been created. Most recently, the road or rail network was intersected with these zones.

For this model approach of gravitational mass movements, a first method for integrating the climate influence was developed.



Source: Schipek & Kallmeier, 2019

Figure 3: Flow scheme (algorithm) for calculation of the resulting modification (specification) of the geo-hazard potential classes by combining the modifiers from the different controlling model input data.

In order to take into account the climatic influence, the climate parameters of: average annual precipitation, average summer precipitation, average winter precipitation, number of days with heavy precipitation events and number of days with freeze-thaw cycles were derived from climate projection data of the German Weather Service (DWD). Related to the IPCC´s Fifth Assessment Report, climate projections for an assumed RCP 8.5 scenario were generated for the "near future" (2031-2060) and "far future" (2071-2100). The generated climate data were also integrated into the present model via a modifier scheme.

The overall result shows that assuming the RCP 8.5 scenario, an increase in the number of federal highway sections potentially threatened by gravitational mass movements is expected (Lohrengel et al., 2020). According to the developed model, about 5% (2.590 km) of the federal highway network is already potentially located in areas with a geo-hazard potential class ≥10, taking into account the climate parameters. In the near future (2031-2060) around 6% (2.980 km) of the federal highway network could already be affected. In the far future (2071-2100), potentially vulnerable trunk roads may increase up to 7% (3.650 km).

Since the process sequences of gravitational mass movements and the resulting hazards have to be evaluated in a differentiated way, first models for the process-specific evaluation of potential impacts have also been developed. It was divided into fall (rock falls) and flow processes. In the current work, they are validated and further developed and the climate parameters are also considered in order to achieve a model of the potential for hazards that is as realistic as possible. The aim is to answer the question of which climatic thresholds can potentially trigger which type of mass movement processes.

2.3. Storms

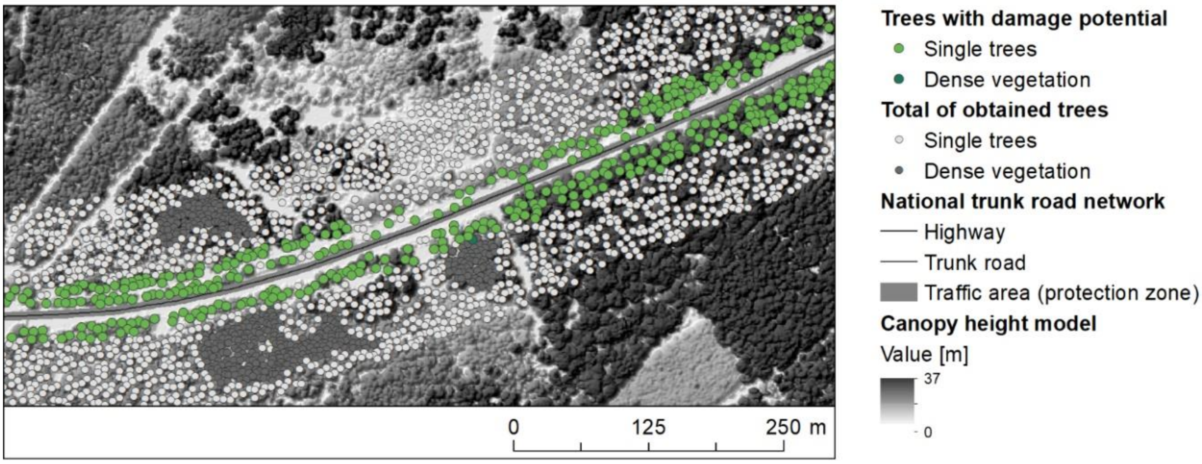
Storms are extreme weather events, that infrastructure operators and traffic are already facing with challenges today, even without taking the climate change into account (Hänsel et al., 2020). Frequently, bent or uprooted trees that fall onto the roadway pose accident hazards for road users or lead to temporary traffic interruptions.

For this reason, it was decided to carry out the exposure analysis of Federal trunk roads to the vegetation of Germany. In the first step, a GIS-based intersection of the roads and Vegetation is made with the Digital Base Landscape Model (Basis-DLM). It contains different types of forests, tree avenues and other woodlands.

As a result, it was found that overall about 25 % of the federal trunk road network pass through forested areas. In addition, about 13% of the federal highway network is exposed to tree avenues. It should be noted that the DLM only contains 2D data. Thus there is no statement about tree heights included. Because this is necessary to give a valid statement about the hazard potential, the research project "Detection of trees using laser scan data (LiDAR data) for exposure analysis along the federal trunk road network of North Rhine-Westphalia" was initiated by the BASt. The aim of this project was the development of a GIS-based method on the basis of free accessible high-resolution LiDAR data as well as RGBI orthoimage data that identifies and parameterises single trees (Steffen et al., 2020). The identification of trees was finally realised by applying both a minimum-curvature technique for single trees as well as a local maxima approach for dense vegetation, especially for coniferous trees. False classification results corresponding to non-vegetation areas have been corrected by the use of the Normalised Difference Vegetation Index. All tree candidates with potential to affect the road infrastructure and road users were retrieved and joined to the respective road sections to deliver a fast and effective method for analysation and visualisation of vulnerable parts of the trunk road network due to tumbling trees.

In order to improve the overall result, both an approach for single tree detection and a method for dense vegetation areas were combined for final application to obtain a more realistic image of the prevailing conditions in different environments (Steffen et al., 2020). The final detection rate achieved with the single tree detection model is assumed to be about 75% for deciduous and about 65% for coniferous trees. The study area was set for a 100 m wide buffer along the road. After the detection of the single trees as well as the dense vegetation, the tree candidates were identified, the height of which exceeds the distance to the road (Figure 4).

Finally, about 16% from the total of ca. 6,100,000 individual trees were classified as relevant for potentially affecting the road infrastructure when it comes to wind throw. In the final step, the number of relevant tree candidates was assigned to the federal highway network. In addition, an exposure map was generated that shows the number of relevant trees per kilometer.



Source: Steffen et al., 2020

Figure 4: Visualization of the single trees as well as the dense vegetation in a 100 m buffer around the federal trunk roads and identification of relevant tree candidates which have a hazard potential.

The results can be made available to infrastructure managers in future, for example to support a sustainable vegetation management concept. The advantage of a GIS-implemented toolbox is that it can be extended individually to other areas. Similarly, newly collected data can be continuously updated for a map visualisation if required.

3. VALIDATION AND FURTHER DEVELOPMENT

The products developed so far are primarily used for climate impact assessment in the BMVI Network of Experts. In addition, however, they could also to be seen as individual components which can be used at least as prototypes for the planning of the construction and operation of the road infrastructure. For the hazards: floodings (heavy rainfall), storms and gravitational mass movements the results are available in different levels of detail.

The results for floodings due to local heavy rainfall represent an innovative contribution to the exposure analysis with regard to possible floodings as a consequence of heavy rainfall events and thus contribute to the further development of the climate impact analysis. They are to be further incorporated into extended exposure analyses on the topic of "flash floods/floodings" across various modes of transport and throughout Germany.

The two tools developed, one for the identification of floodings ("Blue Spot-Tool"), as well as the one for the detection of trees using laser scan data ("LiDAR-Tool") were applied to the federal state of North Rhine-Westphalia. The reason for this is that North Rhine-Westphalia provides particularly high-resolution data bases free of charge, unlike most states in Germany. However, in order to be able to make a statement about the entire federal trunk road network, it would be important to extend the models to all federal states. A first step has already been taken with the application of the LiDAR-Tool to the federal state of Thuringia. The next procedure is to be clarified in coordination with the road construction authorities in order to be able to respond to their needs.

The LiDAR-Tool provided a basis to perform a differentiated exposure of the road to vegetation. But the determination of exposure of road infrastructure to windthrow should be accompanied by additional factors. For example, in addition to the occurrence of vegetation along the routes, its characteristics also play an important role. Other important aspects would be, for example, local wind speeds, tree species and their vitality, soil types and moisture.

Even though the validation shows a high agreement between documented flooding events and the model results of the Blue Spot analysis, the investigations indicate that exact information on the location and capacities of drainage facilities have a high influence on the quality of the hydraulic modelling. Therefore, it is important to obtain more accurate data on culverts, outlets and retention basins in order to further improve the modelling of hydraulic flood paths and potential flood areas in the vicinity of federal trunk roads.

In summary, the Blue Spot analysis, as well as the analysis for the detection of trees using laser scan data, are relatively highly differentiated analyses with high accuracy, even if there is further potential for improvement. However, as mentioned above, these analyses have only been applied to a part of the federal trunk road network. For the exposure analysis of gravitational mass movements, this is exactly the opposite case. Although the model developed was applied to the entire federal territory, a relatively coarse data basis was used. For this reason, ongoing projects are working on the validation and also further development of the model. On the one hand, detailed observations are to be carried out that can better reflect local geological conditions, and on the other hand, the process differentiation of gravitational mass movements is to be further elaborated.

Two projects have recently been started at the Federal Highway Research Institute for this purpose. The first deals with the "Identification of relevant climate parameters and determination of threshold values for gravitational mass movements in Germany including the development of a validation concept" (FE 01.202/2019/NRB). In the present model on general gravitational mass movements, a method was developed for the first time to make a nationwide statement on the influence of climate change on the federal trunk road network. In the first approach, a statistical approximation of the triggering climate parameters was carried out. Since this is only a very generalised approach, the determination of relevant threshold values must be dealt with in more detail. To this end, fundamental mechanisms of action between the basic disposition and the influence of climate or weather, such as the duration and intensity of precipitation or temperature changes such as freeze-thaw cycles, must be further investigated and better understood. So far, mainly the changes of climatic mean values over a period of 30 years have been investigated. For the new project, climatic data with a higher temporal resolution will be used, which will be related to local event data. Furthermore, it is intended to move away from a generalised approach for the whole of Germany. The different conditions of the subsoil (geology) in Germany should be taken into account. In connection with this, a differentiation according to process types (fall, flow and landslide processes) should be taken into account, as there are likely to be considerable differences in the climatic parameters for the triggering of a process.

The second project, which will contribute to the validation and further development of the disposition model, deals with the "Assessment of the stability of potentially landslide-prone slopes along the federal trunk road network" (FE 05.0208/2019/MRB). The aim of the project is to assess the stability of potentially unstable slopes along selected sections of the federal trunk road network. In addition, it is planned to develop impact scenarios related to climatic influences for the selected focus areas. The case studies help to analyse in detail the sections identified as "potentially affected" in the present model on general gravitational mass movements and to validate possible impacts.

Here, too, it is important to move away from a generalised approach and to take into account the regionally different subsurface conditions (geology) that exist in Germany. Furthermore, it is planned to investigate the different processes and forms of gravitational mass movements in a differentiated manner. These aspects must be linked to the local stability calculations in order to provide realistic interactions. In the process of the project, five representative areas are to be selected, for each of which the stability is to be calculated on the two to three most critical profiles. Subsequently, scenarios are to be developed that take into account extreme weather events and climatic changes in order to be able to derive corresponding damage patterns. This includes both short-term extreme rainfall events, which can e.g. trigger fast flow processes, and long-term periods of moisture, which can lead to extreme soil soaking. In order to assess the influence of climatic changes as a consequence of climate change the resulting impact scenarios are finally compared with the stability assessments under current conditions. In the final step, interviews with experts from the locally responsible road authorities will be carried out in order to obtain important information regarding previous traffic obstructions (type, duration) and corresponding damage patterns in the selected scenario and route section. In particular, estimates of costs and construction times from the local road authorities are also used.

In the first research phase of the BMVI Network of Experts, the focus of the work was strongly placed on the exposure analysis in order to generate a first overview of the sections of federal trunk roads that could potentially be affected by climate impacts at all. The second step of the climate impact assessment, the sensitivity analysis, was already considered in the first studies. It identified which route- or object-specific characteristics influence the vulnerability of the transport infrastructure to climatic impacts. The next step is to identify which adaptation measures can be used to reduce this vulnerability. In order to determine how sensitive the road infrastructure is, an enhanced understanding of the interaction between event and damage is essential. Data availability is often a limiting factor for the climate impact assessment. For this reason, existing event and damage data are being expanded and new data sources for infrastructure-specific route and object data are being acquired. The sensitivity of the road mostly relates to structural characteristics. In regulations and standards in the road sector, climatic influences are often already taken into account for the design of infrastructure elements. However, up to now there has been no collected overview of which regulations are climate-relevant and whether there is a need for adaptation if climate change is to be taken into account. An internal working group was formed in the BASt on this topic, which prepared a corresponding report for the BMVI (Federal Ministry of Transport and Digital Infrastructure).

The third step of the climate impact assessment, the criticality analysis was mainly limited to traffic volume related indicators in the first phase. In order to further extend these investigations, the first step in the second research phase will be to establish further indicators that are not related to traffic volume such as ecological, topological or other spatial planning indicators.

The fact that the different modes of transport are in many cases interlinked in the supply chain as well as in passenger transport shows the importance of also taking these relationships into account when assessing criticality. Therefore, a new research project will be carried out to analyse the cross-modal interrelationships of the federal transport system by identifying, for example, intermodal freight transport routes or interconnection nodes between rail, road and waterways. This should enable the derivation of inter-modal criticality indicators in order to understand the complex transport relationships.

Finally, the three sub-steps of the climate impact assessment are to be combined. This will provide an overview of which stretches of road are exposed to the various hazards, may be particularly vulnerable and are located in particularly highly frequented sectors of the federal trunk road network. This information can make an important contribution to prioritisation in the planning of adaptation measures.

The results of the climate impact assessment are to be integrated as a partial aspect into the overarching framework of the resilience management concept that is being developed within the BASt across departments. In a previous project, the framework was defined as follows: "Resilience management builds on a circular interplay of goal definition, identification of critical elements, hazard analysis, resilience screening and action identification, in combination with information transfer and periodic reviews provide the overarching framework, to ensure the resilience of a system in a conceptual and strategic way. In resilience management existing methods and management systems can be integrated." (Mayer et al. 2020). Accordingly, the methodology and results of the climate impact assessment should be used for being integrated into an overarching resilience management system.

In summary, the climate impact assessment is a multifunctional method to prepare road infrastructure adaptation for climate change and extreme weather events. It can be used as an integrated product, i.e., a combination of exposure, sensitivity, and criticality, to look at the transportation system as a whole, or only the individual analyses can be used to answer more specific questions.

REFERENCES

[1] Bott, F., Lohrengel, A.-F., Forbriger, M., Haller, M., Jensen, C., Löwe, P., Ganske, A., & Herrmann, C. (2020). Klimawirkungsanalyse des Bundesverkehrssystems im Kontext von Stürmen: Schlussbericht des Schwerpunktthemas Sturmgefahren (SP-104) im Themenfeld 1 des BMVI-Expertennetzwerks. https://doi.org/10.5675/ExpNBF2020.2020.05

[2] Büche, K., Assmann, A., Huber, B., Henrichs, T., & Pulgrabja, J. (2020). Lokale Analyse und Kartierung von potenziellen Überflutungsflächen im Bundesfernstraßennetz von Nordrhein-Westfalen in Folge von Starkregenereignissen (Blue-Spot-Analyse). Schlussbericht zum Forschungsprojekt FE-Nr. 01.0200/2017/NRB. Bergisch Gladbach: Bundesanstalt für Straßenwesen BASt (unpublished).

[3] Hänsel, S., Herrmann, C., Jochumsen, K., Klose, M., Nilson, E., Norpoth, M., Patzwahl, R., & Seiffert, R. (2020). Verkehr und Infrastruktur an Klimawandel und extreme Wetterereignisse anpassen: Ergebnisbericht des Themenfeldes 1 im BMVI-Expertennetzwerk für die Forschungsphase 2016 – 2019. Berlin: Bundesministerium für Verkehr und digitale Infrastruktur (BMVI).

[4] Krauter, E., Kumerics, C., Feuerbach, J., & Lauterbach, M. Abschätzung der Risiken von Hang- und Böschungsrutschungen durch die Zunahme von Extremwetterereignissen. Bericht zum Forschungsprojekt FE 89.238/2009/AP (Berichte der Bundesanstalt für Strassenwesen S 75). Bergisch Gladbach: Bundesanstalt für Straßenwesen.

[5] Larsen, M., Pihl, K. A., Hansson, K., Hellman, F., & Grauert, M. (2010a). The Blue Spot Concept: Methods to predict and handle flooding on highway systems in lowland areas (Report No. 181). Copenhagen: Danish Road Institute.

[6] Larsen, M., Pihl, K. A., Nielsen, N. H., & Rasmussen, S. F. (2010b). The Blue Spot Model: Development of a screening method to assess flood risk on national roads and highway system (Report No. 183). Copenhagen: Danish Road Institute.

[7] Lohrengel, A.-F., Brendel, C., Herrmann, C., Kirsten, J., Forbriger, M., Klose, M., & Stube, K. (2020). Klimawirkungsanalyse des Bundesverkehrssystems im Kontext gravitativer Massenbewegungen: Schlussbericht des Schwerpunktthemas Hangrutschungen (SP-105) im Themenfeld 1 des BMVI-Expertennetzwerks. https://doi.org/10.5675/ExpNLAF2020.2020.06

[8] Lutz, W. (1984). Berechnung von Hochwasserabflüssen unter Anwendung von Gebietskenngrößen. Mitteilung des Instituts für Hydrologie und Wasserwirtschaft (24).

[9] Mayer, G., Großmann, S., Zulauf, C., Roth, F., Deublein, M., Kohl, B., Kammerer, H., & Dahl, A. (2020). Resilienz der Straßenverkehrsinfrastruktur: Stand der Forschung und Potenziale im Management von außergewöhnlichen Ereignissen. Schlussbericht zum Forschungsprojekt FE-Nr. 01.0199/2017/ARB. Bergisch Gladbach: Bundesanstalt für Straßenwesen (unpublished).

[10] Nilson, E., Astor, B., Bergmann, L., Fischer, H., Fleischer, C., Haunert, G., & Helms, M. (2020). Beiträge zu einer verkehrsträgerüber-greifenden Klimawirkungsanalyse: Wasserstraßenspezifische Wirkungszusammenhänge: Schlussbericht des Schwerpunktthemas Schiffbarkeit und Wasserbeschaffenheit (SP-106) im Themenfeld 1 des BMVI-Expertennetzwerks. https://doi.org/10.5675/ExpNNE2020.2020.07

[11] Rauthe, M., Brendel, C., Helms, M., Lohrengel, A.-F., Meine, L., Nilson, E., Norpoth, M., Rasquin, C., Rudolph, E., Schade, N. H., Deutschländer, T., Forbriger, M., Fleischer, C., Ganske, A., Herrmann, C., Kirsten, J., Möller, J., & Seiffert, R. (2020). Klimawirkungsanalyse des Bundesverkehrssystems im Kontext Hochwasser: Schlussbericht des Schwerpunktthemas Hochwassergefahren (SP 103) im Themenfeld 1 des BMVI Expertennetzwerks. https://doi.org/10.5675/ExpNRM2020.2020.04

[12] Schipek, M., & Kallmeier, E. (2019). Validierung und Weiterentwicklung des Dispositionsmodells und der Hinweiskarte zu Hang- und Böschungsrutschungen für das Bundesfernstraßennetz. Schlussbericht zum Forschungsprojekt FE-Nr. 89.0338/2017/. Bergisch Gladbach: Bundesanstalt für Straßenwesen (unpublished).

[13] Steffen, M., Schipek, M., Lohrengel, A.-F., & Meine, L. (2020). Identification of windthrow-endangered infrastructure combining LiDAR-based tree extraction methods using GIS (Manuscript submitted for publication).

DYNAMIC MEASUREMENT SYSTEMS – USE OF FAST DRIVING MEASURING VEHICLES FOR THE NETWORK-WIDE RECORDING OF NIGHT VISIBILITY OF ROAD MARKINGS IN DRY CONDITIONS

Christian Mathea
Federal Highway Research Institute (BASt)
Bruederstraße 53, 51427 Bergisch Gladbach, Germany
mathea@bast.de

ABSTRACT

Because of their guiding function, road markings make a significant contribution to traffic safety. To guarantee their function, requirements like skid resistance and day and night visibility have been defined in European standard (EN 1436) and in further national regulations. Especially considering the fact that road markings are worn products, it can be seen that regular network-wide condition monitoring is required – also in connection with Connected and Automated Driving (CAD). A gapless measuring of the road markings is possible with dynamic measuring devices. The results of the dynamic surveys can be used as a basis for digital, systematic and focused maintenance management to reduce costs and increase road safety.

This paper provides an overview of current research in the use of dynamic measurement systems in Germany. At first the functionality of road markings is explained and then dynamic measuring devices for measuring night visibility are presented. Finally methods of answering the applicability of dynamic measuring devices and their potential to improve road safety are also discussed.

Keywords:
traffic safety, road markings, dynamic measurements systems, digital maintenance management

1. INTRODUCTION

In the past 30 years, the number of road fatalities in Germany has dropped sharply [1]. Most recently, 3,046 road deaths can be recorded (2019), which is the lowest level in 60 years [2]. The reason for this is that various measures have been successfully implemented. This includes, for example, a progressive development in the sector of vehicle technology, intensive driver training as well as improvements in road infrastructure. In order to continue the trend of decreasing the number of deaths in road traffic accidents and to achieve the goal of "Vision Zero" – no deaths in road traffic – , further factors that influence the road safety must be identified and corresponding measures have to be implemented [2].

Road markings, which serve to ensure and order flowing traffic, make a significant contribution to road safety. In order to realize the function of continuous visual guidance, road markings must be visible to road users at any times and under any conditions (such as night and wet conditions), and have sufficient skid resistance. The European standard (EN 1436) and other national regulations define requirements for traffic-related properties (skid resistance, day and night visibility) and describe the allowable measuring methods for recording and checking the individual parameters [3].

Both static and dynamic measuring methods allow checking the night visibility under dry conditions (R_L, dry). Corresponding measuring devices are available on the market. Although portable measuring devices can provide precise and punctual survey data, the low power of static measuring devices compared to fast-moving vehicle-mounted systems makes them unsuitable for a network-wide and continuous assessment of the quality of road markings. In addition, measurements with hand-held devices involve intervention in traffic and constitute a hazard for the measurement personal.

The research on the subject of dynamic measurement systems focuses on answering various questions with the aim of maximizing the potential of fast-moving measurement systems. As road markings are worn products and have to be constantly maintained and renewed according to their condition, the question arises, for example, how often they should be measured in order to achieve the highest possible road safety level. Furthermore, depending on the research results, conclusions could be drawn about the performance of road markings on the basis of verified measured values and cost assessments could be carried out for the life cycle of different road marking systems in order to further develop existing quality assurance systems. Thereby, the life cycle comprises a total of three phases:

1. investment phase:
   application of the road marking system
2. phase of use:
   use of the road marking system
3. deconstruction phase
   removal of the road marking system

In the context of the Connected and Automated Driving (CAD), it is also considered very likely that there will be a need for further knowledge on the net-wide condition detection of markings.

2. ROAD MARKINGS

The requirements, functionality and components of road markings are described as follows.

2.1. Requirements for road markings

Road markings must be clearly visible for the road users as a guidance device both during the day and at night and must have enough skid resistance, especially for motorcyclists. In order to ensure a minimum level of quality, EN 1436 has, among other things, standardised requirements for the so-called traffic-related properties of road markings as well as the related parameters at European level (see Table 1).

Table 1: Overview – Traffic- related properties and parameters

Traffic-related properties	Parameter / Unit
Day visibility	Luminance coefficient for diffuse lighting (Q_d) $$Q_d = \frac{L_V\,[mcd \times m^{-2}]}{E_V\,[lx]} \quad (1)$$ Lv = luminance Ev = illuminance
Night visibility, dry	Luminance coefficient for retroreflection (R_L) $$R_L = \frac{L_V\,[mcd \times m^{-2}]}{E_V\,[lx]} \quad (2)$$ Lv = luminance Ev = illuminance
Night visibility, wet	Luminance coefficient for retroreflection (R_L) $$R_L = \frac{L_V\,[mcd \times m^{-2}]}{E_V\,[lx]} \quad (3)$$ Lv = luminance Ev = illuminance
Skid resistance	[SRT-Value]

Source: The European standard EN 1436: Road marking materials – Road marking performance for road users and test methods

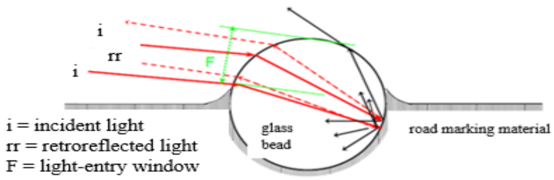
The individual parameters and the following allocation of defined visibility and skid resistance classes shall be determined by means of described measuring methods and test conditions. In addition, this standard also identifies the colour ranges of white and yellow road markings [3].

Based on the classes of EN 1436, minimum requirements for new and used conditions are specified in national regulations in Germany for temporary (yellow) and permanent (white) markings. Furthermore, the most different road marking systems must meet the standards of the relevant application fields. For example, a distinction is made between conventional markings (type I) with an even surface and markings with enhanced night-time visibility in wet conditions due to their drainage supporting surface structure (type II). Only type II markings can be used on highways [4]. In order to realize the highly recognizable value and instinctive traffic guidance, which contributes to road safety, the geometric characteristics of road markings are also regulated. Here, the geometry dependents on the respective design class or road category, such as highway or land road [5].

2.2. Functionality of road markings

With reference to the requirements described before, the functionalities of the road marking are explained below for each of the traffic-related properties.

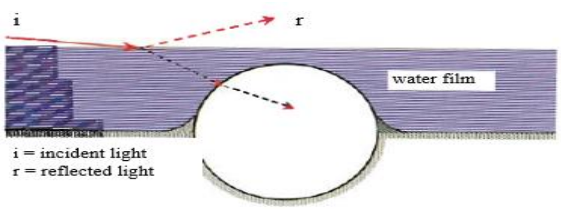
1. Day visibility:
   Day visibility is achieved by a positive contrast of the lane marking to the road surface. Defined colour ranges are specified for this purpose.
2. Night visibility, dry:
   Glass beads (reflective beads) embedded in the road marking are mainly responsible for the night visibility. With an ideal embedding of 60 % in the road marking material, the reflective elements break incoming light, for example the light of a vehicle lamp, which is reflected back to the light source in large parts after entering the road marking material. Non-broken light is often mirror-reflected, so the angle of entry is the same as the angle of exit. Figure 1 explains how retroreflection works.



Source: Drewes, John & Meseberg (2015). Handbuch und Kommentar für Markierungen auf Straßen. Bonn, Germany

Figure 1: Retroreflection – road marking glass bead

3. Night visibility, wet:
   Night visibility in wet conditions is achieved according to the same principle as night visibility in dry conditions. The main difference and the biggest technical challenge is a film of water on the road marking, for example when it rains. Especially with Type I road markings, the small glass beads often are under the water film, which reduces the principle of retroreflection and the water reflects the light of the vehicle lamp in the direction of travel. (see Figure 2). Therefore type I road markings are hard to see for road users in wet conditions.

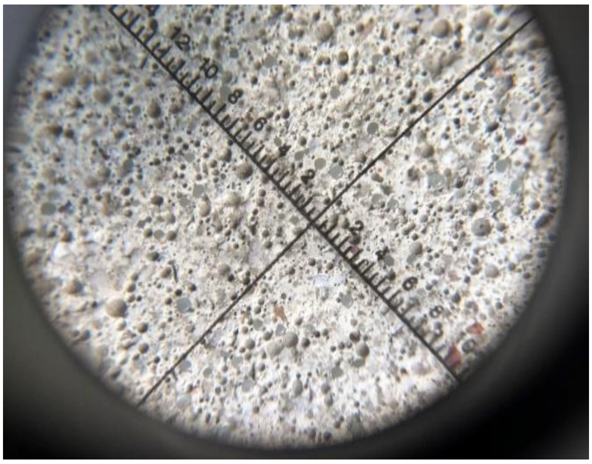


Source: Drewes, John & Meseberg (2015). Handbuch und Kommentar für Markierungen auf Straßen. Bonn, Germany

Figure 2: Reflection – road marking glass bead (Typ I)

The surface of type II road markings, in contrast, is shaped so that the reflective beads are mostly above the water film. This ensures a certain level of retroreflection. The type II effect can be achieved, for example, by profiling or agglomerate structure [6].

Skid resistance:
A sufficient level of skid resistance is given by skid resistance materials also embedded in the marking, which can be made of broken ceramics or glass, for example. Furthermore, calottes formed as a result of torn reflective beads can have a positive effect on the friction resistance and thus on the grip. (see Figure 3).



Source: Federal Highway Research Institute (BASt). Bergisch Gladbach, Germany

Figure 3: Macro photo – Calotte formation

2.3. Components of a road marking system

In addition to the requirements described for road markings, their components are also defined in European standards and national regulations. According to the definition, road markings consist of marking material and the corresponding additional substances, so it is called marking system. Basically, there are two types of marking systems: prefabricated marking systems and non-prefabricated marking systems. In the case of non-prefabricated marking systems, the individual components, such as the marking substance and the the corresponding additional substances (glass beads and/or skid resistance materials) are mixed together during the application [7].

3. DYNAMIC MEASURING SYSTEMS

For the measurement of night visibility, EN 1436 defines a standardised criteria to ensure the comparability of measured values. Regardless of the selected and permissible measurement method (static or dynamic), the so-called 30 m geometry must be used as the basis for determining the night visibility (R_L). In this case, the function of retroreflection described under 1.2 is assumed, according to which the point of illumination and the point of observation are located at almost identical positions (see Figure 4).

The illumination angle (ε) and observation angle (α) are determined according to the following equations:

$$\varepsilon = tan^{-1}\left(\frac{h_s}{e}\right) = tan^{-1}\left(\frac{0.65\ m}{30\ m}\right) = 1{,}24° \quad (4)$$

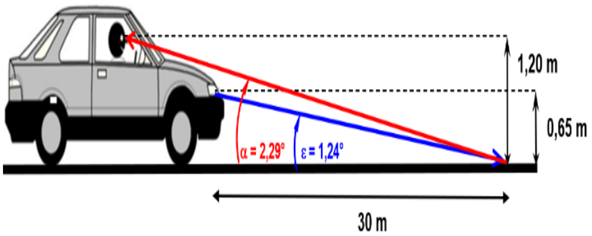
$$\alpha = tan^{-1}\left(\frac{h_d}{e}\right) = tan^{-1}\left(\frac{1.20\ m}{30\ m}\right) = 2{,}29° \quad (5)$$

With:

h_s = Vehicle headlight height

h_d = Eye level of the driver

e = Distance vehicle / road marking



Source: Federal Highway Research Institute (BASt). Bergisch Gladbach, Germany

Figure 4: 30 m Geometry

While static measuring devices are scaled down and detect the reflected light under shadowed conditions to calculate R_L, dynamic measuring devices must in principle meet the requirements of portable measuring devices even at high speeds. Influencing factors such as vehicle movements and changing lighting conditions must be considered [3].

3.1. Requirements for dynamic measuring systems

Because of the challenges of dynamic measuring, EN 1436 gives manufacturers of these devices specific freedom. For example, the distance between the measuring device and the measuring field is not fixed. There are no requirements for the size (length and width) of the measuring field. Furthermore, there are no criteria for positioning the dynamic measuring device on the vehicle. However, the maximum side angle of 0 to ±5° or 0±10° when measuring lane markings on the left and right sides of the measuring vehicle at the same time must be considered. In addition, the measuring instruments shall be designed to detect the range of values between 1 and 2,000 mcd×m-2×lx-1 [3].

Due to the special conditions of dynamic measuring, EN 1436 defines further specifications and special features, which will be explained below. It is allowed to implement various technical solutions, which are not further stated in the standard [3].

- Suppression of daylight signal:
  Compared with the portable measuring devices, when using dynamic measuring devices, with the help of a shielding of the measuring field, the ambient light can not be suppressed, and the ambient light will have an impact on the measuring results. Therefore, special technical equipment measures are required to suppress daylight [3].
- Suppression of changes of signal from tilts and shifts in height:
  Rutting, road curvature and changes in vehicle weight (for example, due to fuel consumption) will affect the expected measuring distance and thus the measured R_L-values. Therefore, measures must be taken to compensate for changes of inclination [3].
- Placement of measured field on the road marking surface:
  The constant positioning of the measuring field on the road marking surface is particularly challenging due to variable road widths and specific steering uncertainties. The user-friendliness of a measuring device depends on the width of the measuring field. The wider the measuring field, the greater the distances to objects, such as road restraint systems, which are located close to the road marking to be measured, can be selected during the drive. In addition, the width of the measuring field determines whether the full width of the road marking is detected or only part of the area. As there are no basic specifications for measuring field width in the standard (EN 1436), it must be specified by the device manufacturer [3].
- Longitudinal coverage of the road marking surface:
  The measuring frequency has a significant influence on the measured value recording. In conjunction with the length of the measuring field and the speed driven, the total measured length per time unit can be determined. Because the measuring frequency is constant and not linked to the speed of the measuring vehicle, the measurement is path-independent. The measuring frequency must be specified by the device manufacturer [3].
- Calibration of measuring equipment:
  The measuring device installed on the vehicle must be calibrated before starting the measuring procedure. The calibration can be performed by using a portable device that has been calibrated with a standard sample. Otherwise, the calibration of the measuring device can also be carried out directly with a standard sample [3].

3.2. Functionality of dynamic measuring systems

In order to solve specific environmental problems, EN 1436 gives manufacturers of dynamic measuring instruments various technical freedoms (see 2.1). For example, the illumination of the road marking required to detect the night visibility can be performed by different types (wavelengths) and illumination times. In practice, LED lamps, halogen lamps or xenon-gas-discharge-lamps are typically used. Various sensors, such as photodiodes, are used to detect and process the retroreflected light. The maximum allowable measuring speeds of systems available on the market vary from 100 to 150 km/h. Depending on the positioning of the measuring device on the vehicle (side or front), the dimensions as well as the distance of the measuring field to the vehicle will also change. The following describes the functionalities of two kinds of measuring devices available on the market as examples. Both of these devices meet the requirements of EN 1436. In addition to features in common such as geo-referenced recording, data structure and processing of the measured values and the possibility of photo or video documentation of the driving, the two measuring devices have significant technical differences [8][9].

3.2.1. ZDR 6020 (Company: Proceq SA)

The measuring device ZDR 6020 from Proceq SA will be installed on the side of the corresponding measuring vehicle (see Figure 5) and can be operated up to a speed of 150 km/h according to the manufacturer. The distance between the measuring field and the vehicle is 6 m. The measuring field is 1 m long and 0.5 m wide [8].

Source: Federal Highway Research Institute (BASt) Bergisch Gladbach, Germany

Figure 5: ZDR 6020

The road markings are illuminated by a halogen lamp, a lens optic and a pinhole disc rotating at a constant speed. The retroreflected light is sensed by 16 photodiodes. A software-supported synchronisation between the perforated disc and the photodiodes can eliminate ambient light. The synchronisation ensures that the measurement values are detected and recorded only when the perforated disc is open. At 600 Hz measuring frequency, up to 600 individual measurements can be recorded every second. The recorded single values can be averaged over selectable measuring blocks (between 10 m and infinity) [8].

3.2.2. Retro Tek-D (Company: Reflective Measurement Systems Ltd.)

The Retro Tek-D measuring device from Reflective Measurement Systems Ltd. is to be installed at the front of the vehicle (see Figure 6) and can be operated up to a speed of 120 km/h according to the manufacturer's specifications. The distance between the measuring field and the vehicle is 12 m. The measuring field is 3m in length and 5m in width. Due to the installation of the measuring device at the front of the vehicle, the Retro Tek-D can measure the left and right road markings at the same time compared with the ZDR 6020 [9].

Source: Federal Highway Research Institute (BASt) Bergisch Gladbach, Germany

Figure 6: Retro Tek-D

The road markings are illuminated by a green LED. Two cameras with green narrow-band filters capture the retroreflected light. The suppression of the ambient light is software supported. Therefore, changes of light intensity caused by activated or deactivated LED illumination are determined. With a measuring frequency of 20 Hz, the Retro Tek-D has a comparatively low measuring frequency, but a direct comparison with the ZDR 6020 is not very useful because of the different sensor technologies [9].

3.3. Factors of influence

In contrast to static measuring devices, it is quite difficult to measure the night visibility of road markings with dynamic devices (cf. 2.1 and 2.2). Within the practical investigations carried out in the past, it was possible to identify factors which influence the measurement results of dynamic surveys, which are listed in Table 2 as examples, differentiated according to constant and random influences. The influence has not yet been quantified [10][11].

Table 2: Factors of influence

Constant influences	Random influences
Curvature	Light conditions (e.g. low sun)
R_L-values	Humidity
Geometry (dotted, continuous)	Speed
Colour of the road marking	
Contrast between road surface an road marking (see Figure 7)	

Source: Goubert, & Lundkvist: Draft Report of the first round robin test for mobile reflectometers

Source: Federal Highway Research Institute (BASt) Bergisch Gladbach, Germany

Figure 7: Contrast – road marking / road surface

3.4. Potential of dynamic measuring systems

The replacement of road markings, for example in the USA, is usually carried out after a fixed time schedule depending on the traffic volume and the type of road markings. This can lead to the renewal of road markings that still meet the minimum requirements [12]. In this case, a complete measurement and network-wide assessment of the condition of road markings is not only important against the background of road safety, but can also contribute to the further development of a systematic maintenance management. For this purpose, analyses of the performance of road markings and life cycle studies can be carried out on the basis of network-wide night visibility values, in order to be able to carry out a targeted renewal of road markings. The efficient use of dynamic measuring systems to record night visibility could help road operators to optimise costs and guarantee an efficient infrastructure. Furthermore, a targeted use of resources is also desirable from an ecological point of view.

4. FIELD TEST – INVESTIGATION OF DYNAMIC MEASURING DEVICES

In order to investigate the influencing factors, determine possible use restrictions and evaluate the performance, field tests were carried out on the basis of previously developed measuring scenarios with the aim of creating a foundation for the usability of dynamic measuring systems, so that the night visibility of road markings can be detected economically and verified. The field tests were carried out in dry weather on motorways in Bremen (Germany) for three consecutive days. A total of three measuring device manufacturers participated in the field tests. Previously defined road sections for the field tests (for example, road sections with different types of road markings) were driven in a plantoon at given speeds (80 km/h and 120 km/h) to achieve almost identical measuring conditions. To ensure comparability of the measured values and to check the GPS accuracy of each measuring devices at the same time, it was planned to create reference points. For this purpose, highly retroreflective road marking tapes (> 1 000 mcd×m-2×lx-1) were used, which were temporarily applied. Because the selected road marking tapes have, especially when they are new, relatively high R_L values, certain peaks (high R_L values) should be visible during the evaluation of the raw data for the purpose of quality assurance. Ideally, single raw data could be verified by using the peak values. In addition, the defined measurement matrix is used for static measurements to quantify differences between static and dynamic devices and to establish their correlation.

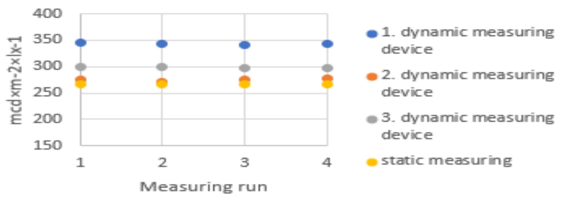
At this time, the evaluation of the measurement data related to standards and quality characteristics, such as accuracy (systematic errors), precision (random errors), uncertainty and repeatability, has not been completed due to pandemic-related delays. Nevertheless, the first investigation approach will be presented here as an example for a measuring section. Independent of the investigation and evaluation of possible influencing factors, firstly, the measuring results of the three dynamic measuring devices were compared, so that any deviation between each measuring device can be quantified. **Error! Reference source not found.**Table 3 gives an exemplary summary of the averaged measured values of the night visibility of each single measuring device in 4 diferrent measuring runs within a measuring section.

Table 3: Dynamic Measuring – comparison of the night visibility of a measuring section

Static reference value [mcd×m-2×lx-1]	Dynamic Measuring device	Night visibility [mcd×m-2×lx-1] / measuring run			
		1	2	3	4
280	1	345	344	340	343
	2	276	271	276	278
	3	299	299	298	297

Source: Federal Highway Research Institute (BASt) Bergisch Gladbach, Germany

The measuring values show that in this case study, the difference between individual measuring devices can reach as high as 20 %. Figure 8 visualises the measurement results.



Source: Federal Highway Research Institute (BASt) Bergisch Gladbach, Germany

Figure 8: Dynamic Measuring – comparison of the night visibility of a measuring section

Further detailed evaluations, results and knowledge of the field tests will be taken up in the further presentation.

5. CONCLUSION AND AN OUTLOOK

The use of dynamic measuring devices to detect the condition of road markings offers significant potential in terms of road safety. Furthermore, measurement data from dynamic measurement systems can contribute to the systematic maintenance management and cost savings. Measurement systems with a diverse range of technology are available on the market which are specially used to measure the night visibility of road markings. The prerequisite for developing potential is that, on the one hand, the night visibility of road markings can be reliably recorded by verified measuring devices, and on the other hand, the measurement data as a basis of road marking condition assessment can be correctly located and uniformly processed.

On the basis of the knowledge acquired in the practical investigations, recommendations for the use of dynamic measuring devices are to be developed in the further process of research, which, in addition to defined requirements for the recording, such as the length of aggregated individual value measurements (measurement blocks), also contain information about suitable measurement periods and measurement intervals. A desirable goal for the future is also the dynamic detection of other traffic-related properties of road markings, such as skid resistance, for which there are no adequate technical solutions yet. Creating standardised conditions and recommendations for the use of dynamic retroreflectometers, together with the progress of sensor technology and findings from research projects, could promote potential device manufacturers to develop and provide suitable measuring instruments.

REFERENCES

[1] Statistisches Bundesamt (2018). Unfallentwicklung auf deutschen Straßen. Wiesbaden, Germany

[2] Bundesministerium für Verkehr und digitale Infrastruktur (2020). Straßenverkehrssicherheit – Alles tun für #DeinLeben. Berlin, Germany

[3] European Committee for Standardisation (2018). The European standard EN 1436: Road marking materials – Road marking performance for road users and test methods

[4] Forschungsgesellschaft für Straßen- und Verkehrswesen (2013). Zusätzliche Technische Vertragsbedingungen und Richtlinien für Markierungen auf Straßen (ZTV M). Köln, Germany

[5] Forschungsgesellschaft für Straßen- und Verkehrswesen (1980, 1993 und 2020) Richtlinien für die Markierung von Straßen (Teil 1, Teil 2 und Teil A). Köln, Germany

[6] Drewes, John & Meseberg (2015). Handbuch und Kommentar für Markierungen auf Straßen. Bonn, Germany

[7] Forschungsgesellschaft für Straßen- und Verkehrswesen (2006). Technische Lieferbedingungen für Markierungsmaterialien (TL M). Köln, Germany

[8] Dr. Meseberg, H. (2009). Prüfzeugnis Nr. 0913-2009-05 über die Eignung des dynamischen Retroreflektometers ZDR 6020 zur dynamischen Messung des Leuchtdichtekoeffizienten bei Retroreflexion R_L von Fahrbahnmarkierungen. Berlin, Germany

[9] Dr. Meseberg, H. (2020). Prüfzeugnis Nr. 0913-2020-02 über die Eignung des dynamischen Retroreflektometers RetroTek-D Mobile Pavement Retroreflectometer zur dynamischen Messung des Leuchtdichtekoeffizienten bei Retroreflexion R_L von Fahrbahnmarkierungen. Berlin, Germany

[10] Goubert, L., & Lundkvist, S-O (2011). Draft Report of the first round robin test for mobile reflectometers. Belgium

[11] Lee, T. (2017). Advanced Methods for Mobile Retroreflectivity Measurement of Pavement Marking, Final Report. New Hampshire, USA: Transportation Research Board

[12] Lee, T. (2011). Advanced Methods for Mobile Retroreflectivity Measurement of Pavement Marking, Final Report for Highway IDEA Project 146. New Hampshire, USA: Transportation Research Board

THE APPROACH OF K-MEANS CLUSTERING AS A TOOL TO ANALYZE THE MULTIFUNCTIONAL ASSESSMENT TOOL FOR THE STRUCTURAL EVALUATION AND THE DESIGN OF PAVEMENTS (MESAS) DATA

Mahdi Rahimi Nahoujy, Dr.-Ing.
Federal Highway Research Institute
Brüderstraße 53, Bergisch Gladbach, Germany
rahimi@bast.de

ABSTRACT

Since 2018, the German Federal Highway Research Institute (BASt) has been using a Multifunctional assessment tool for the structural evaluation and the design of pavements (MESAS) to assess the structural condition of asphalt pavements on network level. MESAS can measure 1,500 km per week with a measuring point each 10 m, where more than 60 values such as bearing capacity, layer thickness, surface picture etc. are collected. This amounts to approx. one terabyte of raw data per week and can thus be called "Big Data". This results in the following problem: the data has to be evaluated, and quality controlled, but there is no established methodology yet for the analysis, especially for these huge amounts of data.

The objective of this study is to develop a new methodological approach for the analysis of the big data collected by MESAS and thus to generate knowledge from it. Data mining is a set of methods applicable to large and complex databases to discover hidden and noticeable (obvious) patterns in the data that were previously unknown. In this study, K-means clustering is used to model the MESAS data. This method is one of the popular unsupervised learning methods, where there is no label (index) for the dataset given. With the help of clustering, the data can be divided into different categories, e.g. "Excellent" or "Good" "Sufficient" and "Failure" or according to other indicators relevant in the assessment of the structural condition of pavements, providing a basis for further analysis and data usage.

Keywords:
Multifunctional assessment tool for the structural evaluation and the design of pavements (MESAS), Non-Destructive Tests (NDTs), Clustering, Machine learning, K-means

1. INTRODUCTION

It is essential to continually inspect and evaluate the pavement of all types of roads, e.g. freeways, highways, major roads, minor roads, etc., after they are built. The process of pavement evaluation follows some principles and goes through several stages; in each stage, collected data is recorded. Regular evaluations at appointed times are fundamental in every method of pavement maintenance; in fact, the data collected from these evaluations and inspections enable a pavement maintenance system to make the required plans for maintenance and rehabilitation. Inspection, similar to pavement maintenance, is initially done in networks, and after the prioritization of projects, the relevant information and details are accurately provided and recorded. There are various pavement evaluation methods encompassing a wide range of parameters; that is why different research centers are not able to employ a single comprehensive evaluation system regarding all types of roads.

Since 2018, the German Federal Highway Research Institute (BASt) has been using a Multifunctional assessment tool for the structural evaluation and the design of pavements (MESAS) to assess the structural condition of asphalt pavements on network level. MESAS can measure 1,500 km per week with a measuring point each 10 m, where more than 60 values such as bearing capacity, layer thickness, surface picture etc. are collected. This amounts to approx. one terabyte of raw data per week and can thus be called "Big Data". This results in the following problem: the data has to be evaluated, and quality controlled, but there is no established methodology for the analysis, in particular concerning slope data. For these huge amounts of data, the usage of methods from the field of data mining, e.g. machine learning is needed.

In general, there is a wide range of research available about NDTs and also about machine learning with Falling Weight Deflectometer (FWD) data, (Jansen, 2009; Chakar, 2011; Wacker, 2020; Rahimi Nahoujy, 2020). MESAS is a new device, though, and not well researched yet. Only a few studies exist about the device, focusing mainly on its general functionality and evaluating its usage for pavement structural testing (Muller et al., 2013; Katicha et al., 2020; Jia et al., 2021). Using machine learning with MESAS data is a fairly new approach with hardly any existing materials, but it is expected that more studies will be published soon.

In this study for the first time, an approach of unsupervised learning is used to model the slope data from MESAS. With the help of K-means, the slope data is divided into different clusters, providing basic labels for further analysis and data usage.

2. BACKGROUND

2.1. Non-Destructive Tests (NDTs)

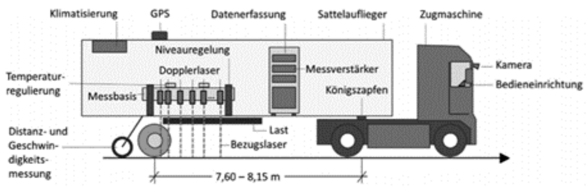
Non-Destructive Tests (NDTs) are employed at different levels to evaluate, repair, maintain, and rehabilitate various pavements. In a network, NDTs are used for pavements subdivided into sections with similar structural capacity. In a project, they are used to detect the location and reason for destructions in flexible pavements.

NDTs evaluating the structural condition use deflection measurement mechanisms. The deflection of pavement layers due to loading is a crucial index in determining the structural performance of the pavement and the strength of different layers. Currently, there are a variety of commercial pavement deflection measurement devices available. The most popular examples are the Benkelman Beam and the FWD.

Recently, NDTs have been further developed with the Traffic Speed Deflectometer (TSD). The TSD is a fast-moving measuring system for recording measured values that can be used to describe the load-bearing capacity of traffic surface pavements (FGSV, 2015). Since 2018, the German Federal Highway Research Institute (BASt) has been using a MESAS to assess the structural condition of asphalt pavements on network level. MESAS's basic function is the TSD high-speed load measurement system, and its additional measurement equipment such as transversal evenness and surface image, longitudinal evenness, ground-penetrating radar (GPR) and front cameras.

2.2. Bearing capacity measurements with TSD

The TSD originally comes from Denmark and was developed to measure deflections at high speed (Rasmussen, 2002). The measurement system with all associated instruments is installed in a truck (see Figure 1). In this way, the load-bearing capacity can be measured at a speed of approx. 80 km/h. To ensure equal general conditions when correcting the individual Doppler laser data, the measuring beam must be resistant to bending and torsion and, if possible, have the same temperature (FGSV, 2015).



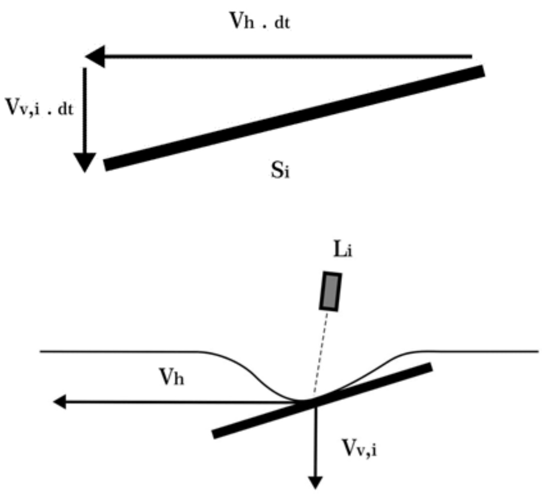
Source:(FGSV, 2015)

Figure 1: Principle sketch of the TSD

The measuring system consists of a towing vehicle with a single-axle semi-trailer in which the measuring device is mounted. The core of the measuring system consists of several sensors (Doppler lasers) arranged at defined intervals on a steel beam, which are aligned approximately perpendicular to the road surface. The lasers continuously record the speed of the short-term deformation of the traffic surface pavement in the direction of the sensor as a result of the load from the trailer axle. The vertical deformation speed vector is determined from the vectorial relationship between the measured deformation speed vector (in the direction of the sensor), the horizontal speed of the road in relation to the measuring vehicle (corresponding to the measured vehicle speed) and the angle to the vertical of the road surface. The Doppler laser Lref (reference laser), which measures in the area in front of the measuring system undisturbed by the load application, is used as a reference to compensate for the vertical vehicle dynamics and the torsion of the measuring beam. This means that the calculation does not consider the actual angle of the Doppler lasers, but the (constant) angular difference to the reference laser. for each sensor position i. The so-called slope value (slope of the deformation at this sensor position) is derived from the corrected vertical speed vector Vv,i and the horizontal speed of the measurement vehicle Vh as follows (Hildebrand et al., 2000; FGSV, 2015):

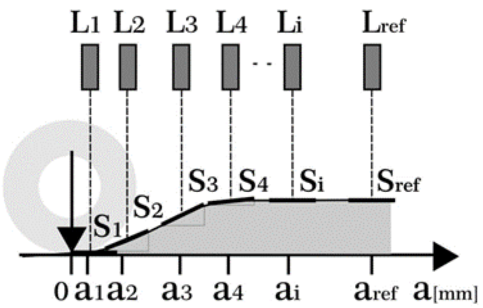
$$Slopei \; = \; Si \; = (Vv, i) \; / \; Vh \qquad (1)$$

The principle of calculating the slope value is shown in figures 2 and 3.



Source:(FGSV, 2015)

Figure 2: Principle sketch for calculating the slope value in a vehicle-related coordinate system (not to scale)



Source:(FGSV 433 B 5, 2015)

Figure 3: Arrangement of the Doppler lasers to determine the slope values (not to scale)

The result files of a TSD measurement contain about 60 different values that are indirectly and directly related to the road condition or the measurement itself. Among them, in the case of BASt's TSD, there are ten different slopes (in distances: -450 (mm), -300 (mm), -200 (mm), 110 (mm), 210 (mm), 310 (mm), 460 (mm), 610 (mm), 910 (mm), 1510 (mm)) and usually at least nine different deflections (in distances: D0 (mm), D200 (mm), D300 (mm), D450 (mm), D600 (mm), D750 (mm), D900 (mm), D1200 (mm), D1500 (mm)).

2.3. Analysis of slope values

As mentioned, 10 slopes are measured by the TSD in each measurement. Since there does not exist any indicator yet to classify the different slopes, the deflection values have to be calculated from them in order to assess the quality of the pavement. But these calculations have error due to curve fitting methods. Either one has to use deflections to evaluate the pavement with some error, or try to find labels for slopes. In this study a method of cluster analysis for slope data is presented to determine labels for TSD data. K-means as one of the popular unsupervised learning methods is used to train a model at network level for slope data.

3. METHODOLOGY

3.1. The method of machine learning

Finding patterns in existing data is only possible for the human brain. But when the volume of data becomes very large and the time required to perform calculations increases, the need for machine learning is introduced as a science that helps people work with massive amounts data in the least amount of time.

The core of the science of machine learning is to design machines that use the examples given to them and learn from their own experiences. In fact, this science tries to create a machine using algorithms in such a way that it can operate without needing plans or prescribed individual actions as input, but that it can learn from itself. In machine learning, instead of programming everything, the data is given to a general algorithm, and it is this algorithm that builds its own logic based on the data given to it. Machine learning has a variety of methods, including supervised, unsupervised, and reinforcement learning. The algorithms used in machine learning fall into these three categories.

3.2. Unsupervised learning

Unsupervised learning is a set of machine learning methods for discovering patterns in data. In supervised learning, the system tries to learn from the previous examples given to it. In unsupervised learning, the data processed with the unsupervised algorithm is non-labeled, which means that the input variable (X) is specified without any corresponding output variable. In unsupervised learning, algorithms are used to discover interesting pattern in the data. Therefore, if the dataset is labeled, it is considered a supervised problem, and if the dataset is unlabeled, the problem is unsupervised. Figure 4 shows an example of supervised and unsupervised learning. The image on the left is an example of supervised learning, in which the regression method is used to find the best fit between the features. In unsupervised learning as on the right, inputs are separated based on characteristics and predictions are based on which cluster they belong to.

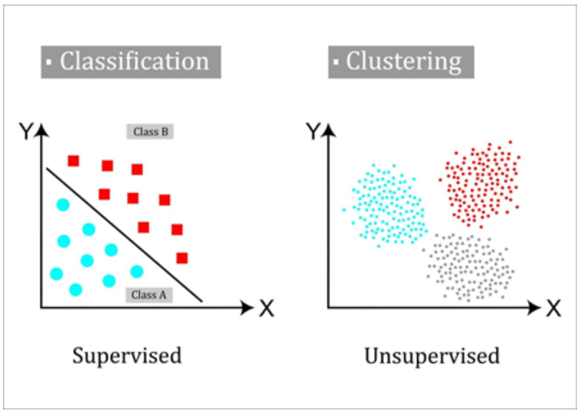


Figure 4: Supervised and unsupervised learning

"Clustering" is the most important method in unsupervised learning.

3.3. Clustering

Cluster analysis, or clustering for short, is the process by which a set of objects can be divided into individual groups. Each division is called a cluster. The members of each cluster are very similar in their characteristics, while the degree of similarity between the clusters is minimal. In such a case, the purpose of clustering is to assign "labels" to the objects, indicating that each object is a member of the cluster.

Although most algorithms or clustering methods have the same basis, there are differences in how similarities or distances are measured, as well as the choice of labels for the objects in each cluster. The main groups for clustering methods are: partitioning methods, hierarchical clustering, density-based clustering, model-based clustering and "fuzzy" clustering.

3.4. K-means

The K-means algorithm is one of the most popular partitioning methods used in machine learning, especially in the field of unsupervised learning. In K-means clustering, the optimization of an objective function is used. The clustering results in this method can be performed using minimization or maximization of the objective function. If the criterion is "distance measure" between objects, the objective function is based on minimization. The answer to the clustering operation is to find clusters where the distance between objects in each cluster is minimal. Conversely, when the dissimilarity function is used to measure the similarity of objects, the objective function is chosen so that the clustering results maximize its value in each cluster.

The K-means clustering method was invented by sociologist and mathematician McQueen in 1965 and developed and optimized by other scientists. In 1957, another version of this algorithm, the standard K-means clustering algorithm, was developed by Lloyd. This version of the clustering algorithm is now standard on most computer software that performs K-means clustering.

3.5. Lloyd algorithm

Suppose that the observations (x1, x2, ..., xn), which have a dimension D, must be divided into k clusters. These clusters are identified as S = {S1, S2, ..., Sk}. The cluster members should be selected from observations that minimize the "within-cluster sum of squares" (WCSS) function. Therefore, the objective function in this algorithm is described as follows (Arthur et al.,2006):

$$arg_S\, min \sum_{i=1 \times \in S_i}^{k} \sum \|X - \mu_i\|^2 = arg_S\, min \sum_{i=1}^{k} |S_i|\, Var\, S_i \qquad (2)$$

where:

µi: the center of clusters Si

| Si |: the number of members in cluster i.

3.6. Training a K-means model

In this study, the K-means algorithm uses an iterative refinement technique that works in the following way (Arthur et al., 2006):

1. First, K centers (mean of the cluster) are quantified randomly, which represent the clusters: $\mu_1^0, \mu_2^0, \ldots, \mu_k^0$
2. Each observation is assigned to the cluster with the closest mean value. In this work, the "squared Euclidean distance" is used as the K-means algorithm:
   $$S_i^t = \left\{x_p : \|x_p - \mu_i^t\|^2 \le \left\|x_p - \mu_j^t\right\|^2 \; \forall j, 1 \le j \le k\right\} \qquad (3)$$
3. the mean values for observations associated with each center are recalculated:
   $$m_i^{t+1} = \frac{1}{|S_i^t|} \sum_{x_j \in S_i^t} x_j \qquad (4)$$
   Then, these two steps (2 and 3) are performed alternately several times so that the means are stable enough or the sum of the variances of the clusters does not change much.
4. Finally, the following mean values of the last phase (in time T) will represent the clusters: $\mu_1^T, \mu_2^T, \ldots, \mu_k^T$

4. RESULTS AND DISCUSSION

4.1. Results

In this study, data measured with the TSD of the Federal Highway Research Institute (BASt) on a highway in North Rhine-Westphalia was used. The measurements were processed every 10 m on more than 52 km of highway with 3 strips in both directions, so that after preprocessing 41,618 data sets for slope 110 (which is the slope value at a distance of 110 mm from the axle load) are available (Figure 5).

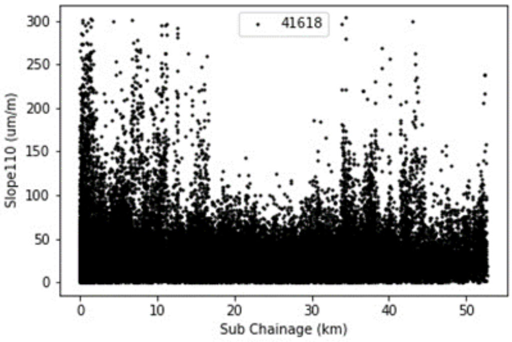


Figure 5: slope 110 data after preprocessing

First, the slope 110 values are divided into four clusters by K-means (one-dimensional clusters). Since the results of clustering could be different in each running, some models were trained and their deviations of results were analyzed. The deviation of averages and limits of the models were very low (less than 1%), showing that the clusters are robust. Four clusters are shown in the following Figure 6.

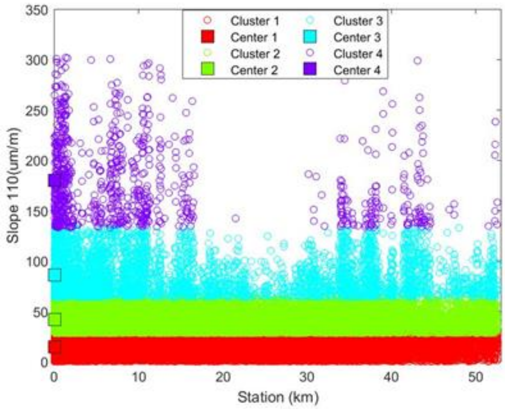


Figure 6: Cluster diagram for slope 110

The center values in this model are positioned as follows:

- Center 1: 14.3641437039793 µm/m
- Center 2: 41.9321957198163 µm/m
- Center 3: 86.2550691677927 µm/m
- Center 4: 179.975589050958 µm/m

All points in each cluster have the smallest distance from the respective center, compared to the other centers. This model results in three boundaries (labels):

- Label 1: ca. 28.15 µm/m
- Label 2: ca. 64.10 µm/m
- Label 3: ca. 133.10 µm/m

By applying these labels, all data is divided into four groups (clusters). The data of each cluster have a special relationship to each other. Based on the existing background of experience in connection with bearing capacity measurements, the TSD data (here: slope 110 values) is then categorized as follows:

- Excellent: 0 – 28.15 µm/m
- Good: 28.15– 64.10 µm/m
- Sufficient: 64.10– 133.10 µm/m
- Failure: more than 133.10 µm/m

The next model is trained for slope 310. In this route there are 40,787 data sets after preprocessing for slope 310 available (Figure 7). The slope 310 data is clustered into four clusters as Figure 8. The center values in this model are positioned as follows:

- Center 1: 12.3095985112341 µm/m
- Center 2: 33.9514330427490 µm/m
- Center 3: 62.5570365505194 µm/m
- Center 4: 106.351718880242 µm/m

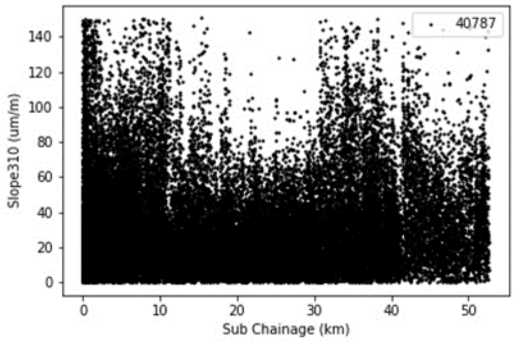


Figure 7: slope 310 data after preprocessing

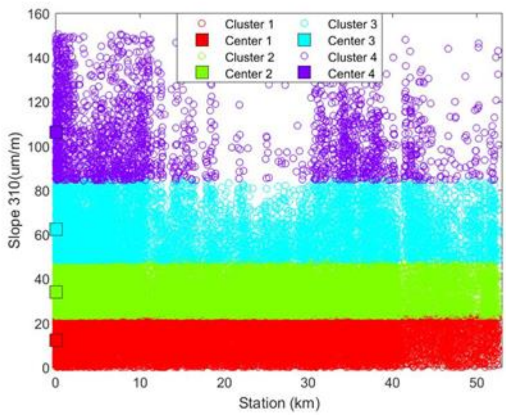


Figure 8: Cluster diagram for slope 310

This model results in three boundaries (labels):

- Label 1: ca. 23.12 µm/m

- Label 2: ca. 48.25 µm/m
- Label 3: ca. 84.46 µm/m

Using these labels, all data is clustered into four groups. The slope 310 data is then categorized as follows:

- Excellent: 0 – 23.12 µm/m
- Good: 23.12– 48.25 µm/m
- Sufficient: 48.25– 84.46 µm/m
- Failure: more than 84.46 µm/m

4.2. Discussion

In this study two different indicators are presented. For slope 110 and for slope 310, both as one-dimensional clusters, the data is divided into four clusters with the labels “Excellent”,“Good”, “Sufficient” and “Failure”.

Since the results of clustering could be different in each running, some models were trained and their deviations were analyzed. The deviation of averages and limits of the models were very low (less than 1%), showing that the clusters are robust.

After the clusters have been positioned by labels, new measured data can be categorized with these labels and it can be determined in which cluster they are located.

Up to now, there has been no existing indicator for slope data, because road agencies and companies have been using deflection values, that have to be calculated from the raw data, in order to assess the quality of the pavement. The presented approach of clustering slope data provides an alternative, that does not need the extra-step of calculating deflection values, which may contain errors. Future research may focus on determining the quantity of error of the calculations of deflections and compare it with the approach presented in this study.

At this point, it is explicitly noted that the boundaries, categories and labels presented are derived from a purely data-based approach and therefore have no general validity and are highly dependent on the data basis.

5. SUMMARY

Non-Destructive Tests (NDTs) are used at different levels to evaluate different pavements in terms of their condition and to derive maintenance needs. The Multifunctional assessment tool for the structural evaluation and the design of pavements (MESAS) is a new device, which can assess the structural condition of asphalt pavements on network level. The Traffic Speed Deflectometer (TSD) is the core part of MESAS and records slope values.

In this paper, an approach of machine learning in road construction was presented. TSD data was modeled by using K-means as an unsupervised learning method. TSD slopes (110 and 310) were modeled in one-dimensional clusters. For each model, four different clusters were defined with certain boundaries as labels (excellent, good, sufficient, failure). The deviation of averages and boundaries of the models were very low at less than 1%.

The clusters thus determined can be used to classify new data sets. That means, for new datasets, the labels function as indicators to categorize the different slope values, enhancing the assessment of the quality of pavement.

When using purely data-based evaluations and classifications, it is nevertheless recommended to check or classify the results using, for example, mechanical models.

REFERENCES

[1] Arthur, D., & Vassilvitskii, S., 2006. k-means++: The advantages of careful seeding. Stanford.

[2] Chakar, T. (2011). Methode zur Klassifizierung von Tragfähigkeitsmessergebnissen des Falling Weight Deflectometers bei Asphaltbefestigungen (Vol. 59). KIT Scientific Publishing.

[3] FGSV, 2015. Arbeitspapier- Tragfähigkeit für Verkehrsflächenbefestigungen Teil B 5. Schnell fahrendes Messsystem – Traffic Speed Deflectometer (TSD): Gerätebeschreibung, Messdurchführung. FGSV-Verl. Köln, 2015, FGSV 433 B 5.

[4] Hildebrand, G., Rasmussen, S. & Andrés, R., 2000. Development of a laser-based high speed deflectograph. In S. D. Tayabji & E. O. Lukanen (Eds.), Nondestructive testing of pavements and backcalculation of moduli. Proceedings of the "Third International Symposium Nondestructive Testing of Pavements and Backcalculation of Moduli", held Seattle, Washington, on 30 June - 1 July 1999 (ASTM STP, Bd. 1375, S. 457–469). Fredericksberg: American Society for Testing and Materials (ASTM).

[5] Jansen, D. (2009). Temperaturkorrektur von mit dem Falling-Weight-Deflectometer gemessenen Deflexion auf Asphaltbefestigungen. schriftenreihe des instituts fuer strassenbau und verkehrswesen, (2).

[6] Jia, X., Woods, M., Gong, H., Zhu, D., Hu, W., & Huang, B. Evaluation of influence of pavement data on measurement of deflection on asphalt surfaced pavements utilizing traffic speed deflection device. Construction and Building Materials, 270, 121842.

[7] Katicha, S. W., Shrestha, S., Flintsch, G. W., & Diefenderfer, B. K. (2020). Network Level

Pavement Structural Testing With the Traffic Speed Deflectometer (No. VTRC 21-R4).

[8] Muller, W. B., & Roberts, J. (2013). Revised approach to assessing traffic speed deflectometer data and field validation of deflection bowl predictions. International Journal of Pavement Engineering, 14(4), 388-402.

[9] Rahimi Nahoujy, M., 2020. An Artificial Neural Network approach to model and predict asphalt deflections as a complement to experimental measurements by Falling Weight Deflectometer. Doctoral dissertation, Ruhr-Universität Bochum, Germany.

[10] Rasmussen, S., 2002. Development of a high speed deflectograph (Report 117). Roskilde: Danish Road Institute.

[11] USDOT, F., 2011. Use of nondestructive testing in the evaluation of airport pavements. Advisory Circular, 150.

[12] Wacker, B. (2020). Zeitraffende Belastungsversuche mit integriertem Einsatz zerstörungsfreier Messverfahren.

A ROUTING PERFORMANCE COMPARISON OF LTE-V2V AND IEEE WAVE FOR EMERGENCY VEHICLE APPROACHING WARNING APPLICATION IN VEHICULAR NETWORKS

Tibor Petrov
Department of International Research Projects – ERAdiate+, University of Zilina
Univerzitna 8215, Zilina, Slovakia
tibor.petrov@uniza.sk

ABSTRACT

Applications based on vehicular communications are becoming an important part of the Intelligent Transportation Systems. They are expected to bring many benefits including increased road safety, reduced environmental impact and decreased socio-economic costs of road transport. Over the years, two dominant communication technologies for vehicular communications have emerged – Dedicated Short-Range Communications (DSRC) and Cellular Vehicle-To-Everything (C-V2X). While many studies investigated the performance of both technologies for specific vehicular applications relying on beaconing, some applications require a network level routing of the communication packets for their operation. In order to assess the routing performance of DSRC and LTE-based C-V2X communication, we prepared a federated traffic-telco simulation study using a grid road network. The vehicles in the simulation run a Privacy Ensuring Emergency Vehicle Approaching Warning application using IEEE WAVE communication standard in the first simulation scenario and LTE-V2V in the second simulation scenario. The unicast performance of both technologies is evaluated in terms of average message end-to-end delay and packet delivery ratio.

Keywords:
V2V, VANET, IEEE WAVE, LTE-V2V, emergency vehicle, warning, routing, simulation

1. INTRODUCTION

Applications based on vehicular communications are becoming an important part of the Intelligent Transportation Systems. They are expected to bring many benefits including increased road safety, reduced environmental impact and decrease of socio-economic costs of road transport. While many communication technologies have been considered for intervehicular communications, two of them have emerged as dominant over the years - Dedicated Short-Range Communications (DSRC) and Cellular Vehicle-To-Everything (C-V2X). In literature, DSRC communication is often used as an overarching term for technologies based on IEEE 802.11 OCB [1] as an Access layer standard of the Station Reference Architecture [2,3]. Two major families of standards based on IEEE 802.11 OCB include ETSI ITS [2] developed for deployment in the European Union and IEEE WAVE [3] to be used in North America and other countries. Communicating nodes using either of these technologies form a Vehicular Ad hoc Network (VANET) – a communication network without any centralized management where nodes are free to enter or leave the network without any prior allocation of communication resources.

On the other hand, C-V2X based systems exploit the already installed communication infrastructure for 4G/LTE, 5G, or other V2X-capable generation of cellular networks. Communication resources in these networks are usually a subject to a centralized network management. C-V2X has been first introduced in 3rd Generation Partnership Project (3GPP) Release 14 [4]. 3GPP Release 14 included a definition of LTE Advanced Pro V2X services, architectures, radio access network regulations and four use case scenarios [5]:

- Vehicle-to-Vehicle (V2V) communication
- Vehicle-to-Infrastructure (V2I) communication
- Vehicle-to-Network (V2N) communication
- Vehicle-to-Pedestrian (V2P) communication

The C-V2X communication can be carried out using LTE Uu and PC5 interfaces. LTE Uu interface is used for communication with the cellular base station and the PC5 interface is used for direct communication between V2X nodes [6,7]. Furthermore, 3GPP

Release 14 defined two additional C-V2X Modes - 3, 4 which are used to allocate communication resources to the V2X nodes. While Mode 3 is used to allocate network resources to the V2X nodes with the support of the cellular base station, Mode 4 allows the nodes to allocate and manage their communication resources fully autonomously in a case when the connection to the cellular base station is not available [8].

With the ever evolving cellular network infrastructure, the use of C-V2X for vehicular applications might seem as a no-brainer choice. However, due to its relative novelty compared to the already mature DSRC, some studies advise more cautious approach [9]. Studying the performance of both technologies in the context of a specific vehicular application may provide valuable insights on their ability to meet the application performance requirements.

While many studies investigated the performance of both technologies for specific Cooperative Intelligent Transport System (C-ITS) applications relying on beaconing, some applications require a network level routing of the communication packets for their operation. For this reason, it is essential to study the performance of both technologies in unicast scenarios and determine the conditions under which one of the technologies has a performance edge over the other. To the best of our knowledge, this issue has not been addressed in the literature yet.

In order to assess the routing performance of DSRC and LTE-based C-V2X communication, we conducted a federated traffic-telco simulation study using a grid road network. The vehicles in the simulation run a Privacy Ensuring Emergency Vehicle Approaching Warning application [10] using IEEE WAVE [3] communication standard in the first simulation scenario and LTE-V2V [4] in the second simulation scenario. We evaluate the routing performance of both technologies in terms of first message delivery delay, average message end-to-end delay and packet delivery ratio.

The rest of this paper is organized as follows. Section 2 explores the relevant literature and similar efforts undertaken by other authors. Section 3 describes the simulation setup, the case study and methodology used to obtain the results. In Section 4, simulation results are presented and discussed. The paper is concluded by Section 5.

2. RELATED WORK

Some studies, for example by Nguyen et al. [12] suggest that LTE provides a significant improvement over DSRC in communication range without suffering in other aspects. The majority of the available works, however, suggest that both technologies come with their own set of strengths and limitations when applied in the vehicular environment.

Bazzi et al. [12] conducted an analytical comparison of IEEE 802.11 OCB and LTE-V2V applied to vehicle-to-vehicle beaconing. Numerical results show that LTE-V2V has been able to support the same beacon generation frequency as IEEE 802.11 OCB but with less dedicated communication resources. The authors, however, observed that the LTE's advantage quickly reduces with the increasing communication range.

Cecchini et al. [13] compared the performance of both technologies by a simulation study considering Cooperative Awareness Service in a highway scenario. They conclude that LTE-V2V outperforms IEEE 802.11 OCB in terms of packet reception ratio while IEEE 802.11 OCB achieves lower communication delay with increasing packet length and communcaition range.

Mir et al. [14] conducted another simulation-based study comparing the performance of both technologies. The results indicate that IEEE 802.11 OCB performs acceptably in sparse networks with limited vehicle mobility while LTE meets most reliability, mobility and scalability application requirements. On the other hand, the study points out that satisfying stringent delay requirements by LTE is a challenge.

While more studies, e.g. [15] tend to support those results as well, overall, there seems to be no consensus in the available literature about the dominance of either of the communication technologies in all target key performance indicators relevant to C-ITS applications. Instead, majority of the studies show that one of the technologies tends to overperform the other in certain aspects when deployed by a specific C-ITS application. Furthermore, to the best of our knowledge, relevant works consider only broadcasting of the C-ITS application messages and the comparison of unicast performance of both technologies is missing in the available literature.

3. SIMULATION EXPERIMENT DESCRIPTION

The simulation experiments were carried out using OMNeT++ discrete event network simulator [16] with Veins [17], INET [18] and SimuLTE [19] simulation frameworks. Simulation of Urban Mobility (SUMO) [20] was used to generate the traffic flows for the simulation.

As a transport network model, we used a Manhattan grid with overall size of 2500x2500 m and 100 m street length. The topology of the simulation experiment is detailed in Figure 1.

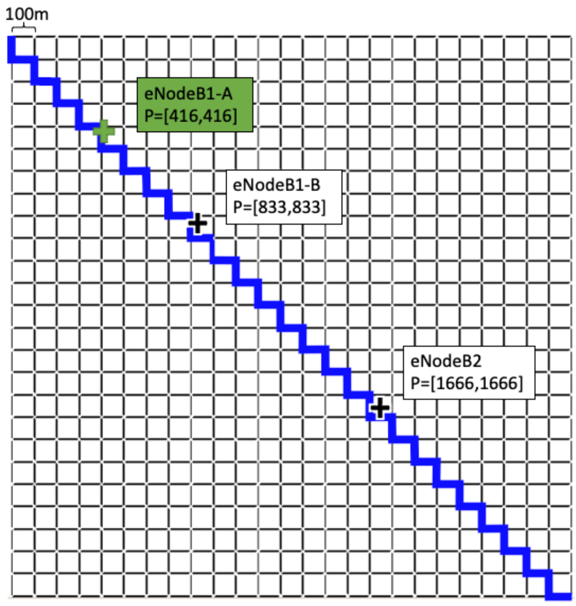


Figure 1: Topology of the simulated network. The two eNodeBs are displayed with their X and Y coordinates in m. The diagonal route of the vehicles is displayed in blue.

The network consists of two LTE eNodeBs. The overall route length is 4800 m. The intensity of the vehicle flow is 1200 vehicles/hour. Maximum vehicle speed is set to 50 kph.

We consider three communication scenarios. In the first communication scenario, the eNodeBs (eNodeB1-B and eNodeB2) are uniformly spaced along the diagonal route of the vehicles, as presented in the Figure 1 (in black). The distance between the sending vehicle and its serving eNodeB1-B d_{eNB1} is 1200m. In the second considered communication scenario, the eNodeB1 (eNodeB1-A) is closer to the sending vehicle, as illustrated in the Figure 1 (in green). The distance d_{eNB1} in this case is 600 m. In the third considered communication scenario, the vehicles communicate directly using DSRC communication. The cellular infrastructure is not used in this case.

The performance of the communication network has been studied in the context of the PEEVA-WS [10] application. This application has been designed to provide warning messages indicating an approaching emergency vehicle (EV) while maintaining the privacy of the sensitive EV data. PEEVA-WS is run by an emergency vehicle which appears after the simulation initialization period of 450 seconds. Immediately after the EV enters the simulation, it starts to transmit communication packets to a selected destination vehicle. The distance between the destination vehicle and EV is varied in the range 100-2800 m. Data packets are exchanged directly between the two vehicles. In the case of LTE technology, the transmission between the EV and the destination vehicle is carried out as an eNodeB-assisted device-to-device (D2D) communication. Hence, the vehicles communicate directly but the scheduling and allocation of the network resources used for the transmissions is managed by the eNodeB. Target metrics for evaluation of the communication network performance are:

- First message delivery delay – it is a time between creation of the first successfully delivered communication packet at the EV's communication module and the moment of reception of the same communication packet by the application layer of the destination vehicle's communication module. The first message delivery delay includes also the time necessary to establish the communication route between the EV and the destination vehicle and is therefore expected to be higher compared to the end-to-end delay of the subsequent communication packets;
- Average end-to-end delay – is the average time between creation and successful reception of all communication packets transmitted between the EV and the destination vehicle during the whole simulation;
- Packet delivery ratio – defined as the ratio between packets sent by the EV N_{sent} and received $N_{received}$ by the destination vehicle during the whole simulation as described in equation:

$$PDR = 100\frac{N_{sent}}{N_{received}}\ [\%] \qquad (1)$$

Each destination vehicle transmission has been simulated using three different random seeds for each of the investigated communication technologies. Detailed values of simulation parameters are summarized in Table 1.

Table 1: Simulation parameters

	DSRC	LTE
Carrier frequency	5900 MHz	2100 MHz
Channel bandwidth	10 MHz	10 MHz
TX Power	20 dBm	eNodeB: 40 dBm UE: 26 dBm
Data rate	6 Mbps	10 Mbps
Packet generation frequency	1 Hz	
Communication distance	100-2800 m	
Traffic intensity	1200 vehicles/hour	
Routing protocol	AODV	-

4. SIMULATION RESULTS

In terms of Packet Delivery Ratio, the results achieved in the three simulated scenarios seem surprising in the context of the available literature. As can be seen in the Figure 2, the DSRC technology achieved the highest PDR regardless of the communication distance, while LTE in both scenarios experienced a sudden drop in PDR in the range between 250 and 1400 m.

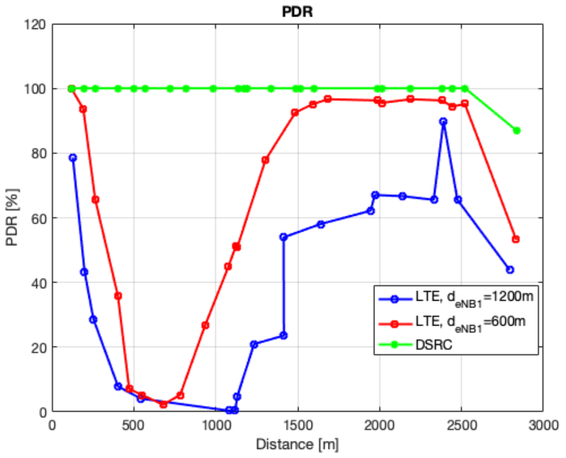


Figure 2: Packet Delivery Ratio.

Figure 3 shows the results in terms of the first message delivery delay. The best results were achieved by LTE with $d_{eNB1} = 600\ m$. The delay in this case does not differ significantly from the average end-to-end delay. In the case of DSRC technology, the induced first message delivery delay is higher since it includes the duration of AODV protocol route discovery. After the communication route is established by the AODV protocol, the subsequent packets are delivered with much lower average end-to-end delay.

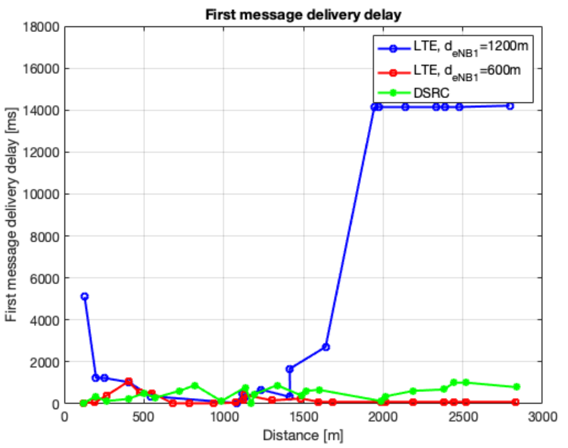


Figure 3: First message delivery delay.

In terms of average end-to-end delay, our observations confim the results achieved by [21]. The results presented in Figure 4 demonstrate that once the communication route is established, the DSRC is able to operate with lower end-to-end delay compared to the LTE. One of the contributing factors for that is the fixed delay induced by the LTE resource selection algorithm which cannot be avoided. On the contraty, the resource access delay in DSRC is given by the CSMA-CA parameters and depends on the network load [21]. Mean end-to-end delay $\bar{L}_{DSRC}$ averaged throughout the whole simulation and throughout all destination vehicles is $(7 \pm 70)\ ms$. In the case of LTE with $d_{eNB1} = 600\ m$, the $\bar{L}_{LTE}^{600m} = (68 \pm 181)\ ms$. The worst performing case in terms of end-to-end delay is LTE with $d_{eNB1} = 1200\ m$, where the $\bar{L}_{LTE}^{1200m}$ is $(2640 \pm 5373)\ ms$.

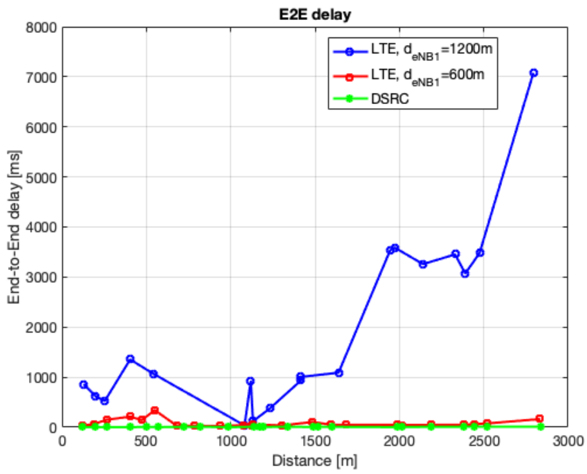


Figure 4: Average end-to-end delay.

5. CONCLUSION

In this paper, we compared the unicast performance pf DSRC and LTE-V2V technology in a vehicular environment in the context of Privacy Ensuring Emergency Vehicle Approaching Warning application. The simulation study has been carried out using a 25x25 Manhattan grid road network model with 100 m street length.

Our results suggest, that in such a scenario, best results in terms of communication delay and packet delivery ratio are achieved by DSRC technology. The LTE-V2V technology has been investigated for two different distances between eNodeB and the transmitting vehicle d_{eNB1} – 600 m and 1200 m. In the case of 600 m d_{eNB1}we observed significantly lower communication delay and higher PDR than in the case of 1200 m d_{eNB1}. However, in both cases, there is a steep drop of PDR in communication distances between approximately 250 and 1400m. Destination vehicles in those ranges are at the boundaries of the transmitting vehicle's communication range or beyond. In such a case, the error rate of the communication increases dramatically and the same communication packet is being retransmitted multiple times until the LTE MaxHarqRtx parameter is achieved. After this point, the direct communication between the vehicles is impossible. The communication is then switched from D2D mode to infrastructure-assisted mode, where the data packets between the transmitting vehicle and the destination vehicle are relayed by the serving eNodeB.

The simulation results further indicate that the DSRC communication in the presented scenario is very reliable. This is caused by the fact that the traffic volume in the simulation is high enough for the routing protocol to find the optimal route to the destination in each case. Moreover, the communication route is rather stable and since there are no significant connection break-ups, the PDR stays very high.

The variability of the results in the available literature indicate that the communciation performance of both technologies is likely very scenario specific. Therefore, an investigation of the communication performance under different traffic scenarios and traffic volumes could be beneficial.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract no. SK-IL-RD-18-005 ICT and smart cars for efficient emergency response and traffic management (SENECA).

This publication was realized with support of Operational Program Integrated Infrastructure 2014-2020 of the project: „Value of travel time from an individual traveller’s perspective“ (ITMS2014+ 313011T241), co-financed by the European Regional Development Fund.

REFERENCES

[1] IEEE Std 802.11p-2010. (2010). IEEE Standard for Information technology Local and metropolitan area networks Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments.

[2] ETSI EN 302 665 V1.1.1. (2010). Intelligent Transport Systems (ITS); Communications Architecture.

[3] IEEE Guide for Wireless Access in Vehicular Environments (WAVE) – Architecture. (2014). In IEEE Std 1609.0-2013, pp.1-78, doi: 10.1109/IEEESTD.2014.6755433.

[4] Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2. (2016). In 3GPP TS 36.300.

[5] Kiela, K., Barzdenas, V., Jurgo, M., Macaitis, V., Rafanavicius, J., Vasjanov, A., Kladovscikov, L., & Navickas, R. (2020). Review of V2X–IoT Standards and Frameworks for ITS Applications. Appl. Sci. 2020, 10, 4314. https://doi.org/10.3390/app10124314

[6] Qualcomm Technologies. (2019). Introduction to Cellular V2X. Available online: https://www.qualcomm.com/media/documents/files/introduction-to-c-v2x.pdf (accessed on 8 June 2020).

[7] Wang, J., Shao, Y., Ge, Y., & Yu, R. (2019). A survey of vehicle to everything (v2x) testing. Sensors 2019, 19, 334.

[8] Molina-Masegosa, R., & Gozalvez, J. (2017). System Level Evaluation of LTE-V2V Mode 4 Communications and Its Distributed Scheduling. 2017 IEEE 85th Vehicular Technology Conference (VTC Spring) (pp. 1-5). Sydney, NSW, Australia. doi: 10.1109/VTCSpring.2017.8108463.

[9] Yu, T., Zhang, S., Cao, S., & Xu, S. (2018). Performance Evaluation for LTE-V based Vehicle-to-Vehicle Platooning Communication. 2018 24th Asia-Pacific Conference on Communications (APCC) (pp. 15-20). Ningbo, China. doi: 10.1109/APCC.2018.8633551.

[10] Petrov, T., Pocta, P., Roman, J., Buzna, Ľ., & Dado, M. (2020). A Feasibility Study of Privacy Ensuring Emergency Vehicle Approaching Warning System. Appl. Sci. 2020, 10, 298. https://doi.org/10.3390/app10010298

[11] Nguyen, T.V, Shailesh, P., Sudhir, B., Kapil, G., Jiang, L., Wu, Z., Malladi, D., & Li, J. (2017). A comparison of cellular vehicle-to-everything and dedicated short range communication. 2017 IEEE Vehicular Networking Conference (VNC) (pp. 101-108). Turin, Italy. doi: 10.1109/VNC.2017.8275618.

[12] Bazzi, A., Masini, B.M., Zanella, A., & Thibault, I. (2016). Beaconing from connected vehicles: IEEE 802.11p vs. LTE-V2V. 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC) (pp.1-6). Valencia, Spain. doi: 10.1109/PIMRC.2016.7794598.

[13] Cecchini, G., Bazzi, A., Masini, B.M., & Zanella, A. (2017). Performance comparison between IEEE 802.11p and LTE-V2V in-coverage and out-of-coverage for cooperative awareness. 2017 IEEE Vehicular Networking Conference (VNC) (pp.109-114). Turin, Italy. doi: 10.1109/VNC.2017.8275637.

[14] Mir, Z.H., & Filali, F. (2014). LTE and IEEE 802.11p for vehicular networking: a performance evaluation. J Wireless Com Network 2014, 89. https://doi.org/10.1186/1687-1499-2014-89

[15] Vukadinovic, V., Bakowski, K., Marsch, P., Garcia, I., Xu, H., Sybis, M., Sroka, P., Wesolowski, K., Lister, D., & Thibault, I. (2018). 3GPP C-V2X and IEEE 802.11p for Vehicle-to-Vehicle communications in highway platooning scenarios. Ad Hoc Networks. 74. 10.1016/j.adhoc.2018.03.004.

[16] OMNeT++ Discrete Event Simulator, https://omnetpp.org

[17] Sommer, C., German, R., & Dressler, F. (2011). Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis. IEEE Transactions on Mobile Computing (TMC), vol. 10 (1), 3-15.

[18] INET Framework, https://inet.omnetpp.org

[19] Virdis, A., Stea, G., & Nardini, G. (2016). Simulating LTE/LTE-Advanced Networks with SimuLTE. In M.S. Obaidat, J. Kacprzyk, T. Ören, J. Filipe, (Eds.), Simulation and Modeling Methodologies, Technologies and Applications (Volume 402 of the series Advances in Intelligent Systems and Computing, pp. 83-105). Springer, doi: 10.1007/978-3-319-26470-7_5

[20] Alvarez Lopez, P., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wießner, E. (2018). Microscopic Traffic Simulation using SUMO. 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 2575-2582). Maui, USA.

[21] Mannoni, V., Berg, V., Sesia, S., & Perraud, E. (2019). A Comparison of the V2X Communication Systems: ITS-G5 and C-V2X. 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring) (pp. 1-5). Kuala Lumpur, Malaysia. doi: 10.1109/VTCSpring.2019.8746562.

THE IMPACTS OF FLASHING GREEN AT SIGNALIZED INTERSECTIONS – CASE STUDY OF BŘECLAV

Eva Pitlová, Ing. PhD.
Lucie Vyskočilová, Ing.
CDV - Transport Research Centre
Líšeňská 33a 636 00, Brno, Czech Republic
eva.pitlova@cdv.cz, lucie.vyskocilova@cdv.cz

ABSTRACT

The red signal violation may lead to traffic conflicts that often result in death or serious injuries at signalized intersections. Therefore, the flashing green was designed to work as a psychological measure to eliminate this violation. It works on the principle of dilemma zone reduction as it simplifies drivers' decision to smoothly stop or safely drive through the intersection. As side effects of the flashing green introduction, the increase of premature stops and rear-end collisions were monitored in recent research. The aim of this study is to demonstrate the flashing green effect on the intersection safety and capacity. For this purpose, a pilot study of 6 consequent intersections in Břeclav was carried out. Drivers' behavior surveys were performed in two ongoing steps. During traffic conflict observation, video from a drone situated above the intersection was recorded. The results show decrease in vehicles' speed at time of yellow onset. On the contrary, a slight increase in conflicts at the time of flashing green onset was observed. The survey shows that drivers stop earlier before yellow onset which leads to a 1,8s reduction of effective green time. Interviews with local drivers about the signal change show people´s positive attitudes towards the flashing green introduction (84%). Furthermore, results show that a quarter of respondents do not know how to correctly respond on flashing green signal, so a nationwide campaign would be needed before flashing green´s widespread introduction.

Keywords:
intersection, signal plan, flashing green, driver's behavior, survey.

1. INTRODUCTION

Traffic signal change from green to yellow informs the drivers about the end of free flow. According to that they can adapt their speed and therefore stop at intersection. Still, there exist many drivers that pass through the intersection during red signal because the length of yellow signal is insufficient to safely stop at the intersection. In a recent CDV - Transport Research Centre survey (Ambros, 2013) in Czech Republic the stop obligation during the intersection red signal is violated in 15% of the cycles and the obligation to stop during the yellow signal is violated in up to 30%. As results show, most of the drivers violated the yellow signal on purpose. They approach the intersection in such a great distance when the yellow signal onsets that they can safely stop but instead of it they decide to take the risk and get through the intersection unsafely. Such traffic violations occur in more than 1/2 of the traffic signal cycles, in which traffic accidents occur. The driver's failure to comply with the red or yellow signal stop obligation exposes the drivers themselves as well as the other road users to the risk of an accident. Therefore, the intermittent green signal, also known as flashing green, was designed to work as a psychological measure to eliminate the red and yellow signal violation at signalized intersections. It works on the principle of reducing the dilemma zone and thus makes it easier for drivers to decide whether they can smoothly stop or safely drive through the intersection. Its introduction is also recommended by traffic psychologists, who rely on research and experiences in many countries (Austria, 1960). However, drivers' behavior among countries may vary, and therefore, what is good for one country may not be useful for another. The aim of this study is to demonstrate the effect of the flashing green on the safety and capacity of signalized intersections in the Czech Republic.

1.1. Flashing green experience

For the first time in Europe the flashing green was applied at signal controllers in Austria, in 1983. After successful implementation, many countries followed their example such as Switzerland, Germany, Estonia, Latvia, Lithuania, and others (Köll, 2002). Extensive research in Germany and Switzerland

(Factor R., 2012) have shown that, in addition to its positive impact, the flashing green can also cause the driver´s earlier stops at intersection which may lead to a capacity reduction and rear – end collisions. This is due to the driver´s sudden and premature stop as a reaction to the signal change into the flashing green. The research results show different numbers of driver´s earlier stops throughout the countries. In Austria, the increase rises to 80%, in Switzerland 35% and in Germany 28%.

Later in 2009, there was a study of flashing green impacts at 9 intersections and 3 crossings in Bratislava, capital of Slovakia (Novotný T., 2013). After its completion, the introduction of flashing green at all intersections continued for three more years with recommendation to apply the flashing green at all intersections in the city. After the three years, the city decided to set it aside, although there was no negative effect monitored on traffic flow fluency and the intersections safety was increased. It turned out that at several intersections in Bratislava there are obsolete traffic light controllers and for the flashing green introduction they would have to be replaced with new ones, which would lead to high financial demands for the city. Due to the disintegration of the intersection's coordination, the city even decided to cancel the flashing green at the existing intersections.

In Poland, extensive survey of driver's opinion about innovations in traffic light signals shows that only 12% were in favor of flashing green, 29% of countdown signals, and 10% of those asked preferred a combination of both (Łopuszyński, 2018). According to the interview results we can say that drivers in Poland are quite positive about the signal modification of traffic lights, but rather than flashing green they prefer traffic lights with information about the remaining signal duration, which is currently the safest, but more expensive option and has its limitations.

Outside Europe, the flashing green exists mostly in the US, Mexico, Turkey, Israel, Russia, China, India, and Cambodia. The most significant research took place in America and China (Wu, 2014) (Tang, 2015). The results show that the flashing green has a positive psychological effect on the driver, which reflects in the intersection safety. However, currently the traffic lights with signaling countdown are being introduced much more instead of the flashing green. The countdown signal provides drivers with information about the duration of the current phase when they approach an intersection and so it is easier for them to adjust their behavior. In addition, the red signal countdown significantly reduces the vehicle´s delay time at the start of green phase, thus increasing the overall capacity of the intersection. To introduce the countdown device at an intersection, it is necessary to completely replace the current traffic lights or equip it with a counting device, which is expensive especially for a nationwide application. In addition, currently most of the traffic lights operate under the dynamic traffic management, to accurately react on the actual traffic flow and, for these signal plans, the countdown traffic lights cannot be used.

International studies (Köll, 2002) (Novotný T., 2013) (Tang, 2015) (Wu, 2014) confirm that flashing green has a positive psychological effect on the driver, which reflects in intersection safety enhancement. However, it has been found that in addition, flashing green may result in capacity decrease or in a slight accident increase in the specific cases of rear-end collisions (Koll, 2004). Therefore, it is necessary to verify these possible negative effects on a selected sample of intersections before its widespread implementation.

1.2. The law behind its application

The flashing green introduction preceded the legislation change. This change has been already made in neighboring countries such as Austria, Slovakia, and Poland (Austria, 1960) (Slovak Republic, 2009) (Poland, 2002). There is a difference in the signal meaning and legislation among the countries, specifically in the definition of the yellow signal. While in Austria there is hard restriction passing on the yellow signal, in Slovakia, as in the Czech Republic, the law allows drivers to pass through the intersection when the yellow signal onsets and the car approaching the intersection is in such a small distance that it is impossible to safely stop at intersection. In Austria, the flashing green means end of the free driving through phase, and although its effect on the driver behavior has been demonstrated, still it is not considered in the capacity calculations of traffic lights (Factor R., 2012).

2. DATA AND METHODS

The flashing green impacts were determined based on the driver behavior pilot study at selected intersections. The survey was performed in two steps. While there was a drone recording current intersection from above, the conflicts on selected approaches were concurrently monitored. In addition, a questionnaire was created to observe the information about driver's knowledge, manner, and attitudes towards the intermittent green signal at the traffic lights.

2.1. Case study in Břeclav

Based on the research abroad and with regards to technological requirements, 6 intersections in Břeclav, that are part of the 1st class road number 55, were selected for monitoring driver behavior. The city is located near the Austrian and Slovak borders and therefore it is considered that the drivers are aware of the flashing green operation. The signal plan was modified from 1st September to 29th October 2019 at all six intersections shown in Fig. 2, after obtaining all the necessary permits. The test operation of the modified signal plan was allowed only for 2 months, and therefore the driver's behavior was monitored in two steps - by analyzing the image from camera recordings taken from drones and traffic conflict observation on the site.

Figure 1: Location of Břeclav in the Czech republic and neighboring countriies

2.2. Data collection

To provide complex data analysis of traffic flow, there was a drone recording the intersection from above, while the potential conflicts were monitored. In addition, during the investigation the questionnaire interview of local drivers was performed.

The drivers' reactions were monitored in three phases and compared to each other:

- F1 before the flashing green introduction,
- F2 shortly after the flashing green introduction (first week after introduction),
- F3 about a month after the flashing green introduction (6th - 9th week after the introduction).

Since the drone can only fly at a specific height above buildings and roads, most of the intersections were recorded from an angle, and therefore some approaches or whole intersections were hidden behind the trees or buildings. Thus, two out of six intersections were excluded from the survey (Fig. 2). Nevertheless, recording was performed at 4 intersections (Table 1) during the saturated traffic flow and in good weather conditions (without raining and strong wind). The monitored approaches of monitored intersections in Břeclav.

Table 1: The monitored approaches of monitored intersections in Břeclav

Int.	Road crossing the I/55	Approach 1	Approach 2
K1	Sovadinova	I/55 – Vienna	Sovadinova
K2	17.listopadu	I/55 – Brno	II/425
K3	Pod Zámkem	I/55 – Brno	I/55 - Vienna
K4	U Nemocnice	I/55 – Brno	U Nemocnice

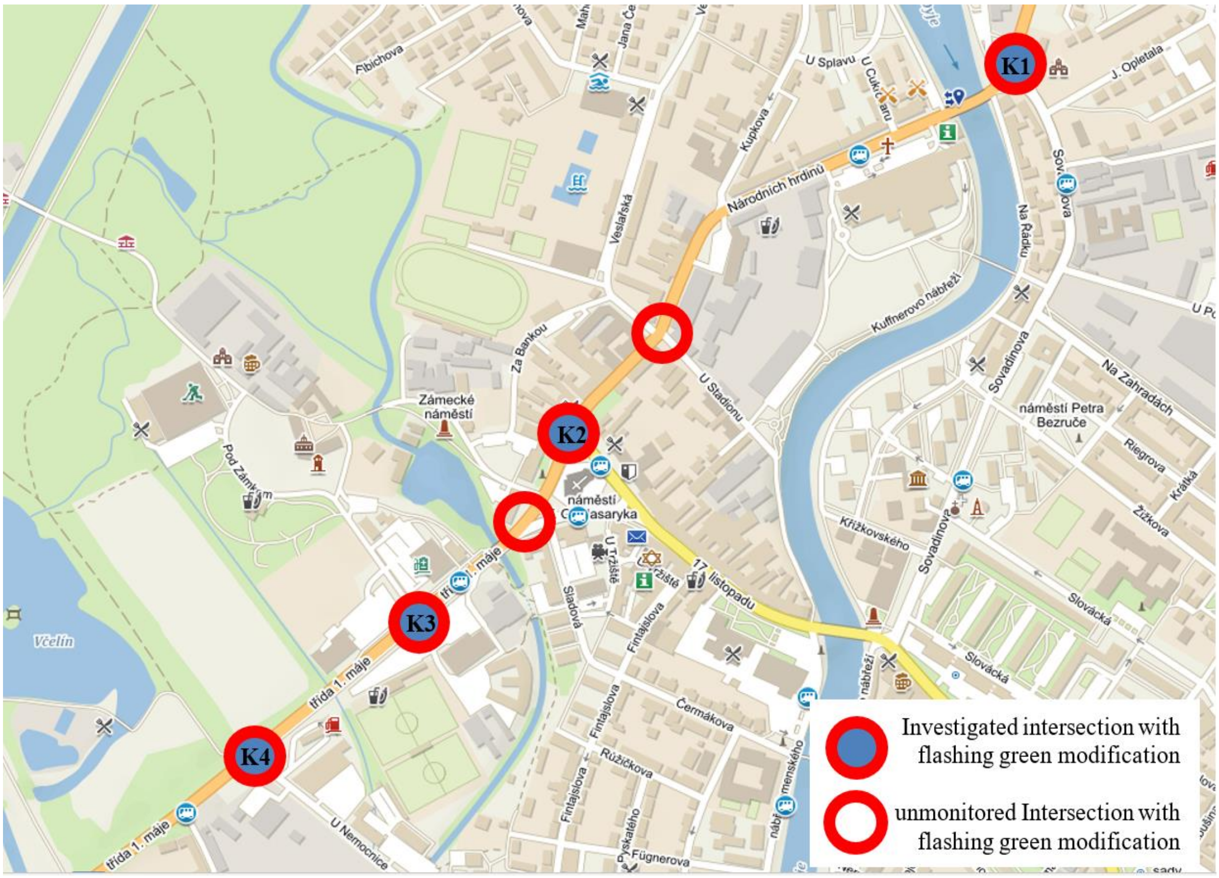


Figure 2: Map of monitored intersections with flashing green modification in Břeclav

2.3. The survey

For traffic accident comparison before and after the flashing green introduction, the long-term driver behavior monitoring would be the appropriate option. However, the test run of the modified signal plan was allowed only for 2 months, which is a very short period for this type of comparison and the results of this method would be insufficient. Therefore, the traffic conflict monitoring methods were used to monitor driver behavior. At the time of the drone recording, the traffic conflicts at selected approaches were also monitored according to the methodology (Ambros, 2013). At each intersection, monitoring was performed at two approaches at a time. Of all possible conflicts, only rear-end types of conflicts in 1–3 gravity were monitored. The conflicts of zero severity (violation of red signal) were not reported. Monitoring focused especially on the signal change from green to flashing green.

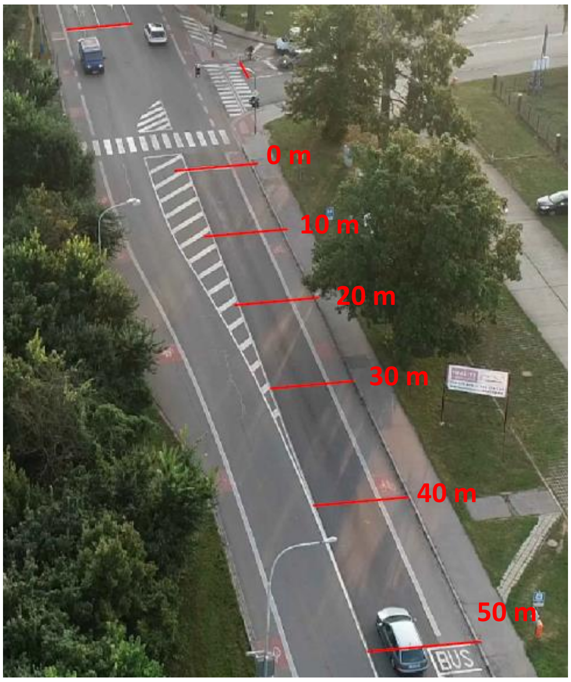


Figure 3: Example of profiles at an intersection K4

The drone recording option was chosen, after surrounding inspection that there was not any suitable building or structure for static camera fixation to record necessary area of whole intersection with its approaches. The drone was able to record approximately 18 minutes video with loaded battery (depending on weather and wind) and after this time it needed a battery change. The battery changes were repeated as many times as necessary to achieve the desired intersection recording time. The drone always recorded the view from the same point and angle even after the battery change (based on GPS coordinates). For each intersection, image analysis was performed to obtain text files with time data (accurate to thousands of seconds) of the vehicles passing through predefined profiles. On the approaches, profiles in 10m distances were formed starting from the stop line, continuing up to 50m (Fig. 3). Based on this information, the average speeds of each recorded vehicle passing between the individual profiles were calculated.

The vehicle passages were time synchronized with the signal plans for each traffic movement. The three phases of the survey were compared to each other and evaluated due to:

- vehicle speeds that were affected or were not by signal change, measured in individual zones (between profiles – 10 m),
- speed during the yellow signal and acceleration immediately after the yellow signal onset,
- speed on the STOP line for vehicles passing an intersection at a yellow or red signal,
- the proportion of cycles in which at least one vehicle passed the intersection on a red or yellow signal, relative to the total number of cycles.

To analyze the knowledge, manner, and attitudes of drivers, local interviewing was carried out using two combined questionnaire surveys. The first survey took place in Břeclav where data were collected directly close to the investigated intersections and the second survey was conducted as a classical representative survey. Field data collection took place near the investigated intersections after about a month of its signal plan modification, while drivers were asked about their knowledge of the flashing green and the change in their behavior. The sample consisted of a random selection of drivers. The national data collection was carried out using a statistical survey conducted based on quota sampling. The aim of the combination of these two methods was to verify the influence of flashing green on drivers' behavior and knowledge.

3. RESULTS

According to atypical values obtained from conflict monitoring, the individual F1, F2, F3 phases were compared with each other based on the rate of conflicts (number of conflicts on selected approach related to the traffic volume, expressed as a percentage). Special attention was given to the results comparison of phase 1 and phase 3.

The results of the relative conflict rate in each phase for investigated intersections and its approaches are shown in Table 2 colored for easy comparison (green - lowest value, yellow - mean value, red - highest value). It is obvious that at some intersections after the flashing green introduction, the gradual development of the increase in conflicts was monitored for each phase. This means that even a month after the signal modification, the drivers were not used to it and their uncertainty has risen. Due to the short testing period (2 months) the further development in driver behavior cannot be predicted.

Table 2: Conflicts monitoring evaluation

Intersection approach		Monitored conflicts			Relative confliction [%]		
		F1	F2	F3	F1	F2	F3
K1	1	9	10	2	0,413	0,465	0,082
	2	0	0	0	0,000	0,000	0,000
K3	1	0	2	2	0,000	0,101	0,136
	2	0	0	0	0,000	0,000	0,000
K4	1	1	5	10	0,038	0,233	0,386
	2	0	4	7	0,000	0,123	0,224
K5	1	1	6	7	0,043	0,235	0,269
	2	1	4	1	0,165	1,361	0,208

The image analysis from drone recording shows that the relative number of vehicles passing the intersection at the red or yellow signal decreased at all intersection approaches comparing the first phase F1 with the phase F3 (at the K5 intersection was the decrease even from 35% to 4%). There was just slight increase at the K1 intersection comparing phase F1 and phase F2 but again in phase F3 it decreases. According to the vehicle speeds, after the flashing green introduction at all investigated intersections, the speeds decreased at the time of the yellow signal beginning. This means that the drivers drove at a lower speed when the signal changed to yellow and so they were more prepared to stop at the red signal. That leaded to a reduction in the speed of vehicles that passed the intersection on yellow or red signal at all intersections (Table 3), which reduces the gravity of the consequences in case of a vehicle collision. Based on the fact that more than one vehicle can pass on the red or yellow signal in one cycle, the relative number of cycles that were violated (at least one vehicle passed on the red or yellow signal) were also evaluated. To better illustrate that, Table 3 demonstrates the violated cycles. At two intersections (K2, K3) the percentage of red signal violated cycles keep constant in all phases (around 5%), while at the remaining two intersections (K1, K4) the percentage of cycles decrease comparing phase F1 with third phase F3.

The values for each intersection in the three phases (F1, F2, F3) are colored for easy comparison (green – positive improvement, yellow – middle stage, red – negative deterioration).

Table 3: Video analysis results for each intersection in three phases (F1, F2, F3)

Monitored phase of flashing green	F1 before introduction				F2 immediately after introduction				F3 one month after introduction			
Intersection	K1	K2	K3	K4	K1	K2	K3	K4	K1	K2	K3	K4
Number of investigated cycles	143	146	138	161	153	106	132	165	156	93	219	156
vehicles investigated during flashing green	-	-	-	-	52	47	64	90	57	26	95	67
% of violated cycles on yellow signal	4	41	18	33	10	8	18	28	8	9	11	3
% of violated cycles on red signal	10	7	5	11	6	12	4	5	6	5	5	1
% of vehicles that violated yellow or red signal	12	48	22	35	16	19	16	26	9,7	15	16	3,7
Average speed at the yellow onset [m/s]	9,3	13	11	9,8	8,7	11	9,5	7,9	7,6	10	9,6	7,5
V85 at the yellow signal beginning [m/s]	11	15	15	13	11	15	13	11	11	14	11	11
Speed at stop line during red or yellow [m/s]	8	13	12	9,9	8,7	7,1	9,3	7,7	6,1	9,2	7,9	7,5
Acceleration at stop line during red /yellow [m/s2]	0,8	1,1	1,5	-0,7	-1,1	-0,2	-0,2	-0,9	-0,4	-2,4	-1	-1,3
% of early stopped vehicles due to flashing green	-	-	-	-	33	17	13	22	33	31	41	58
% of vehicles that passes during flashing green	-	-	-	-	67	83	88	78	67	69	59	42

3.1. Flashing green impacts

What was demonstrated in this paper is that, due to the flashing green at all intersections, there was an increase in the number of vehicles that reacted on signal change approximately in 30-10 m before the STOP line and stopped. This distance is a critical value, as the necessary car stopping clearance in driving speed of 50 km/h (13.9 m/s) is around 35 m. At the time of the yellow signal onset, the drivers were therefore able to better evaluate whether they would safely pass through the intersection or not. On the other hand, sudden deceleration of vehicle speed that is approaching the intersection can cause rear-end type of conflicts. Furthermore, surveys showed that about a third to half of drivers stop prematurely in front of the intersection during the flashing green. This, together with the fact that drivers reduce their speed, could negatively affect the capacity of intersection.

From the results of conflict monitoring there was a slight increase in the monitored type of conflicts after the introduction of the flashing green. These were mainly the low severity conflicts (numbered as 1), which corresponds to the expected normal maneuvers of the front vehicle and the subsequent reaction of the following vehicle. In Table 2 the relation between monitored conflicts and the current traffic flow as a relative confliction for each intersection divided in three phases is determined. Despite a slight increase in rear-end type of conflicts in the second and third phases (F2, F3), all intersection approaches still present the safe state according to Figure 4, so the introduction of flashing green did not significantly affect the intersections safety. In case of vehicle speed, after the flashing green introduction, there was a decrease in speed at all evaluated intersections at the time of the yellow signal beginning. The drivers drove at a lower speed when the signal changed to yellow and prepared to stop at the red signal. At the same time, at all intersections there was a reduction in the vehicle speed that passed the intersection on a yellow or red signal, which reduces the gravity of the collision consequences.

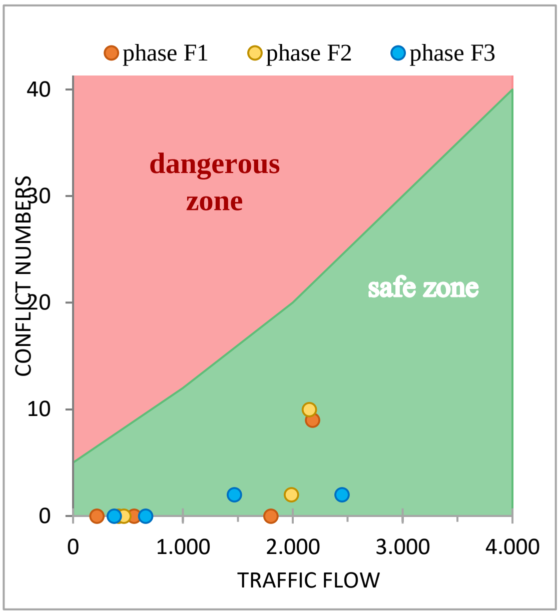


Figure 4: Conflicts gravity evaluation according to the current intersection traffic volume

According to the Czech guidelines for the intersection capacity calculation called TP 188 (Bartoš, 2018) the effective green is considered depending on the duration of the green signal, and its minimum value is the length of the green signal itself (z '= z at values $z \geq 11$ s). It is commonly expressed as follows (Bartoš, 2018):

$$z' = z - r + ž \qquad (1)$$

where:

z ‘- effective green signal time [s],

z - green signal time[s],

r - vehicle start-up time (usually 1 s is considered) [s],

ž - the yellow signal length (approx. 2 s) [s].

In the case of the flashing green implementation, it is necessary to consider its impact on the effective green. It will be necessary to consider this effect of the vehicles that stop at the flashing green in the calculation. According to the results, a third to half of drivers stop prematurely at the stop line during the flashing green. The aim of this study is to evaluate the weight of the impact on effective green at investigated intersections. Based on the obtained data, the real time of the effective green was analyzed. Due to the dynamic controlled signal plans at the investigated intersections, the green length varies and so every cycle monitored in the second and third phase (F2, F3) was counted including the flashing green. The real time of effective green was counted based on the time difference between the first and last vehicle that passed in one cycle (regardless of whether the last vehicle passed the intersection on the green, yellow, or red signal). Only cycles with saturated flow were selected for analysis.

This means that the green time has been fully utilized without large vehicle distances. In the following Figure 5 the relation between real effective green time and the green length of all analyzed cycles is shown.

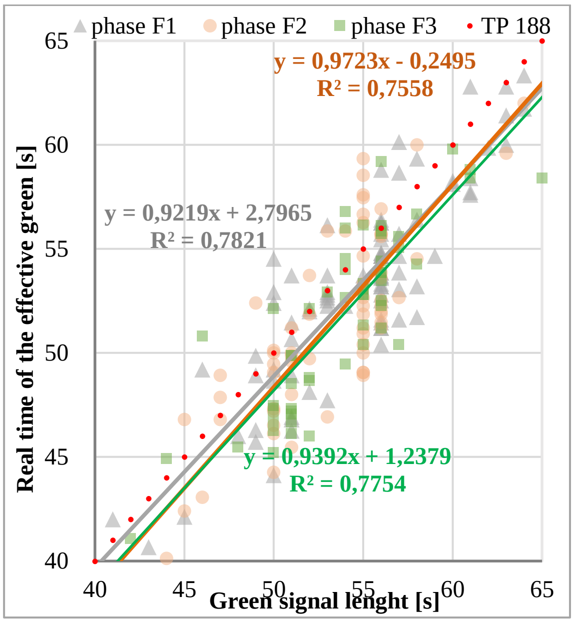


Figure 5: Efective green comparison for each phase and with the TP regulations (Bartoš, 2018)

The value of effective green determined according to the valid regulation TP 188 (Bartoš, 2018) is marked with dotted red line. From the plotted curves the influence of the flashing green on the effective green real time value was determined. Throughout the analyzed range, the effective green real times were lower than the considered value. Furthermore, we can say that the values of effective green real times differ in whole analyzed range in constant value of 1,8 s comparing the first (F1) with the third phase (F3). Comparing the development of effective green real times monitored in the third phase (F3) with the value calculated according to the TP188 (Bartoš, 2018), the value is not constant in the monitored range, but the average value is reduction of about 3,1 s. These reductions in the effective green real time as a consequence of flashing green introduction are quite significant and it would be appropriate to be considered in regulations capacity calculations (Bartoš, 2018). To accurately determine the impact of flashing green on capacity, it will be necessary to perform long-term monitoring of the light-controlled intersections with flashing green implementation, with respect to the effective green time monitoring.

3.2. Driver´s interview

The national survey shows that less than a quarter of respondents (24%, 601 people) have already encountered flashing green at traffic lights, while more than half of the respondents (60%, 1508 people) have not yet encountered this signal. About 16% of respondents (419 people) do not remember. The high percentage of people who do not have experience with the flashing green means that a nationwide campaign needed to be held before it widespread introduction in Czech Republic. In addition to a nationwide survey, a local driver´s experiences with flashing green were investigated. The people were interviewed directly around the investigated intersections and with the help of the city itself by using their web portal. The survey was held about a month after the introduction of flashing green testing. The sample had the character of a driver's random selection who experienced the flashing green testing in Břeclav city.

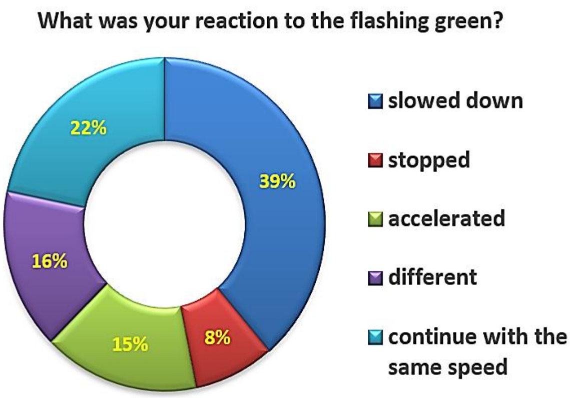


Figure 6: National survey results about flashing green expierence

The results show that when the flashing green occurs, up to 39% of respondents answered that they would rather slow down and 8% would stop completely. 15% of people had the opposite reaction and rather accelerated when flashing green occurs and 16% of people said that they would react differently. 22% of respondents said that the flashing green has no effect on their driving behavior. According to the safety side, 78% of respondents evaluated it positively. The 8% of respondents think that flashing green has a negative effect. This also results in a people's positive attitude towards its nationwide introduction in the Czech Republic presented by 84% of respondents.

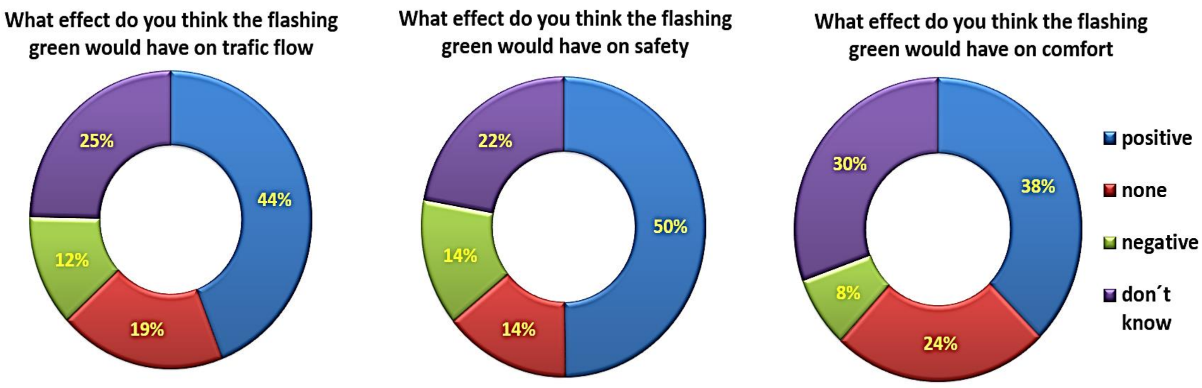


Figure 7: National survey results about flashing green attitude

In Figure 7 the complex results of the survey in terms of the driver's opinion about flashing green impact at the fluency, safety, and comfort are presented. The largest proportions of positive answers were given by a group of drivers with B +. On the contrary, the most negative answers were given by a group of small drivers under 24 years of age.

4. DISCUSION

Drivers' reaction on signal change may vary throughout the countries. According to that, most of the countries carried out their own studies about flashing green before its widespread introduction. As side effects, the premature stops of vehicles at intersections resulting in slight increase in rear-end type of conflicts when the flashing green onset were reported.

In Austria, the side effect on traffic flow fluency the 3% capacity reduction was determined. In this pilot the negative effect was proven in 1,8 s reduction of effective green measured one month after the flashing green introduction and comparing it to the calculated effective green (Bartoš, 2018) the reduction is even 3,1 s.

In addition, an increase in rear-end type of conflicts was observed too but the number and gravity of monitored conflicts still preserve the intersections in safe side according to (Ambros, 2013) as it was established as common maneuver of the consequence vehicle reacting to sudden stop of previous one. The reduction in speed has also been demonstrated and due to that the number of red lights violated cycles decrease.

Most of the surveys beyond Europe about flashing green, point out that even the dilemma zone has decreased, but the drivers incorrectly react on it and so it is necessary to enlighten people before its widespread introduction. For this purpose, as a part of this study the local's knowledges, reactions and opinion about flashing green was carried out. Their reaction was positive about its introduction but mostly the younger group of drivers (18-24 years) mistakenly marked meaning of it which proof the previous studies about their lack of knowledge. As for their reaction, 39% of asked said they would rather slow down when flashing green occurs, which corresponds to 41% of the premature stopped vehicles, monitored one month after the flashing green introduction. For example, in Poland the interview results show people's positive attitude towards flashing green but it also shows that they prefer more the countdown devices so they can see the remaining time of red, which could improve the capacity as the green begins or with the green signal when they approach intersection, they feel better as they know the remaining time which affect the safety. However, the countdown device as traffic light cannot be used at the intersections with dynamic signal plans which are nowadays common in most cities. It is also more expensive if we compare it nationwide introduction with the flashing green introduction expenses. In this case Flashing green is more appropriate.

5. CONCLUSION

The survey shows an increase in the rear-end type of conflicts after the flashing green introduction in almost all monitored intersections approaches (5 out of 8) . This means that the drivers had a higher confidence in their behavior when the flashing green begin. These were mainly the conflicts of 1st severity level which correspond to the expected maneuvers of the previous vehicle and the subsequent reactions of the next vehicle. Despite a slight increase in conflicts in second and third phases (F2, F3), the intersections are still in a safe side. The evaluation of speeds after the flashing green introduction shows, that there was a decrease in speeds at all evaluated intersections at the time of the yellow signal beginning. This means that, after the introduction, the drivers drove at a lower speed when the signal changed to yellow and therefore were better prepared to stop at the red signal. That leads to reduction in the speed of vehicles that passed the intersection on a yellow or red signal at all intersections, which reduces the gravity of the consequences in case of a vehicle collision. Also, the results show that there was a decrease in number of the red signal violation cycles one month after the flashing green introduction at all intersection. Furthermore, the survey showed that about a third to half of drivers stop prematurely at flashing green. This could negatively affect the signal-controlled intersection capacity. According to this fact, the real and calculated effective green was compared for each intersection monitored in the three phases. Based on the monitored data the 3 seconds reduction of effective green was determined compared with the calculated values according to TP 188 (Bartoš, 2018) and about 2s reduction was determined comparing the first and third phase (F1 – before flashing green introduction, F3- one month after introduction). Nevertheless, these are the data from the research at 4 intersections and the effect of flashing green was monitored only for the permitted two-months duration of signal plan modification. Due to the type of intersections surveyed for this study, it would be appropriate to carry out an additional survey at intersections with a higher permitted speed (due to the longer braking distance of vehicles) and less congested traffic. Long-term monitoring is necessary to better quantify the capacity reduction, together with a focus on selected characteristics of the traffic flow distinguishing vehicles categories. To accurately determine the flashing green impact on capacity, it will be necessary to perform long-term monitoring of signalized intersections with flashing green introduction, focusing on monitoring the effective green time. The observations need to be during the fully saturation of the monitored approaches.

The nationwide survey showed that a quarter of respondents have already experienced the flashing green. The high percentage of people who did not have such experience means that a nationwide campaign will be necessary before its introduction and integration into the legislation of the Czech Republic. The local survey focused on people who had to respond directly to the flashing green as a driver. According to the local survey results, up to 39% of respondents lowered their speed and 8% stopped completely when the flashing green occurred. Even 78% of respondents evaluated the flashing green impact positively in terms of safety. This also results in a positive attitude towards its nationwide introduction in the Czech Republic presented by 84% of respondents.

Acknowledgments

The data presented in this paper were obtained in study financed by the Ministry of Transport based on Decision No. 2 / RVO / 2018 File no. 5 / 2018-710-VV / 1 as the institutional support for the long-term conceptual development of a research organization and implemented with the support of the Břeclav city, which permitted the survey to carry out the necessary observations.

REFERENCES

[1] Ambros, J. et al., (2013). Methodology for Monitoring and Evaluating Traffic Conflicts, Transport Research Center, v. V. I., Czech republic: Brno

[2] Austria (1960). Federal Ministry for Digital and Economic Affairs: Act Concerning the Granting of Asylum 1960. In BGBl. Nr. 159/1960 zuletzt geändert durch BGBl. I Nr. 116/2010; § 38. Bedeutung der Lichtzeichen. Vienna, Austria: Bedeutung der Lichtzeichen.

[3] Bartoš, L. et al. (2018). TP 188 Technical regulation for capacity calculaton of the intersection and roads profiles. Plzeň: Koura Publishing.

[4] Factor R., Prashker J.N., Mahalel D.(2012). The flashing green light paradox. Transportation Research Part F: Traffic Psychology and Behaviour.Volume 15, (pp. 279-288).

[5] Köll, H., Axhausen, K., Bader, M. (2002). Decision-making behavior at traffic lights with and without flashing green as an announcement of the transition time yellow. Straßenverkehrstechnik, pp. 339–345.

[6] Koll, H., Bader, M., Axhausen, K.W. (2004). Driver behavior during flashing green before amber: a comparative study. Accident Analysis and Prevention, pp. 273-280.

[7] Łopuszyński, M., Pendzich, I. (2018). Need to use advance notice for urban public transport drivers. Scientific Journal of Silesian University of Technology, pp. 81-90.

[8] Novotný T. (2013, august 6). The flashing green disappears from Bratislava, some experts do not agree. Retrieved from Auto SME: https://auto.sme.sk/c/6887860/blikajuca-zelena-sa-z-bratislavy-vytrati-niektori-odbornici-nesuhlasia.html

[9] Poland (2002). Regulation of the Minister of Infrastructure and Internal Affairs and Administration.:Signs and Signals lights. Poland: Ministers of Home Affairs and Administration.

[10] Slovak Republic (2009). Ministry of the Interior Regulation: Road Act 9/2009. In M. o. Regulation. Bratislava: Ministry of the Interior Regulation.

[11] Tang, K., Xu, Y., Wang, P., Wang, F. (2015). Impacts of Flashing Green on Dilemma Zone Behavior at High-Speed Intersections: Empirical Study in China. Journal of Transportation Engineering, pp. 1-13.

[12] Wu, Y. (2014). Comparative Analysis of Different Dilemma Zone Countermeasures at Signalized Intersections based on Cellular Automaton Model. Orlando, Florida: Dissertation

EFFECTIVENESS OF ADVANCED EMERGENCY BRAKING SYSTEMS IN TRUCKS: AN ANALYSIS OF REAR-END COLLISIONS ON MOTORWAYS IN GERMANY

Leon Straßgütl, M.Sc.
Federal Highway Research Institute (BASt)
Brüderstraße 53, 51427 Bergisch Gladbach, Germany
strassguetl@bast.de

ABSTRACT

Advanced emergency braking systems (AEBS) are advanced driver-assistance systems that can prevent or mitigate rear-end collisions. Since 1st November 2015, these systems are mandatory for newly registered heavy goods vehicles (HGV) and buses above eight tones (vehicles of categories N2, N3 and M3) according to EU-Regulation No. 347/2012.

This paper aims at evaluating the effects associated with the Regulation in Germany. It addresses two research questions: (1) How did the number of rear-end collisions of HGV and buses develop after the Regulation was introduced? (2) Did the severity of rear-end crashes change after the implementation of the Regulation?

To examine the impact of the Regulation, the analysis used individual data of police records on crashes as well as data on technical vehicle information from the German Federal Motor Transport Authority (KBA). At first, relevant vehicles and crashes were identified in the data. Subsequently, treatment and control groups were defined that were compared before and after the Regulation came into effect. The treatment group were rear-end crashes of new vehicles since AEBS mostly affect these crashes. Other crashes and crashes of older vehicles served as control groups. Odds ratios and chi square tests were used to estimate the impact associated with the Regulation.

The results indicate a substantial decrease of rear-end collisions on German motorways associated with the Regulation. Moreover, the results further suggest a reduction in the severity of crashes.

Keywords:
Advanced emergency braking systems (AEBS), Active safety systems, Heavy goods vehicles (HGV), Road safety, Crash analysis.

1. INTRODUCTION

Crashes involving heavy goods vehicles (HGV) have particularly severe consequences. This is especially true for rear-end crashes. Approximately two thirds of rear-end crashes of HGV occur on motorways. This crashes account for more than 90% of all fatalities in rear-end crashes with HGV (Panwinkler 2018).

Advanced emergency braking systems (AEBS) offer the potential to prevent or mitigate rear-end crashes. These systems detect collision risks using camera, radar and/or laser. The driver is first alerted about the possible collision. If the driver does not react by braking or steering, the system autonomously initiates an emergency braking. For HGV and buses, these systems are mandatory as regulated in EU-Regulation No. 347/2012 (European Commission 2012a). The effect associated with the Regulation was evaluated in this paper.

Several studies have estimated the potential effect AEBS have on the number of rear-end crashes. For Germany, for instance, Kuehn et al. (2011) estimated the potential of autonomous emergency braking systems in trucks on the basis of crashes with personal injury or property damage above 15,000 € in 2004 to 2006. They inferred that these systems could have had an effect in 27% of rear-end crashes on all roads if they only detected moving targets and 52% if they additionally detected standing targets.

In an in-depth analysis of crashes of HGV with personal injury in Brandenburg in 2016, Trabert et al. (2018) estimated that 21 out of 25 crashes (84%) could have been prevented or mitigated with an ideal AEBS. An ideal AEBS was presumed to detect an obstacle if it was visible and in the same lane when the gap between HGV and obstacle was at least 150 m. The analysis included all road types although the majority of crashes occurred on motorways.

Petersen et al. (2020) analysed rear-end collisions with severe personal injury by HGV on motorways in Lower Saxony and concluded that 23 out of 40 rear-end collisions of HGV without AEBS could have been prevented if the vehicle had had AEBS. This corresponds to 58%. The high effectiveness of these systems has also been found in international studies. Jermakian (2012) estimated that 37% of all rear-end crashes of large trucks between 2004-2008 could have been prevented in the US if the vehicle had had a forward collision warning system. In 83% of the crashes that were assessed as preventable the driver did not brake. For non-fatal injury and fatal rear-end crashes respectively, the author estimated a decline by 29% and 17%.

Woodrooffe et al. (2013) forecasted the effect of an autonomous emergency braking systems (collision-mitigation braking and forward-collision warning) that detects stationary and moving targets to 28-40% for the US.

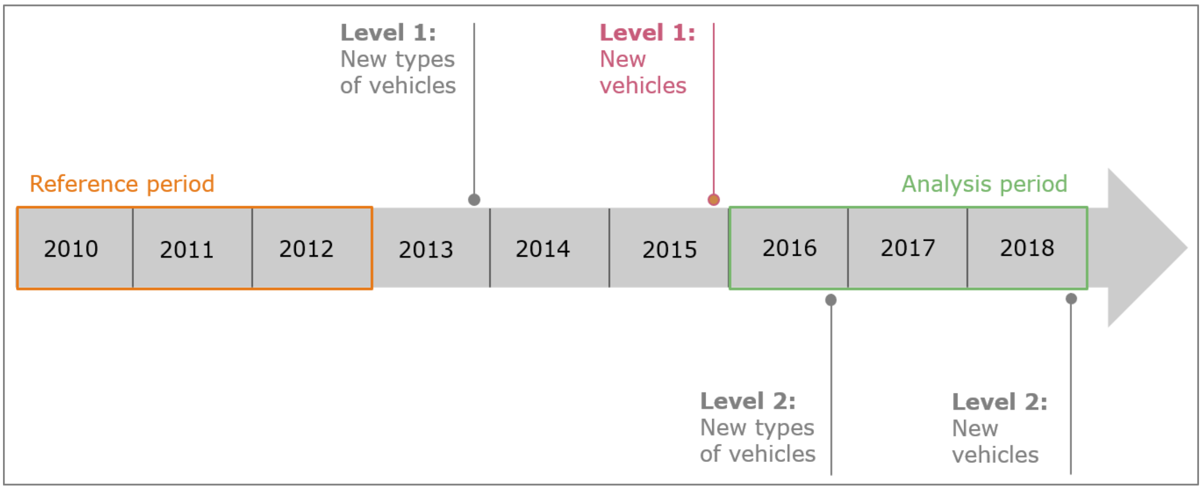


Figure 1: Timeline.

These studies estimated the potential benefit of AEBS. They all assumed that the system was activated at all times and that the driver always reacted appropriately. Petersen et al. (2020) and Seiniger et al. (2020), however, pointed out that drivers could instinctively react to an imminent collision by braking or steering and thus inadvertently overriding the system.

To factor in potential shortcomings in the interaction between AEBS and driver, an ex-post analysis of real-world crashes is necessary. For the US, Teoh (2020) analysed crashes of large trucks on motorways between 2017 and 2019 and found that the crash rate of rear-end collisions of trucks equipped with automatic emergency braking systems had been reduced by 41%. The analysis of Petersen et al. (2020) also suggests that HGV with AEBS were disproportionately less involved in rear-end crashes on motorways in Lower Saxony in Germany than those vehicles without AEBS.

An investigation of the ex-post effectiveness of AEBS in Germany is missing so far. This analysis is supposed to close this gap. This paper examines the effect on HGV crashes associated with the EU-Regulation.

The article thus addresses two research questions:

How did the number of rear-end collisions of HGV and buses develop after the Regulation was introduced?
Did the severity of rear-end crashes change after the implementation of the Regulation?

These questions were analysed using individual data of police records from 2010 to 2018. The remainder of the article is organised as follows: Section 2 illustrates the details of the Regulation. Section 3 describes the data and method used to analyse the effects associated with the Regulation. Section 4 presents the results. In section 5, the findings and limitations are discussed. Section 6 concludes.

2. BACKGROUND

In the European Union, AEBS are mandatory for new HGV and buses with a few exceptions. The details are regulated in EU-Regulation No. 347/2012 (European Commission 2012a).[1] The Regulation

[1] Identical regulations were issued at UNECE level (UNECE 2013).

came into effect at two approval levels each first for new types of vehicles and two years later for new vehicles. The first approval level became effective for new types of vehicles on 1st November 2013 and for new vehicles on 1st November 2015. The second approval level came into force for new types of vehicles on 1st November 2016 and for new vehicles on 1st November 2018. There is no obligation to retrofit already registered (older) vehicles.

At the first approval level, AEBS were obligatory for vehicles of the following categories as defined in Directive 2007/46/EC (European Commission 2007):

- Category M3: Vehicles designed and constructed for the carriage of passengers, comprising more than eight seats in addition to the driver's seat, and having a maximum permissible gross vehicle weight exceeding 5 tonnes.
- Category N3: Vehicles designed and constructed for the carriage of goods and having a maximum permissible gross vehicle weight exceeding 12 tonnes.
- Category N2 (AEBS obligatory only if exceeding 8 tonnes): Vehicles designed and constructed for the carriage of goods and having a maximum permissible gross vehicle weight exceeding 3.5 tonnes but not exceeding 12 tonnes.

At the second approval level, AEBS became additionally mandatory for vehicles of these categories:

- Category M2: Vehicles designed and constructed for the carriage of passengers, comprising more than eight seats in addition to the driver's seat, and having a maximum permissible gross vehicle weight not exceeding 5 tonnes.
- Category N2 ≤ 8 tonnes.

Exempted from the Regulation are vehicles with more than three axles, off-road vehicles, certain special purpose vehicles, semi-trailer trucks with a maximum permissible gross vehicle weight not exceeding 8 tonnes and buses in which the carriage of standing passengers is possible. The specifications of the Regulation are summarised in Table 1.

Table 1: EU-Regulation No. 347/2012.

Approval level	Type of vehicles	Categories of vehicles*	Effective as of
1	New types of vehicles	M3, N3, N2 > 8 t	01/11/2013
	New vehicles	M3, N3, N2 > 8 t	01/11/2015
2	New types of vehicles	M2, M3, N2, N3	01/11/2016
	New vehicles	M2, M3, N2, N3	01/11/2018

* Except vehicles > 3 axles, off-road vehicles, certain special purpose vehicles, semi-trailer trucks > 8 t, buses in which the carriage of standing passengers is possible.

Source: European Commission (2012).

AEBS detect collision risks using camera, radar and/or laser. If the system detects a collision risk, it alerts the driver. The system issues a second warning if the driver does not react to the first warning by braking or switching lanes. During this warning phase, the system reduces the speed of the vehicle only slightly. If the driver does not respond to the alerts, the system initiates the emergency braking phase. Towards a stationary target, the system had to reduce its speed by 10 km/h at the first approval level and by 20 km/h at the second approval level.[2] If the target is moving, the collision must be prevented regardless of the approval level.

To avoid confusion due to false warnings, the driver can manually deactivate the system. The driver can furthermore override the system during an emergency brake by accelerating for instance.

3. DATA AND METHODS

Individual German police records from 2010 to 2018 was used for this analysis. Additionally, technical vehicle information was used that the German Federal Motor Transport Authority (KBA) provided.

3.1. Study object

As illustrated in section 2, the Regulation became effective at two approval levels first for new types of vehicles and two years later for new vehicles. This study examines the effect of approval level 1 for new vehicles. The reason for that is twofold. First, there is only information available on the year in which the vehicle was originally registered. The date of the vehicles' type approval cannot be derived from the data. Second, since approval level 2 did not come into force for new vehicles until the end of 2018, there was simply not yet (sufficient) data to analyse this effect.

As AEBS became obligatory by the end of 2015 for new vehicles, the analysis period began in 2016 and ended in 2018 – the last year for which data was available. The number of crashes in this period was compared to a reference period from 2010 to 2012.

[2] For vehicle of categories M2 and N2 < 8 t, a speed reduction of 10 km/h was sufficient at the second approval level.

AEBS for new types of vehicles were only introduced by the end of 2013. Thus, between 2010 and 2012, there was no obligation for HGV or buses to be equipped with AEBS and the number of crashes was not influenced by the measure. Both, analysis and reference periods represent the examination period.

Figure 1 shows the timeline of the Regulation's introduction and the periods under examination.

3.2. Vehicles included in the study

At approval level 1, AEBS were – with some exceptions – mandatory for vehicles of categories M3, N3, N2 > 8 t. Information on the category of vehicles involved in a crash is not included in the data. Thus, the category was derived using the variables "category of road user" and "maximum permissible gross vehicle weight". To avoid confounding, the categories were narrowly defined. Meaning, the analysis only included those vehicles that most likely were subject to the Regulation. The variable "category of road user" can take on 37 different values. Fourteen values refer to HGV. Of these, nine values were included in the analysis. These vehicles most likely comply with the definition of the Regulation. The other values indicate for instance off-road vehicles, special purpose vehicles or vehicles with a maximum permissible gross vehicle weight lower than 8 t. Hereinafter, the term "HGV" refers only to those vehicles that have been included in the study.

The analysis includes only one value for buses from the five values that refer to this kind of vehicle. All other values also contain vehicles that carry standing passengers and were thus excluded to get a narrowly defined group. Henceforth, the term "bus" thus indicates only those vehicles included in the study.

The technical vehicle data contains information on maximum permissible gross vehicle weight. However, this data is only available for a subset of crashes. In 2018, for example, this data was missing for 20.5% of all HGV involved in a crash on motorways. It is mainly missing if the vehicle was registered abroad. In 2018, this applied to 15.5% of HGV involved in a crash. Data on maximum permissible gross vehicle weight is also not available if the driver fled the scene of the crash or if the police officer recorded the wrong licence number. The study excluded crashes with missing technical vehicle information. This ensured that the analysis only took vehicles into account which were covered by the Regulation. Additionally, all vehicles with more than three axles were excluded from the analysis.

3.3. Crashes included in the study

The analysis was limited to crashes on motorways. This had several reasons. First and foremost, crash situations on motorways are generally less complex. It is thus easier to identify the situations in which AEBS (would have) had an effect on the outcome. Second, AEBS were originally designed for motorways (Seiniger et al. 2020). Field tests showed that false warnings of AEBS on motorways could only be observed near motorway construction sites (ibid.). Thus, drivers might be less inclined to deactivate the systems while driving on motorways compared to other road types. Third, most rear-end collisions occur on motorways in Germany. According to Panwinkler (2018), around two thirds of all rear-end collisions were located on motorways. Rear-end collisions on motorways are typically also more severe than those on other road types.

The analysis considered all crashes with casualties (killed or injured) as well as crashes with property damage only and with at least one vehicle that is not roadworthy.

Furthermore, the analysis included only those crashes in which the party mainly responsible for the crash was an HGV or bus. In rear-end collisions, police record the vehicle that crashes in the vehicle in front of it as the mainly responsible party. AEBS in this vehicle could have helped prevent or mitigate the impact.

3.4. Definition of treatment and control groups

The analysis is conducted using one treatment and all in all three control groups. These were distinguished by two criteria: vehicle age and crash scenario. Rear-end collisions of new HGV and buses formed the treatment group (Treatment). Rear-end collisions by older HGV and buses served as a first control group (Control I). Other crashes of new HGV and buses acted as a second control group (Control II). Other crashes of older HGV and buses were the third control group (Control III). The latter group was only needed to calculate the overall effect of the measure.

The definition of the groups is summarized in Table 2. Below, the criteria that have been used to define the groups are described in more detail.

Table 2: Definition of treatment and control groups.

		Crash scenario	
		Rear-end collision	Other crash
Age of vehicle	New vehicle	Treatment group	Control (I) group
	Older vehicle	Control (II) group	Control (III) group

A vehicle was defined as new if it was first registered during the period under examination. Thus, in the reference period, the registration was between 2010 and 2012. Equivalently, the registration of a new vehicle was between 2016 and 2018 in the analysis period. Consequently, in the first year of each period, only vehicles registered in the same year were considered. For example, in the analysis period in 2016, only vehicles registered in 2016 were defined as new. In the third (and last) year of each period, vehicles registered in the same year and in the two preceding years were defined as new. In the analysis period in 2018, for instance, those vehicles registered between 2016 and 2018 meet the criteria of the definition (Figure 2). There are two reasons for this differentiation. First, the definition had to be constructed in a way that all vehicles in the treatment group must be equipped with AEBS. Second, sufficient data for the analysis was required.

A vehicle was perceived as older if it was registered six years before a new vehicle. In the first year of the period under examination, only vehicles registered in one year satisfy the criteria. For instance, in the analysis period in 2016, these were vehicles registered in 2010. In the third year, the definition covers vehicles six to eight years old – thus, vehicles registered between 2010 to 2012 in the analysis period in 2018. This definition prevented confounding due to differently defined age groups. Otherwise, crash years would have been weighted differently in the age groups and external factors like cold or heat waves could have influenced the results.

There is no flag for rear-end crashes in the accident data. Therefore, the combination of the variables "type of accident" and "kind of accident" were used to identify rear-end collisions. The type of accident describes the situation which leads to the crash. The kind of accident specifies the course of the crash, e.g. the collision or involuntary leaving the road.

Rear-end collisions were defined as the combination of driving accidents or accidents moving along in carriageways (type of accident) and collisions with a vehicle moving ahead or waiting (kind of accident). This definition was based on the classification in Panwinkler (2018).

Crashes that were not considered as rear-end collisions were defined as other crashes – albeit with some exceptions. A lane departure warning system became mandatory in HGV and buses at the same time as approval level 1 took effect (European Commission 2012b). To avoid confounding, lane departure crashes were excluded from the analysis. These were defined as the combination of driving accidents or accidents moving along in carriageways (type of accident) and collisions with vehicles moving parallel in the same direction or leaving the carriageway to the right or left (kind of accident). Additionally, collisions with (voluntary) stopping vehicles (e.g. for parking) and collisions with oncoming vehicles were exempted from the analysis. These crashes are special cases when located on motorways and could have potentially confounded the results. The same applies to single vehicle crashes that were also excluded from the analysis.

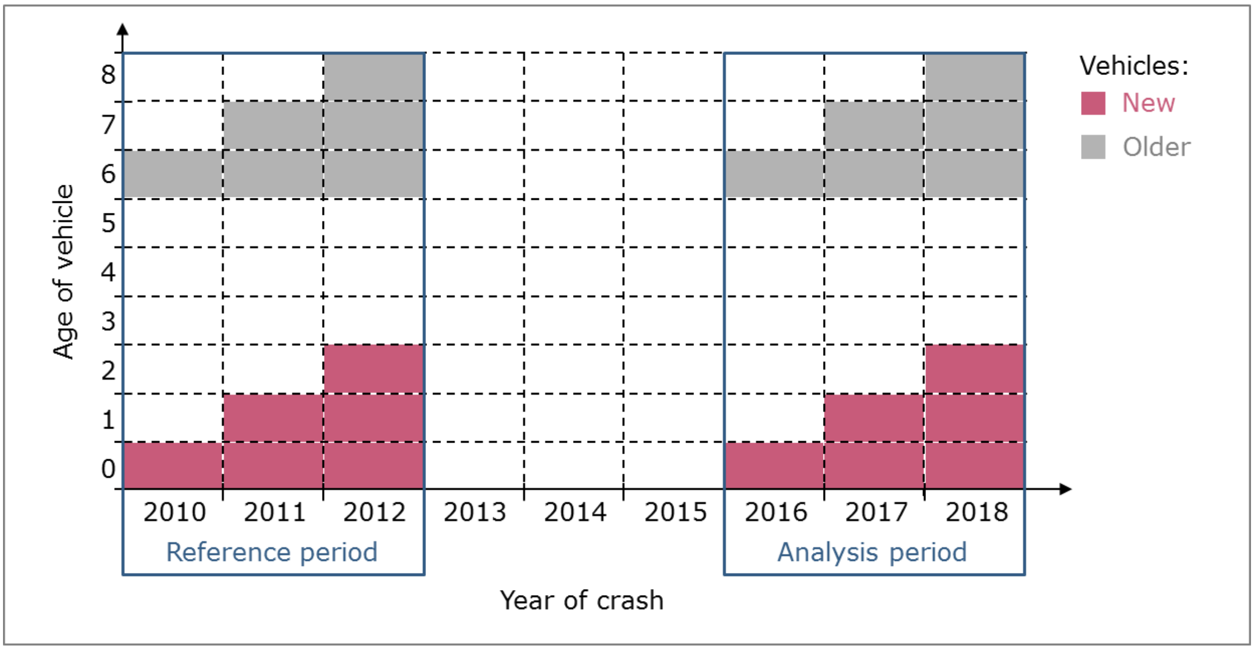


Figure 2: Definition of new and older vehicles.

The hypothesis was that the Regulation was associated with an effect on the number of crashes of HGV and buses on motorways. Odds ratios (OR) and chi square tests were computed to test this hypothesis. To estimate the overall effect of the system in AEBS-relevant crash situations, changes between reference period and analysis period in both variables – crash scenario and age of vehicle – were considered. This corresponds to the interaction term. It is the same as the ratio of the OR of rear-end collisions in the reference and analysis period and the OR of other crashes in both periods. In this way, the model allows for changes in the composition of crash scenarios (e.g. shifts in situations that could lead to rear-end collisions but no such variations for other crashes) as well as changes in the composition regarding vehicle age (e.g. different developments in the kilometres driven on motorways by older vehicles).

AEBS might affect more severe and less severe crashes disproportionally. The analysis was thus repeated for different levels of crash severity separately to test the hypothesis if the Regulation was associated with changes in the severity of rear-end collisions.

4. RESULTS

This section presents the results. In sub-section 4.1, the changes in the number of rear-end collisions associated with the Regulation were analysed. In sub-section 4.2, it was examined if the severity of rear-end collisions changed after the implementation of the Regulation.

4.1. Changes of the level of rear-end collisions

The treatment group developed more positively than the control groups. The number of rear-end collisions of new HGV and buses decreased by -13.8 % between 2010-2012 and 2016-2018 (Treatment). In contrast, the number of rear-end collisions of older HGV and buses increased by about 13.1% (Control I). The odds ratio is significantly lower than 1 (OR = 0.762, 95% CI = 0.591, 0.982) (upper half of Table 3). The chi square test statistic is 4.402 (p = 0.036). These results suggest that the number of rear-end collisions of new HGV and buses developed significantly better than those of older HGV and buses. The hypothesis that there is no difference in the development between treatment group and control group I can thus be rejected.

The number of other crashes of new HGV and buses grew by 12.6% between the reference and the analysis period (Control II). The odds ratio between treatment group and control group II is again lower than 1 (OR = 0.766, 95% CI = 0.595, 0.988) (lower half of Table 3). The chi square test statistic is 4.235 (p = 0.040). These findings suggest that the number of rear-end collisions by new HGV and buses dropped significantly more than the number of other crashes. The hypothesis that the development of the treatment group is the same as of control group II can therefore be rejected.

The ratio of the OR of rear-end collisions and of other crashes is 0.630 (95% CI = 0.420, 0.945). This corresponds to an overall effect associated with the Regulation of -37.0% on the number of rear-end collisions on motorways.

Table 3: Results.

Crash scenario	Vehicle age	Group	Period		Trend	Odds ratio (95% CI) Treatment vs. Control
			Reference	Analysis		
Rear-end	New	Treatment	311	268	-13.8%	
	Older	Control I	191	216	13.1%	0.762 (0.591, 0.982)
Other	New	Control II	191	215	12.6%	0.766 (0.595, 0.988)
	Older	Control III	129	120	-7.0%	-
Interaction effect						0.630 (0.420, 0.945)
Effect of measure						-37.0%

BASt-U2l-6/2021

Table 4: Crash severity.

Crash severity	Crash scenario	Vehicle age	Group	Period		Trend	Odds ratio (95% CI) Treatment vs. Control
				Reference	Analysis		
Severe injury (fatal/seriously injured)	Rear-end	New	Treatment	99	66	-33.3%	
		Older	Control I	50	74	48.0%	0.450 (0.289, 0.724)
	Other	New	Control II	27	33	22.2%	0.545 (0.300, 0.990)
Slight injury	Rear-end	New	Treatment	162	169	4.3%	
		Older	Control I	114	127	11.4%	0.936 (0.672, 1.305)
	Other	New	Control II	72	64	-11.1%	1.174 (0.787, 1.750)
Material damage only	Rear-end	New	Treatment	50	33	-34.0%	
		Older	Control I	27	15	-44.4%	1.188 (0.551, 2.564)
	Other	New	Control II	92	118	28.3%	0.515 (0.307, 0.863)

BASt-U2l-6/2021

4.2. Changes in the severity of rear-end collisions

In addition to the effect on the total number of crashes, the analysis also investigated the effect on the severity of crashes. The crash's severity level is assigned based on the most severe consequence – thus, a crash with slightly and seriously injured persons is counted as a serious injury crash. Due to the otherwise small number of cases, fatal crashes and serious injury crashes were combined into one category (severe injury crashes). For crashes with property damage only, solely those crashes were considered in which at least one vehicle was not roadworthy.

The number of crashes in the treatment group with severe injury decreased considerable after the Regulation came into effect compared to before (-33.3%). In control group I and II, on the other hand, the number increased substantially (48.0% and 22.2%). The odds ratio between treatment and control group I and treatment and control group II, respectively, were both significantly below 1 implying that the number of crashes had been reduced after the Regulation was implemented (Table 4). The null hypothesis that there was no change in the number of crashes with severe injury associated with the Regulation can be rejected.

For crashes with slight injury, the results are less clear. In the treatment group as well as in control group I, the number of crashes grew slightly. In control group II, in contrast, the number of crashes sank slightly. No odds ratio is significantly different from 1. The null hypothesis cannot be rejected.

The number of crashes with property damage only declined drastically in the treatment group and in control group I (-34.0% and -44.0%). In control group II, the number rose substantially (28.3%). The odds ratio of treatment vs. control group II is significantly lower than 1 while the odds ratio of treatment vs. control group I does not deviate significantly from 1. For crash scenarios, the null hypothesis of no change in the number of crashes with property damage only after AEBS became mandatory can thus be rejected. For vehicle age, the null hypothesis cannot be rejected.

The declining numbers of crashes in AEBS situations could thus be primarily attributed to the decreasing number of crashes with severe injury. These results suggest that the Regulation is associated not only with a decline of the overall number of crashes but also with a reduced severity of crashes. However, especially for crashes with property damage only, the numbers are rather low. These results should therefore be interpreted with caution.

5. DISCUSSION

The level of crashes declined after the Regulation came into effect for those vehicles for which AEBS became mandatory. This reduction is significant compared to control groups. The analysis further suggests that not only the number of crashes declined but also the severity of crashes decreased for the vehicles affected.

To check whether the groups are substantially different, the frequency distribution across selected variables were examined. There were no structural differences detected between new and older vehicles with regard to age of driver, time of crash, crash severity and proportion of crashes in close proximity to construction sites. Rear-end collisions and other crashes also did not differ substantially with regard to age of driver and proportion of crashes in close proximity to construction sites. As seen in section 4.2, rear-end crashes were more likely severe than other crashes. Rear-end collisions were also more prevalent during weekdays than other crashes. However, the great majority of HGV and bus crashes occurred during weekdays – regardless of crash scenario they made up more than 90% of crashes. Even if crashes on weekends were excluded from the analysis, the number of crashes had been reduced significantly more in the treatment than in the control group.

The kilometres travelled on motorways by HGV registered in Germany were approximately constant in the period under consideration (BAG 2019). The vehicle fleet of HGV increased substantially between 2012 and 2018 with a higher growth rate for older vehicles (+29%) than for new ones (+10%) (KBA 2012, 2018). However, it is not clear whether this translates into a higher proportion of kilometres travelled for older vehicles. Unfortunately, kilometres travelled differentiated by age of vehicle are not collected on a regular basis. A study conducted in 2014 analysed this aspect (Bäumer et al. 2017). The authors did not distinguish between road types however. They found that vehicle kilometres of new HGV were 1.5 times higher than those of older HGV. However, differences in the development of the vehicle fleet only affect the comparison concerning vehicle age not those regarding crash scenario. Since both variables were combined when calculating the overall effect, changes in kilometres travelled would have had no effect on this value.

The analysis period and the treatment group were constructed in a way so that 100% of the vehicles in this group and in this period were equipped with AEBS. In all other constellations between period and group, AEBS were not obligatory. However, even before the Regulation, manufacturers equipped vehicles with AEBS. Thus, in all groups in the reference period as well as in the control groups in the analysis period, equipment with AEBS was higher than 0%, but below 100%. Unfortunately, no data is available on the percentage of HGV and buses not affected by the Regulation that were nevertheless equipped with AEBS. Thus, the effect estimated in this study was not the effect that came with equipping the vehicle fleet with AEBS. Rather, it was the effect on the number of crashes associated with the introduction of the Regulation and thus the increased equipment of the fleet.

Additionally, requirements of approval level 1 were rather moderate compared to today’s state of technology. Some vehicles in the treatment group were already equipped with more efficient AEBS. The effect estimated in the study is thus at least partly associated with these more efficient systems. This should be kept in mind when interpreting the results. The proportion of vehicles already equipped with AEBS of approval level 2 or more is unknown.

Furthermore, the data provides no information if the system was activated during the crash. If a substantial number of drivers would have deactivated the system, the true effect of the Regulation had been larger than the effect estimated in this study. However, in their study, Seiniger et al. (2020), examined the relevance of false warnings. They tested the systems in real traffic over a distance of 1,300 km and found that false warnings were rather uncommon on motorways apart from construction sites. Additionally, in a survey of HGV drivers in Germany conducted by Trabert et al. (2018), the majority of respondents (92%) stated that they deactivated the system rarely or never. However, this result is based on a sample of only 85 drivers. The German Federal Ministry of Transport and Digital Infrastructure plans to ban the possibility of deactivating the system in the future BMVI (2020).

During the period under examination, lane departure warning systems became mandatory. To avoid confounding, the analysis excluded lane departure crashes. There were no other improvements regarding safety systems in HGV that could have had an effect on the number of crashes.

Overall, the results suggest that the Regulation was associated with a positive effect on the number of rear-end crashes on motorways. Future research is needed on the percentage of AEBS equipment in the entire vehicle fleet of HGV and buses. Furthermore, the effectiveness of AEBS on other road types is unknown – although for this kind of analysis, more information on the frequency of system deactivation is needed. A comparison with the effectiveness in other European countries can also be useful.

6. CONCLUSION

In this paper, the effect associated with EU-Regulation No. 347/2012 was analysed. This Regulation made advanced emergency braking systems (AEBS) mandatory for heavy good vehicles (HGV) and buses. It came into effect at two approval levels each first for new types of vehicles and two years later for new vehicles.

This study examined the effect associated with the first approval level for new vehicles which was effective since November 2015. The paper addressed two research question: (1) How did the number of rear-end collisions of HGV and buses develop after the Regulation was introduced? (2) Did the severity of rear-end crashes change after the implementation of the Regulation?

The analysis used individual police records of HGV and buses between 2010 and 2018 and additional technical vehicle information by the German Federal Motor Transport Authority (KBA). Rear-end crashes of new vehicles formed the treatment group. The number of crashes in this group was compared to three control groups: (I) rear-end collisions by older HGV and buses; (II) other crashes of new HGV and buses; and (III) other crashes of older HGV and buses. These groups were observed in two periods: a reference period before AEBS became mandatory in HGV and buses and an analysis period after the introduction of approval level 1. The analysis was limited to crashes on motorways because crash situations on motorways are less complex. Due to missing technical vehicle information on foreign vehicles, the analysis was further limited to vehicles registered in Germany. Odds ratios and chi square tests were used to estimate the impact associated with the Regulation.

The number of crashes in the treatment group developed more positively than in the control groups. The overall effect associated with the measure was estimated to -37.0%. This effect is significant. Additionally, the results suggest that the severity of crashes declined disproportionately in the treatment group.

Since the treatment group was narrowly defined (all vehicles had to be equipped with AEBS, only rear-end collisions), the magnitude of the effect seems plausible. It should be mentioned, however, that the effect associated with approval level 1 of the Regulation is likely to be slightly overestimated in the analysis. Requirements of this level were relatively low compared to today's state of technology. The treatment group thus most likely also included vehicles that were already equipped with more efficient systems. Nevertheless, the findings imply that there was a substantial decline in the number of crashes associated with the Regulation which corresponded to an accelerated AEBS equipment of the vehicle fleet. AEBS can thus contribute to road safety.

REFERENCES

[1] BAG (2019): Mautstatistik. Jahrestabellen 2018. Köln.

[2] Bäumer, Marcus; Hautzinger, Heinz; Pfeiffer, Manfred; Stock, Wilfried; Lenz, Barbara; Kuhnimhof, Tobias; Köhler, Katja (2017): Fahrleistungserhebung 2014 - Inländerfahrleistung. Bremen: Fachverlag NW (Berichte der Bundesanstalt für Straßenwesen Verkehrstechnik).

[3] BMVI (2020): We will prohibit switching off advanced emergency braking systems for HGVs to improve road safety. Online verfügbar unter https://www.bmvi.de/SharedDocs/EN/Articles/StV/Roadtraffic/emergency-braking-systems.html, zuletzt geprüft am 15.02.2021.

[4] European Commission (2007): Directive 2007/46/EC of the European Parliament and of the Council of 5 September 2007 establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles (Framework Directive).

[5] European Commission (2012a): Commission Regulation (EU) No 347/2012 of 16 April 2012 implementing Regulation (EC) No 661/2009 of the European Parliament and of the Council with respect to type-approval requirements for certain categories of motor vehicles with regard to advanced emergency braking systems.

[6] European Commission (2012b): Commission Regulation (EU) No 351/2012 of 23 April 2012 implementing Regulation (EC) No 661/2009 of the European Parliament and of the Council as regards type-approval requirements for the installation of lane departure warning systems in motor vehicles.

[7] Jermakian, Jessica S. (2012): Crash avoidance potential of four large truck technologies. In: *Accident Analysis & Prevention* 49, S. 338–346. DOI: 10.1016/j.aap.2010.10.033.

[8] KBA (2012): Fahrzeugzulassungen (FZ). Bestand an Nutzfahrzeugen und Kraftfahrzeuganhängern nach technischen Daten (Größenklassen, Motorisierung, Fahrzeugklassen und Aufbauarten). FZ 25. Flensburg.

[9] KBA (2018): Fahrzeugzulassungen (FZ). Bestand an Nutzfahrzeugen, Kraftfahrzeugen insgesamt und Kraftfahrzeuganhängern nach technischen Daten (Größenklassen, Motorisierung, Fahrzeugklassen und Aufbauarten). FZ 25. Flensburg.

[10] Kuehn, Matthias; Hummel, Thomas; Bende, Jenoe (2011): Advanced driver assistance systems for trucks - benefit estimation from real-life accidents. In: *Proceedings of the 22nd International Technical Conference on Enhanced Safety of Vehicles*.

[11] Panwinkler, Tobias (2018): Unfallgeschehen schwerer Güterkraftfahrzeuge. Bremen: Fachverlag NW (Berichte der Bundesanstalt für Straßenwesen M, Mensch und Sicherheit, Heft M 277).

[12] Petersen, Erwin; Scholze, Carmen; Falke, Christoph (2020): Notbremsassistenzsysteme im Lkw. Eine Analyse niedersächsischer Autobahnunfälle in den Jahren 2015 bis 2019 und der Einfluss aktueller Systeme (Advanced Emergency Braking Systems for Trucks – an analysis of truck accidents on Lower Saxonys motorways from 2015 to 2019 and the influence of actual AEBS) (4), S. 245–256.

[13] Seiniger, Patrick; Heinl, Felix; Bühne, Jan-André; Gail, Jost (2020): Lkw-Notbremsassistenzsysteme. Bremen: Fachverlag NW (Berichte der Bundesanstalt für Straßenwesen F, Fahrzeugtechnik, Heft F 133).

[14] Teoh, Eric R. (2020): Effectiveness of front crash prevention systems in reducing large truck crash rates.

[15] Trabert, Tobias; Elio, Iryna; Müller, Gerd; Malczyk, Axel (2018): In-depth Analyse schwerer Unfälle mit schweren Lkw: Unfallforschung der Versicherer (Nr. 54). Online verfügbar unter https://digital.zlb.de/viewer/metadata/34092585/1/-/.

[16] UNECE (2013): Regulation No 131 of the Economic Commission for Europe of the United Nations (UN/ECE) — Uniform provisions concerning the approval of motor vehicles with regard to the Advanced Emergency Braking Systems (AEBS).

[17] Woodrooffe, John; Blower, Daniel; Flannagan, Carol A. C.; Bogard, Scott E.; Bao, Shan (2013): Effectiveness of a Current Commercial Vehicle Forward Collision Avoidance and Mitigation Systems. In: *SAE Technical Papers* 9. DOI: 10.4271/2013-01-2394.

EFFECTS OF CROWDING ON THE OPTIMAL SUBSIDY OF PUBLIC TRANSPORT

Tamás Strommer, junior scientific researcher
KTI Institute for Transport Sciences
Than Károly Street. 3-5., Budapest, 1119 Hungary
strommer.tamas@kti.hu

ABSTRACT

Subsidizing public transport companies is a worldwide used policy due to the scale economies and low specific travel costs of these kind of transport services. Frequent use of subsidies, however, does not mean uniformity, as there are many methods of subsidizing transport companies. Recent studies has shown that the amount of subsidy applied varies greatly depending on the attributes of the city, service provider and transport mode. The purpose of the present study is to reconsider a model that optimizes fares, and modify the model's computation method by using new parameters that simplifies the management of crowding costs. The study presents the theoretical background of the topic, the important modifications made in the model, and also demonstrates usability of the model by examining the pricing of the Hungarian interurban transport system. The results recalculated with the modified model showed that crowding cost is a critical component of optimal pricing, as even at a moderate congestion level travel costs are increased by 30–50%. This means, that the optimal fares aimed to reduce the costs of crowding are higher than they were previously set. In case of Hungarian interurban transport, a subsidy rate of around 90% should be applied, with the exception of peak period bus transport, for which the calculated subsidy rate is 74%.

Keywords:
aggregate analysis, pricing policy, vehicle-size optimization, transit fares optimization, public transport regulations

1. INTRODUCTION

The extreme increase in the number, size and population of cities and megalopolises raises serious questions about their livelihood and sustainability. These questions can only be answered through prudent and watchful urban planning, by reaching the sufficient level of human flexibility (e.g. in housing, mobility), and improving the quality and efficiency of public and social services. The transport system can be able to compensate the shortcomings of urban design by providing high quality mobility services.

The smooth operation of the transport system is a cornerstone for the well-functioning cities and megalopolises. Since the recognition of this idea, greater attention has been focused on researching optimal planning and operational strategies for tasks of transport planning and management. Based on the commonly used microeconomical approach, travel pricing is a well-established and frequently used method to influence traveller decisions.

Our study examines the optimal subsidy of public transport operators. Although the daily operation of these companies are often supported by public funds, the accurate definition of subsidy—a more precise and easily usable formulae of the usual rate—is still missing. The appropriate rate of subsidies and the suitable financing method of public transport is important, because of these services' attractiveness and reliability—since both feature is strongly dependent on financial resources, and they are some of the key aspects of the transportation process. (The substantially unprofitable operation at a social level can also bring optimal welfare conditions, therefore subsidies are essential to the operation of transportation companies.) However, based on the results of recent studies, significant (proportional) differences can be observed between subsidies granted to companies providing services in similar circumstances (see for example the results of Boss and Rosenschon 2008; Doll and van Essen 2008; Parry and Small 2009; Tscharaktschiew and Hirte 2011). Although it is hard to draw far-reaching conclusions because of the variance between the compared systems' and countries' transport policy, economic situation, and the limits of subsidization, the topic of subsidies given to transportation operators functioning in similar circumstances is still worth examining.

The basis of our research was the computational model presented by Parry and Small (2009), which investigates the (ideal) pricing of travel. The model can be used to determine the subsidy rate and the expected operational-economic characteristics belonging to it under optimal conditions. The aim of our study is to extend this model with parameters describing crowding in public transport, and also the application of the model for a new context, the Hungarian interurban public transport system. The mode and purpose of adding new parameters to the model is to make the model able to handle both new and already-built-in parameters in the same environment, and thus be able to account for more detailed and accurate results of the effects of crowding. (Nowadays crowding is studied from many aspects from engineering and economics to psychology and ergonomics).

The paper is structured as follows: first the paper deals with the optimal pricing of public transport—the most influential factors of the calculation, and the basics of the model. The following sections introduce the needs that call for the modifications of the model, the use of crowding parameters and the further use of optimized parameters. Then the paper presents the basic data for the calculation of interurban traffic in Hungary and also the application of the model for the examined region. Finally we summarize all the results and conclusions that can be drawn from the results of the model and we suggest some further improvements.

2. OPTIMAL PRICING OF PUBLIC TRANSPORT

The benefits of public transport and the positive effects (e.g. lower local emissions, moderate fuel consumption and land use) are most dominant in networks with significant traffic and near bottlenecks. Even when utilizing the benefits of public transport, it is advisable to strive for an optimal operational situation and strategy, where the benefits of the daily use are not outweighed by the inherent negative phenomenon. All systems should avoid the state, when:

- the presence of public transport vehicles cause a major congestion in the road traffic of the city (or even cause significant delays of the public transport service itself),
- the level of crowding on public transport vehicles grow so high, that it causes a major decrease of utility, or when
- the contamination of vehicles and the decrease of passenger safety cannot be avoided (see the results of Perone 2002 on passenger preferences and mode choices).

2.1. Scale economies of scheduled transport services

In the case of public transport systems operating on a frequent basis and providing scheduled services, the phenomenon of scale economies prevails (Mohring 1972, 1976; Turvey and Mohring 1975). The additional operation costs of passenger services at higher passenger numbers, under idealized conditions, are balanced by the benefits of passenger time savings through the reduction of the average waiting time, i.e. the net marginal social cost of travel is lower than the marginal cost of operation. In other words, the marginal cost of using a service that operates under scale economies is always lower than the average cost, and therefore the optimal price, which is equal to the marginal cost, leads to a loss that must be compensated with subsidy.

Regarding the optimization of the social costs, the sum of operating and user costs, by the means of service frequency, we obtain simple relation that the optimal frequency—number of vehicles departing per unit of time—increases proportionally with the square root of the number of passengers, assuming a proportional relationship of vehicle amount and operating costs, and that passengers arrive to stops at random (Mohring 1972).

In addition, in public transport, due to fixed operating costs and waiting times, scale economies can also be observed in operating and user costs. Since both lead to subsidies, Mohring's model requires that the service should be subsidized even if the economies of scale in operation do not materialize, as the scale economy still applies to user costs.

2.2. Marginal cost pricing

Where user cost is a non-negligible part of social cost, as in the case of scheduled public transport services, the marginal cost based pricing should include two types of costs at the optimum, the service provider's and the passengers' costs (Jansson 1979; Vickrey 1980). This duality can be observed both in the amount of vehicles on the road, proportional with user costs and the size of the service area, and the size or capacity of the vehicles, proportional with the costs of the service provider.

The uniqueness of the transport market is that the consumer (passenger) is also involved in the process of producing the product, since the locomotion only can be performed by the passenger's time spent. It is precisely because of passenger involvement that the right approach is to escape the notion that the only costs which are relevant to optimization are those of the transport operator. The time-costs of the passengers must be included too, and fares must be equated with marginal social costs. (Turvey and Mohring 1975) The significant change brought by the proliferation of information and communication technologies (ICTs) also has to be highlighted, as it has made multitasking a realistic option during travel (though learning, working, various communication forms etc., see Keserű et al. 2015; Keserű and Macharis 2018; Munkácsy, Keserű, and Siska n.d.). This change made travel time almost as valuable as other activities—in an appropriate travel environment—the value of passengers' travel time savings could drop significantly (International Transport Forum 2018).

2.3. Overview of the model of Parry and Small

One of the significant works of the recent years on the field of transport economics, subsidization and pricing policy was the paper of Parry and Small (2009). The paper highlights the importance of the question whether the subsidy of public transport companies needs to be reduced (or even abolished), or increasing subsidy level brings the system closer to the optimal operating conditions.

The authors stated that despite the differences between the subsidy rates of transport companies, there is no generally accepted, practicable calculation method that can be used to determine the ideal financial strategy for a given transport operator, i.e. the rate of subsidy. In most cases, only complex, and therefore location-specific simulation models are able to provide data on traffic flow. Since pricing policies are often based on these results of modelling it is difficult to determine whether public transport pricing fits well with a city or region's transport system.

The aim of the authors was to create an analytical model that is simple enough to use without highly specific data, i.e. it requires data that are typically available at all transport providers, statistical officest etc., and at the same time the model can manage the most important parameters describing the transport system. It was also targeted that the calculation method could be applicable to transport systems of significantly different size, structure, operation method, and vehicle occupancy, as well as to cities and countries with different levels of development.

2.3.1. Model structure

The model was built using aggregated data that incorporate peak and off-peak results of all transport modes. The advantage of this model structure is that it can replace network models significant in size and computational supplies, their need for detailed calibration, the necessarily integrated decision methods of traffic models etc.. It is important to note that aggregate data are sufficient for this level of analysis, and it is neither necessary nor advisable to subdivide the data, as in many cases aggregated results can be obtained from local transport companies, road transport organizations, statistical offices, etc., or censuses, surveys have already done these researches—which also simplifies the collection of data and thus the investigation and modelling, too.

One of the model's most important methodological feature is that all parameters are derived to vehicle-miles and the calculations are defined by these kind of parameters, too. The key components of the model can be calculated using the following parameters:

- user benefits (consumption of numeraire good and the utility of travelling);
- vehicle occupancy, service frequency, crowding and travel time;
- external pollution and accident costs;
- household budget constraint, the balance of income and expenditure;
- ways in which companies adapt the constraints of the service.

The model calculates values of costs and benefits and then aggregates these results using passenger-miles "travelled" in the system—passenger-miles per travel mode and time periods. Similarly, in-vehicle travel time (T), waiting time (W), access time (A), and crowding measure (C) for the extra cost of vehicle-crowding can be calculated as the product of specific values multiplied by the number of passenger-miles travelled during the time period under investigation, calculated for each transport modes. A combination of these factors can be used to produce the generalized (non-money) cost of travel, which is proportional to the time spent traveling: $\Gamma = \Gamma(T, W, A, C)$. The Γ function establishes the relationship between the costs received in time dimension.

The way to determine utility is to sum all the components in cost dimensions, which can help balancing the benefits and losses. According to the model, the user preferences and the benefits of their activities (U) can be modeled as the difference between the value of the utility function—based on the consumption of numeraire good (X), the sub utility from passenger miles travelled (M), and the generalized (non-money) cost of travel—and factor Z, which takes the magnitude of pollution externalities and accidents costs into account. In addition to the utility function, the balance of the representative household must be determined. This approach assumes that all households spend the tax-deductible portion of their income on traveling and consuming numeraire good.

The indirect utility function of households is calculated as the value of the maximized utility function within the budget constraints, see the first line of Equation (1), but an indirect utility function of a user (passenger) can also be used to measure social benefits by replacing the transit agency's budget constraint. Then, the amount of tax (TAX) expressed as a function of operating costs and ticket revenue can be inserted in the second expression of Equation (1). In which the sum of operating costs (OC) can also be divided into fixed and variable (service-dependent and independent) costs.

$$\tilde{U} = \tilde{u}(\{p^{ij}, t^{ij}, w^{ij}, a^{ij}, c^{ij}\}, TAX) - Z = \max_{X,M,\Gamma} \begin{Bmatrix} u[X, M, \Gamma(T, \mathrm{W}, \mathrm{A}, \mathrm{C})] + \\ +\lambda \cdot \left(I - TAX - X - \sum_{ij} p^{ij} M^{ij}\right) \end{Bmatrix} = \max_{X,M,\Gamma} \begin{Bmatrix} u[X, M, \Gamma(T, \mathrm{W}, \mathrm{A}, \mathrm{C})] + \\ +\lambda \cdot \left(I + \sum_i \tau^{iH} M^{iH} - \sum_{ij \neq iH} OC^{ij} - X - \sum_i p^{iH} M^{iH}\right) \end{Bmatrix} \quad (1)$$

As it can be seen from the presented model structure, one can analyse the journeys by transport mode and travel time period. This makes easier to examine the journeys' components in detail, but because of the aggregate level an optimal strategy can also be determined based upon the same data.

2.3.2. *Formulas derived from the model*

The net marginal social cost of each trip is obtained by totally differentiating the indirect utility function of Equation (1) with $-p$. The extreme is where marginal reduction or increase of fares no longer increases or decreases social welfare. For example, one can determine the extremes of the indirect utility function for peak-period rail travels by differentiating Equation (1) with $-p^{PR}$. The result is Equation (2), which shows how the value of utility changes for each component as a result of a single reduction in the travel fees of peak-period rail transport.

$$MW^{PR} \equiv -\left(MC^{PR}_{supply} - p^{PR}\right)\left(-M^{PR}_{PR}\right) + +\left(MB^{PR}_{scale} - MC^{PR}_{occ}\right)\left(-M^{PR}_{PR}\right) + \sum_{ij=PR,iH}\left(MC^{ij}_{ext} M^{ij}_{PR}\right) + + \sum_{ij=OR,PB,OB}\left(MC^{ij}_{supply} + MC^{ij}_{ext} + MC^{ij}_{occ} - MB^{ij}_{scale} - p^{ij}\right) M^{ij}_{PR} \quad (2)$$

In Equation (2) the first component (MC^{PR}_{supply}) reduces social welfare, since the greater the difference between the cost of transporting a passenger and the price of a ticket, greater losses the transport agency accumulates. The effect of the transport system's scale economies impacts through the difference between the marginal benefit of scale economies (MB^{PR}_{scale}) and the marginal cost of congestion (MC^{ij}_{occ}). When the Mohring effect applies, a marginal reduction of fares has a positive outcome, since the operator increases the service frequency, thereby reduces the waiting time of all the other passengers. In contrast, crowding has an opposite effect. With the marginal reduction of fares the number of passengers will increase, and in parallel the increase of crowding will cause losses of well-being and social benefits. The remaining components of Equation (2) expresses the benefits of changing externalities (MC^{ij}_{ext}) due to the expected reduction in the number of cars, and the impact of lower peak-period rail travel fees on bus and off-peak-period rail travel characteristics (via scale economies, appearance of new passenger, crowding on vehicles etc.).

2.3.3. *Results of previous work*

Parry and Small examined the transport systems of three cities having drastically different transportation systems (London, Washington DC, and Los Angeles). The subsidy rates for each city were arbitrarily modified by time periods and transport modes, because a one time intervention in the model would have caused unrealistic change in service levels, as well as users' expected decisions and mode choices.

The paper's results show that in most cases significant rates of subsidies are recommended for transport agencies. In 11 out of 12 cases, the optimal value of subsidy is more than two-thirds of the operating cost, and in more than half of the cases it reaches 90%, while in one case the model suggested a minor reduction. The model proves that new passengers appearing would have a negative impact on utility due to the difference between average and marginal cost. The results of the model also show that "revenues" are primarily come from scale economies and decreasing externalities.

The model predicts a major increase in passenger miles, more than 50% increase in off-peak periods. This result is perfectly rational, since it is particularly practical to facilitate the best use of capital (i.e. fleet of vehicles, stations and infrastructure) in less frequented periods. With significantly lower off-peak period transit fares and due to greater spatial and temporal coverage, on account of the Mohring effect, more passengers will choose public transport.

Based on the examples of the examined cities, it can be concluded that in such cases public transport should be given a decisive role, and the use of public transport should be significantly supported. However, it is worth emphasizing that this is a secondary optimum. Not only in terms of pricing, as this model ignores the expected impact of other actions taken (e.g. infrastructure or vehicle development). The primary optimum could be achieved through appropriate pricing of car use, but this kind of pricing is deliberately omitted by the study.

3. EXAMINATION OF THE CROWDING ELASTICITY

Determining the values of the parameters built in the model, Parry and Small (2009) made a suspicious finding that the proportion of congestion losses are "negligible" compared to other factors. However, studies over the last 10 years have shown that perceived travel costs can rise by around 50%, or even more according to some measures at a crowding level of 3 passengers/m^2. In addition, the original model used a nonlinear relationship between crowding and value of time, which is an inappropriate simplification based on the results of the relevant literature. (Björklund and Swärdh 2015; Hörcher, Graham, and Anderson 2017; Kroes et al. 2014; Tirachini et al. 2016; Whelan and Crockett 2009)

It is difficult to determine the actual characteristics of crowding, therefore our approach trace it back to the travel time. When supplemented by a multiplier term the effect of crowding can be calculated with the formula $c^{ij} = t^{ij} m(l^{ij})$. Since the relationship between vehicle occupancy and crowding factor can already be approximated by a linear function (see the results of the above-mentioned literature), a multiplier function of the form $m(l) = \alpha_c \cdot l$ has been used, where l expresses the degree of vehicle occupancy. We assume that the level of crowding does not affect travel time (i.e. neglecting the relationship between the crowding level and the aligning time). Thus the elasticity of α_c depends solely on the shape of the multiplier function, which in this case, due to the linear relationship $\eta_c = (\partial m(l)/\partial l) \cdot (l/m(l)) = 1$ (see the results of the relevant measures (Björklund and Swärdh 2015; Hörcher et al. 2017; Kroes et al. 2014; Tirachini et al. 2016; Whelan and Crockett 2009).

The model of Parry and Small (2009) takes only the most important parameters of the transport system into account, but the model is still simpler, and uses less data than the detailed traffic demand models. The analytical relationships of the model counts for different modes of traffic (rail, bus, and private car), and uses a simple method of welfare optimization while maintaining economic equilibrium. As well as the optimum depends on the costs of travel, it relies on the consumption of the numeraire good, the passenger miles travelled in the whole system, and the model also calculates the effects of pollution, accidents and other externalities.

The model traces back cost variables (waiting, access, crowding) to travel characteristics, such as waiting time to service frequency, accessibility to route density, and it gives the crowding costs as a function of load factor. The model takes travel time patterns into account specialized for every mode of transport.

By the time of Parry and Small's research there were no available measurement with valid results to quantify the access costs and the effects of crowding on vehicles, therefore the paper replaced this indicators with some simplifications. The applied assumption was that the service provider (travel agency) adjusts the vehicle route density (D) and service frequency (f^{ij}) according to the changes of travel demand, while responds to the changing intensity of crowding by optimizing vehicle size (l^{ij}) and congestion level (o^{ij}). These optimization conditions take the following form (see Parry and Small 2009, p. 708):

$$\varrho^W w^{ij} \eta_w^{ij} = \varrho^A a^{ij} \eta_a^{ij}, \quad (3)$$

$$\varrho^C c^{ij} \eta_c^{ij} o^{ij} = t^{ij} k_2^{ij} n^{ij}, \quad (4)$$

where the factors ϱ^k express the marginal (monetary) cost of each type of loss, and η^k their elasticity ($k = W, A, C$).

These equations can be turned into the relation that route density can be increased as long as the decreasing costs of access, through the fixed quantity of passenger miles can balance the effect of the decreasing service frequency, and thus the increasing waiting costs. Similarly, the increase in vehicle size pays off as long as the gains from the reduction of crowding can cover the rising operating costs. Based on these relationships a generalized user cost gas been expressed that can summarize all factors as a function of (travel) time:

$$q^{ij} = p^{ij} + \varrho^T t^{ij} + \varrho^W w^{ij}(1 + \eta_w^{ij}/\eta_a^{ij}) + t^{ij} k_2^{ij} n^{ij}/\eta_c^{ij} o^{ij}. \quad (5)$$

In the next chapter, we introduce the suggested modifications and changes in the aforementioned relationships and optimization parameters.

4. OPTIMIZATION PROBLEMS

Crowding-related costs are difficult to measure and quantify in the way it was used in the model of Parry and Small. Based on assumptions confirmed by the literature we have modified and extended the original model using a linear multiplier function. These parameters and formulas measuring crowding can be used to optimize vehicle size and determine optimal fares. In this chapter we will present these questions, and also the related modifications in details.

4.1. Vehicle-size optimization

In the initial model, similarly to the other parameters, crowding is defined as a specific parameter of distance travelled, i.e. passenger miles. Despite the simplicity of this approach, in our opinion, the effects of crowding should rather be compared to travel time, since the extent of profit loss does not depend on the distance travelled, but rather on the duration of discomfort.

An important simplification of the original model is that the effects of crowding can be determined as a prerequisite for optimizing vehicle size with the balance of operating and crowding costs. According to this assumption, in order to increase passenger miles, the service provider must optimize the size of vehicles in such a way, that the resulting passenger-side benefits could compensate the increasing operating costs of larger vehicles. This relation would be a very practical solution, but its feasibility is highly questionable: it is not possible to react to the constant changes in passenger traffic by continuously changing the vehicle fleet, or intervening flexibly at a fixed infrastructure (e.g. by enlarging the stations of metro networks). Based on this consideration, it is advisable to abandon this equation and approach the question from the viewpoint of travel time.

The cost function was used by Parry and Small (2009) in the following form:

$$\Gamma\left(\sum_{ij} t^{ij}M^{ij}, \sum_{ij\neq iH} w^{ij}M^{ij}, \sum_{ij\neq iH} a^{ij}M^{ij}, \sum_{ij\neq iH} c^{ij}M^{ij}\right) \quad (6)$$

However, after substituting the term $c^{ij} = t^{ij}m(l^{ij})$, only the parameters in time dimension will remain:

$$\Gamma\left(\sum_{ij} t^{ij}\left(1 + m(l^{ij})\right), \sum_{ij\neq iH} w^{ij}M^{ij}, \sum_{ij\neq iH} a^{ij}M^{ij}\right) \quad (7)$$

In Equation (7) the measure of congestion is already time-related, hence the marginal monetary cost of travel time equals crowding's ($\varrho^C = \varrho^T$). If we derive the utility function supplemented with the crowding expression in a similar way to the term covering all travel costs expressed in Equation (5), and we derive the optimal vehicle size from this transformed utility function, then the maximum of the indirect utility function is as follows:

$$0 = \frac{\partial\tilde{U}}{\partial n} = -\lambda\varrho^T tM\frac{\partial m}{\partial l}\frac{\partial l}{\partial n} - \lambda Vt\frac{\mathrm{d}K}{\mathrm{d}n}. \quad (8)$$

By substituting $\partial l/\partial n = -l/n$ and $\mathrm{d}K/\mathrm{d}n = k_2$ in Equation (8), we can get a simplified form:

$$0 = -\lambda\varrho^T tM\frac{\partial m}{\partial l}\frac{l}{n} - \lambda Vtk_2. \quad (9)$$

According to the formula (A7a) given in the appendix of Parry and Small (2009), the elasticity of crowding cost is:

$$\eta_c^{ij} = \frac{\partial c^{ij}}{\partial l^{ij}}\frac{l}{c} = \frac{t\,\partial m(l)}{\partial l}\frac{l}{t\,m(l)} = \frac{\partial m(l)}{\partial l}\frac{l}{m(l)} = \frac{\partial m}{\partial l}\frac{l}{m}. \quad (10)$$

If we use Equation (8) and (10) together, we get:

$$0 = -\lambda\varrho^T tM\eta_c^{ij}\frac{1}{n}m - \lambda Vtk_2. \quad (11)$$

Rearranging Equation (11), dividing both sides by λM, then rearranging it again to get $\varrho^T tm$, which is practically speaking equals with $\varrho^C c$, one can get the formula that can be substituted in Equation (5) of the generalized costs.

$$\varrho^C c = \varrho^T tm = ntk_2/\eta_c o. \quad (12)$$

Actually, we can express the result by using other terms, if we use the multiplier function $m(l) = \alpha_c \cdot l$ directly, since:

$$\lambda\varrho^T tM\frac{\partial m}{\partial l}\frac{l}{n} = \lambda Vtk_2$$

$$\lambda\varrho^T tM\alpha_c l/n = \lambda Vtk_2$$

$$\varrho^T t\alpha_c lo/n = tk_2$$

$$\varrho^T t\alpha_c ml = tk_2$$

$$\varrho^C c = \varrho^T tm = tk_2/l. \quad (13)$$

Naturally the formula of generalized cost can be used with the result relationship from both types of derivation (only if a linear multiplier function is used):

$$q^{ij} = p^{ij} + \varrho^T t^{ij} + \varrho^W w^{ij} + \varrho^A a^{ij} + \varrho^C c^{ij} = p^{ij} + \varrho^T t^{ij}(1+m) + \varrho^W w^{ij}(1 + \eta_w^{ij}/\eta_a^{ij}) = p^{ij} + \varrho^T t^{ij} + \varrho^W w^{ij}(1 + \eta_w^{ij}/\eta_a^{ij}) + tk_2/l. \quad (14)$$

It must be mentioned that a tractable formula can be derived for the costs of crowding, which directly gives the cost of crowding using only the travel time and the parameters m or l (latter is the widely used load factor).

4.2. Determining the optimal transit fares

The other formula where crowding plays a significant role is the one to determine optimal fares and subsidy rates. In this case it is necessary to examine where the extreme value of the indirect utility function is, where the equation $\partial\tilde{U}/\partial p^{PR} = 0$ is met.

In addition to vehicle size optimization, crowding parameters also play important role in the marginal welfare formula's (A6) CROWD + VEHSIZE component. This sum contains the formula of the marginal welfare's crowding-dependent part, and the equation derived from the tax-relationship ($\mathrm{d}TAX/\mathrm{d}p^{PR}$), which is an indicator of the transport agency's operating cost, and therefore also depends on the vehicle size:

$$\sum_{ij\neq iH}\left[\varrho^C\frac{dc^{ij}}{dp^{PR}}M^{ij} + t^{ij}V^{ij}k_2\frac{\mathrm{d}n^{ij}}{\mathrm{d}p^{PR}}\right]. \quad (15)$$

The further transformations of the equation require the relationships for elasticity that can be found in the appendix of Parry and Small (2009). Using these formulae, Equation (15) can be transformed:

$$= \sum_{ij\neq iH}\frac{ntk_2}{o}(1-\varepsilon_V)\frac{\mathrm{d}M}{\mathrm{d}p^{PR}}. \quad (16)$$

From this equation, the form of the crowding factor (marginal cost of increased vehicle occupancy) is already apparent:

$$MC_{occ}^{ij} = \frac{ntk_2}{o}(1-\varepsilon_V). \quad (17)$$

If we approach the question from the viewpoint of crowding multiplier, we get a similar result, but the parameters are divided into several other factors. At first, with only the CROWD component counted in, we will get the

$$\varrho^C\frac{\mathrm{d}c}{\mathrm{d}p^{PR}}M = \varrho^T m\frac{\partial t}{\partial p^{PR}}M + \varrho^T t\frac{\partial m}{\partial p^{PR}}M \quad (18)$$

equation, where the first term represents the crowding costs due to variable travel time, which, due to the similarity can be linked to the USERTIM designation in the marginal utility equation of the model. Thus, the crowding multiplier also appears in the relationship of marginal congestion costs

$$MC_{cong}^{iH} = \sum_{k=H,B} t_H^{ik} \varrho^T (1+m) M^{ik} + t_H^{iB} K^{iB} V^{iB}, \tag{19}$$

that can later be used to calculate the marginal congestion costs of bus transport: $MC_{cong}^{iB} = \alpha_B MC_{cong}^{iH}$.

The second component of the sum expresses that the change of fares also has an impact on the crowding multiplier; the equation can be further adjusted as follows:

$$\varrho^T t \frac{\partial m}{\partial p^{PR}} M = \varrho^T t \alpha_c \frac{\partial l}{\partial o} \frac{\partial o}{\partial M} \frac{dM}{dp^{PR}} M. \tag{20}$$

By substituting the modified Equation (20) in the CROWD + VEHSIZE component used in Equation (15), the derivation can be continued:

$$\sum_{ij \neq iH} \left[\varrho^T t \frac{\partial m}{\partial p^{PR}} M + tVk_2 \frac{\partial n}{\partial p^{PR}} \right] = \sum_{ij \neq iH} \left[\varrho^T t \alpha_c \frac{\partial l}{\partial o} \frac{\partial o}{\partial M} \frac{dM}{dp^{PR}} M + tVk_2 \frac{\partial n}{\partial o} \frac{\partial o}{\partial M} \frac{\partial M}{\partial p^{PR}} \right]. \tag{21}$$

All the partial derivatives, $\partial l/\partial o = (1-\varepsilon_n) \cdot l/o$, $\partial n/\partial o = \varepsilon_n \cdot n/o$ and $\partial o/\partial M = (1-\varepsilon_V) \cdot o/M$ can be substituted in Equation (21), we get a quite compact relationship for the marginal welfare effects:

$$= \sum_{ij \neq iH} \left[\left(\varrho^T \alpha_c \frac{\partial l}{\partial o} M + Vk_2 \frac{\partial n}{\partial o} \right) t \frac{\partial o}{\partial M} \frac{dM}{dp^{PR}} \right] = \sum_{ij \neq iH} \left[(\varrho^T \alpha_c (1-\varepsilon_n) l + k_2 \varepsilon_n n/o) t (1-\varepsilon_V) \cdot \frac{dM}{dp^{PR}} \right]. \tag{22}$$

It is important to emphasize that in order to carry out the optimization, it is necessary to determine the occupancy of the vehicles, i.e. the load factor (l). Since the crowding level is often expressed by fraction of the number of passengers and the vehicle's (floor) size, it is necessary to determine this quotient in order to the further use of these parameters. Accordingly, the quotient of the average number of passengers and the average surface area of the vehicles (calculated for the vehicles of each city), or even the formula ($ml = k_2/\varrho^T = \alpha_c l^2$) retrieved from Equation (13) can give us the followings:

$$l = \sqrt{k_2/(\varrho^T \alpha_c)}. \tag{23}$$

Applying one of the two ways of approach (calculation of vehicle capacity for the whole vehicle fleet, or using Equation (23) with the parameter values defined in the model) can determine the required values of vehicle occupancy and the load factor, so the value of the crowding factor can also be calculated.

5. APPLICATION OF THE MODEL IN A HUNGARIAN INTERURBAN ENVIRONMENT

The uniqueness of the Hungarian settlement system is its strong capitalization, the effect of which can be obserbed both in the structure of the transport network and also in the main connections of transport services. Due to the country's radially structured transport network, the lack of transverse connections, and the poor permeability of the Danube River—which divides the country in halves—the vast majority of road and rail traffic flows directly through, or in the immediate vicinity of the capital. This effect puts a huge weight on the road and rail network of Central Hungary. One quarter of the country's population lives in Budapest and its metropolitan area (Hungary had a population of 9.77 million in 2019). Beside the complex and overcongested transportation system of Budapest and its agglomeration, the biggest traffic appears on the roads and rails between the central region of Hungary and the bigger cities of the country. Although of lesser importance in terms of traffic volume, railway lines and bus routes serving rural settlements also play an important role in the country's transport system.

An important feature of public transport in Hungary is that state-owned service providers still dominate the market. The Hungarian state-owned company group, the MÁV-Volán Group's subsidiaries are responsible for the passenger transport on rail and for the interurban bus transit—and in some cities for the local transport service also, on the basis of a contract between the municipality and the company. The MÁV-Volán Group has an important role to play in providing public transport in cities and metropolitan areas, between the regions of the country, and as well as in connecting villages, towns and cities to the transportation system.

The examination of the subsidization issue is complicated by the fact that in the European Union, hence in Hungary also, transport service providers often operate on a regulated market within the framework of public service. In a regulated market like this, there is competition when one comes to entering the market, but companies already in the market are providing the service on a yet exclusive basis. In addition, the state, the procurer of the transport services may impose a public service obligation (according to Regulation (EC) No 1370/2007), but in this case the transport company's burdens resulting from this service has to be balanced. (Jászberényi and Pálfalvi 2009) In other words, it is a practical solution for a state to sustain public transport services through subsidization, but this method of operation also fixes the dependence of the service providers on state resources.

Table 1: Parameter values used by the calculations for the Hungarian interurban transport system

Parameters	Hungary		Unit	Source of data
	Rail	Bus		
Median wage rate	7340	7340	HUF/hour	KSH, 2017.
Number of unlinked trips	146.9	490.6	millions/year	KSH, 2017.
Annual passenger miles	7366	7397	millions	ITM, 2017.
Annual rail car / bus miles	82	367	millions	ITM, 2017.
Fleet size	1677	5239	–	ITM, 2017.
Transit speed	51.9	35.5	km/h	MÁV-Start, 2016., KTI
Purchase cost of rail car or bus	350	62	million HUF	webpages of vehicle manufacturers
Total operating cost	243	161	billion HUF	ITM, 2017.
Total fare revenues	039	57	billion HUF	ITM, 2017.
Parameters	**Road transport**		**Unit**	**Source of data**
	Peak	Off-peak		
Annual vehicle miles	82.2		million/year	(Magyar Közút 2018)
Average trip length	41.1		km	KTI, 2016.
Fuel tax	228.0		HUF/liter	estimation
Occupancy	1.32	1.32	pass/vehicle	previous results of Hungarian measurements (from 2016)
Auto average speed	50	60	km/h	estimation
Fuel efficiency	6		l/100 km	estimation

Among the European Union member states Hungary has one of the highest rates of public bus transport (Statistical Office of the European Communities 2019), a kind of service that is available at all settlement. This level of coverage is probably caused by the fact, that the motorization level is well below the European average, and the fares are cheaper compared to other countries'. Moreover, the Hungarian State provides travel discounts to several social groups. For example, citizens under the age of 6 and over 65 can use almost all kind of public transport for free, and people with a student card also receive a 50% discount on their travel.

Although this paper primarily deals with the development of methodology of calculating optimal public transport subsidies, the Hungarian adaptation of the modified calculation method was also an important element of the research. The application of the methodology for Hungary has been carried out on a national level, primarily in order to ensure that the local characteristics of the large-scale transport system do not significantly influence the results of the calculations.

By the time of our research, the operating and budget data of the national service providers were available for the years 2016/17, so we were able to perform domestic calculations with quite recent data. To describe the operating conditions of the Hungarian transport system we used the parameters and the values shown in Table 1, the other parameter values were the same as in the original model. We used primarily the data provided by the Ministry of Innovation and Technology, transport service providers, the Central Statistical Office (KSH) and the Institute of Transport Science (KTI).

It is important to emphasize that the Hungarian adaptation mainly remains on theoretical level, and does not include a detailed analysis of the domestic circumstances and characteristics of the local transport system. For the purpose of calculation and the correct interpretation of results, it should be also noted that the data used for the calculations mainly describe the characteristics of the interurban transport system. These data also describe all settlements where the service was ran by state-owned bus and train operators.

6. RESULTS OF THE CALCULATION

The results of the calculations are discussed in two sections, differentiating the effects of modifying the crowding formulas and the results of the model's application for the transport system of Hungary.

6.1. Adaptation of crowding relationships

Reconsidering the relationships and formulas used in the model was a practical decision. The original study stated that the costs of congestion were "relatively small", and since then a couple of measurements found the exact opposite, i.e. that the economic losses caused by crowding cannot be neglected, so a revision of the crowding relationships and the model was highly appropriate. After the modifications, the costs of crowding can be included as a stand-alone element of travel decisions, rather than a proportion of other travel-time components. As a result, the previously presented crowding parameters also appears in the MW formula (see Eq. (2)), creating a direct relationship between user costs and objective congestion metrics.

The results obtained by the model using the modified cost and crowding formulas are summarized in Table 2. Compared to the results of Parry and Small (2009), it can be seen that the discomfort caused by crowding already accounts for a significant proportion of travel

Table 2: Results of modelling with modified cost and congestion relationships

	Washington, DC				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak
Current subsidy, percent of operating costs	47	55	80	76	83	82	79	69	67	72	59	40
Marginal welfare effects												
MW/W at current subsidy	–0,21	–0.24	–0.51	–0.50	0.34	–0.10	–0.18	4.21	0.51	0.01	0.09	1.38
Marginal cost/price gap	–0.03	–0.16	–0.94	–1.42	–0.57	–1.22	–0.87	–2.36	–0.20	–0.55	–0.25	–0.09
Net scale economy	0.09	0.41	0.51	1.99	0.18	0.86	0.45	5.90	0.04	0.28	0.30	1.74
Crowding costs	–0.08	–0.06	–0.20	–0.14	–0.17	–0.14	–0.22	–0.55	–0.07	–0.06	–0.21	–0.17
Externality	0.20	0.07	0.13	–0.05	0.79	0.32	0.46	0.44	0.57	0.35	0.14	–0.52
Other transit	0.04	–0.02	–0.02	0.13	0.10	0.07	–0.01	0.77	0.18	–0.01	0.12	0.41
MW/W at 50% subsidy	–0.21	–0.26	–0.05	–0.56	0.37	0.18	0.09	3.25	0.49	0.17	0.14	1.45
Optimum subsidy, percent of operating costs	>90	80	42	>90	>90	78	64	>90	>90	73	71	>90
Proportion of subsidy due to												
Marginal cost/price gap	0.45	0.57	0.40	0.40	0.38	0.48	0.42	0.49	0.31	0.55	0.50	0.39
Net scale economy	0.18	0.44	0.77	0.61	0.12	0.41	0.40	0.45	0.04	0.22	0.42	0.74
Crowding costs	–0.16	–0.06	–0.29	–0.04	–0.11	–0.07	–0.19	–0.04	–0.07	–0.04	–0.29	–0.07
Externality	0.45	0.08	0.15	–0.02	0.54	0.15	0.38	0.04	0.55	0.28	0.20	–0.30
Other transit	0.08	–0.03	–0.02	0.04	0.07	0.03	0.00	0.07	0.17	–0.01	0.17	0.24
Percent change is passenger miles	23.9	34.1	–48.3	28.6	11.6	–12.8	–23.6	51.0	20.3	1.7	12.6	142.5

costs (+20–60% increase), even at moderate congestion levels. It can also be observed that a significant increase in crowding-related (marginal) costs is to be expected, regardless of city, transport mode and travel period. The growing number of travels in the off-peak period may be more likely to be caused by the “replenishment” of crowding costs that were completely neglected before. The results show that marginal welfare gains, optimal subsidies and the expected passenger number are reduced in most of the studied cases. The direction (and magnitude) of the subsidy change is in line with previous expectations, as the transport system is able to optimally operate with fewer passengers due to the increase of travel costs, where the losses are mostly caused by crowding disutility.

The exception is the off-peak bus service of Los Angeles—we can also observe a minimal increase in London’s peak rail marginal welfare gains—the model predicts significant subsidy and passenger growth compared to previous results. This may be caused by the fact that decreasing peak-time subsidies and significantly increasing off-peak period subsidies may lead to a higher proportion of trips taking place in the latter time period.

Comparing the results with the unmodified model’s, it can be clearly seen that even in cases of moderate congestion and crowding level, the passengers’ travel costs significantly increased. As a result, in most cases a substantial part of the benefits arising from scale economies are neutralized by crowding losses. In all the modelled cities crowding losses approach the level of peak rail transport’s benefits coming from scale economies, and in the case of London it also outweighs it. Therefore, as a results of the model modification, we conclude that at all (public) transport systems with heavy traffic the negative effects of crowding, and all the related losses should always be taken into account.

6.2. Results of the model’s application in Hungary

With the use of the presented parameter values (Table 1), the results of the optimization show that a subsidy rate of around 90% should be applied at almost all of the cases. There is an exception of the peak period bus transit, where a subsidy rate of 74% is recommended. This is 2 percentage points lower than the applied value of subsidy.) The forecast for change of passenger miles are in line with the change of subsidy rate. It predicts lower growth in the rail sector, while the rise by off-peak period bus travel, with the highest increase of subsidy that exceeds 90%. (A smaller decline in passenger miles is expected at the peak period bus transport, because of the optimal subsidy rate is smaller than the current one.)

The disaggregated results of Table 3 also show that similarly to the studied cities abroad, the benefits of the reduction of externalities also play a decisive role in the Hungarian transport system. Compared to the other transport systems however, in Hungary relatively high subsidy rates are applied, so the situation does not provide the opportunity for a significant increase of subsidy—with the exception of off-peak bus transport. Therefore the congestion losses can virtually neutralize the benefits of scale economies. It can be undoubtedly concluded, that similarly to other cities, a high subsidy rate is a necessity of the Hungarian interurban public transport system.

The results of the modelling with Hungarian data show many similarities with the characteristics of foreign cities, although the model basically aims to examine the transport system of the urban and suburban environment. The similarity between the results of the model variants (the original and the modified version) shows that the model can be

Table 3: Results of the application of the model in Hungary

	Hungary			
	Rail		Bus	
	Peak	Off-peak	Peak	Off-peak
Current subsidy, percent of operating costs	87	87	76	64
Marginal welfare effects				
MW/W at current subsidy	0.27	0.01	–0.02	0.38
Marginal cost/price gap	–0.85	–1.74	–0.71	–0.69
Net scale economy	0.02	0.13	0.12	0.69
Crowding costs	–0.11	–0.09	–0.15	–0.12
Externality	1.01	0.64	0.58	0.05
Other transit	0.20	1.07	0.15	0.46
Optimum subsidy, percent of operating costs	**>90**	**88**	**74**	**>90**
Percent change is passenger miles	5.7	3.8	–3.2	92.4

adapted and used in a geographical, social, economic, technical environment that is significantly different from the original cities. Naturally, the accuracy of the results could be improved by specifying the data, or by a detailed review of the factors adjusted to the features of other (Hungarian) cities. Nonetheless, the calculation method and results are perfectly consistent with the aims of the study.

7. SUMMARY, CONCLUSIONS, OPPORTUNITIES OF DEVELOPMENT

The results of the paper has been presented in two clusters. Firstly, we described the results of the transformations of the crowding formulae and all the modifications carried out in the model. Second, we presented the conclusions of the Hungarian adaptation. Finally, the suggestions and possible directions of further development of the model and the research area are presented.

7.1. Effects of crowding

The revision of the original study have shown that by taking the costs of crowding into account, it would be advisable to apply a smaller subsidy at peak periods. This would obviously lead to an increase in fares, resulting in fewer passengers and thus less crowding on public transport vehicles. At the same time, higher off-peak subsidy levels and cheaper fares would allow some travels to be shifted to this period, which would provide better overall transport conditions—naturally these provisions would require time-differentiated pricing which is currently unavailable in Hungary. It is also important to note that in Hungary this theory may be hampered by the fact that those social groups, whose travels could be relatively easily redistributed, such as people over the age of 65, currently do not have to pay fares, and hence the changes of pricing would have no effect on their travel habits.

From the results of the model modification and supplementation, we concluded that in these cases of heavy traffic and crowded transport systems the negative effects of the crowding phenomenon and the resulting losses should always be taken into account.

7.2. Interpretation of results in Hungary

The results of adapting the model for Hungary showed that although the model is primarily designed for urban and suburban traffic, it can be applied to national level service providers as well. As the collected technical and economic data primarily described the operation of the national railway and bus transport system, we could analyze the data of interurban traffic at a national level. In our opinion, the most important requirement for the model is that data shold come from cities, counties, regions etc. with similar economic, social, and transport conditions—this limitation prevents important features from being “averaged”, otherwise the data would already bring distortions into the calculation.

The results for the Hungarian transport system are clustered around very high subsidy rates. For example, providing a larger subsidy rate in the off-peak period could cause the higher occupancy of vehicles outside peak hours—for example, by shifting a part of peak period bus travels to the off-peak period. According to the model this transmission could utilize idle capacities, and thus it would make a substantial improvement of the transport system and would also achieve the reduction of social losses.

We suggest that the state should also promote the more even use of capacities in other ways to prevent harmful levels of crowding. Gradual shift start supported by the state, postponing the beginning of school, or by streamlining the flow of information by organizing and standardizing a real-time information management systems—e.g. presenting the usual occupancy of vehicles, or the actual values using ticket purchase or data of vehicle sensors. Additionally, transformations caused by the coronavirus epidemic coluld be also good examples. Changing working, commuting and travel habits, accelerating digitalization and the possibility od home office provide an opportunity to a major review of transport strategies.

7.3. Directions for further improvement

By the further development of the model, it could more effectively predict the impact of strategic pricing decisions at a given city or at a regional (national) level. Currently the model cannot really account for the capacity constraints of the transport system. Thus, for example with increasing peak frequencies the user can't be certain that the system can manage the increasing demand without costly infrastructure (or traffic organizational and controlling) investments. Take the rail network of Budapest for example, which is currently suffering from the lack of capacity, so the model would also require a parallel examination of road and rail capacities. Despite the model can control the occupancy of a particular traffic mode and in a given period by optimizing crowding, but there may be situations where this method is no longer sufficient and enough for an optimal solution.

The main question of the modelling carried out during the research was to examine the kind of pricing that could be used to achieve the (second best) social optimum. Though, changing prices necessarily imply some kind of traffic reorganization, e.g. the increasing occupancy of public transport due to the reduction of cars in areas prone to congestion, but there are indirect effects that do not appear in the model. The calculation does not include the gains from detached and unused parking and transport spaces in downtown areas due to reduced road traffic, or the potential benefits of using these yet free spaces for other purposes—the benefits of increased local trade due to bigger pedestrian and bicycle traffic, or the rise in property prices.

Further questions are raised, since due to the use of aggregated data, it is not possible to select or further divide users, i.e. the impact of pricing cannot be examined separately on several social groups. It would be important to analyse in detail how social groups respond to a transport policy pricing decision, since in most cases there is a correlation between the quality of transport system and the settlement environment, thus housing prices, and so indirectly with the financial situation of residents. For these reasons, it is not certain that groups which are still distinguishable in reality will behave according to the model as a result of pricing. Therefore, it is questionable whether the social and the political-economic optimum do not slip, or intentionally can be slipped away. For example, the price-sensitive, but typically more mobile people can be "supported" by the overcharging of transport services for people with less income and reduced mobility. But this method can be used for good reasons as well, if we influence the travel patterns of the more price-sensitive social groups by pricing to disencumber the overcongested and crowded peak periods.

Acknowledgement

I am grateful to my supervisors, András Munkácsy and Dániel Hörcher for their help and support.

REFERENCES

[1] Björklund, Gunilla, and Jan-Erik Swärdh. 2015. 'Valuing In-Vehicle Comfort and Crowding Reduction in Public Transport'. *Working Papers in Transport Economics* 2012(12).

[2] Boss, Alfred, and Astrid Rosenschon. 2008. *Der Kieler Subventionsbericht: eine Aktualisierung*. Kiel: Inst. für Weltwirtschaft.

[3] Doll, Claus, and Huib van Essen. 2008. *Road Infrastructure Cost and Revenue in Europe*. Delft, The Netherlands: CE Delft.

[4] Hörcher, Daniel, Daniel J. Graham, and Richard J. Anderson. 2017. 'Crowding Cost Estimation with Large Scale Smart Card and Vehicle Location Data'. *Transportation Research Part B: Methodological* 95:105–25. doi: 10.1016/j.trb.2016.10.015.

[5] International Transport Forum. 2018. 'Zero Value of Time Roundtable'. *ITF*. Retrieved 6 November 2019 (https://www.itf-oecd.org/zero-value-time-roundtable).

[6] Jansson, Jan Owen. 1979. 'Marginal Cost Pricing of Scheduled Transport Services'. *Journal of Transport Economics and Policy* 13(3):268–94.

[7] Jászberényi, Melinda, and József Pálfalvi, eds. 2009. *Nemzetközi közlekedés és turizmus*. Budapest: Aula Kiadó.

[8] Keserű, Imre, Jeroen Bulckaen, Cathy Macharis, Joeri Minnen, Ignace Glorieux, and Pieter Theun van Tienoven. 2015. 'Is Travel Time Wasted? Evidence from a Time Use Survey in Flanders, Belgium'. Windsor, United Kingdom.

[9] Keserű, Imre, and Cathy Macharis. 2018. 'Travel-Based Multitasking: Review of the Empirical Evidence'. *Transport Reviews* 38(2):162–83. doi: 10.1080/01441647.2017.1317048.

[10] Kroes, Eric, Marco Kouwenhoven, Laurence Debrincat, and Nicolas Pauget. 2014. 'Value of Crowding on Public Transport in Île-de-France, France'. *Transportation Research Record: Journal of the Transportation Research Board* 2417(1):37–45. doi: 10.3141/2417-05.

[11] Mohring, Herbert. 1972. 'Optimization and Scale Economies in Urban Bus Transportation'. *The American Economic Review* 62(4):591–604.

[12] Mohring, Herbert. 1976. *Transportation Economics*. Ballinger Publishing Company.

[13] Munkácsy, András, Imre Keserű, and Miklós Siska. n.d. 'Travel-Based Multitasking on Public Transport: An Empirical Research in Hungary'. *Periodica Polytechnica Transportation Engineering* (megjelenés alatt).

[14] Parry, Ian W. H., and Kenneth A. Small. 2009. 'Should Urban Transit Subsidies Be Reduced?' *The American Economic Review* 99(3):700–724. doi: https://doi.org/10.1257/aer.99.3.70.

[15] Perone, Jennifer S. 2002. *Advantages and Disadvantages of Fare-Free Transit Policy*. Tampa, FL: University of South Florida.

[16] Statistical Office of the European Communities. 2019. *Energy, Transport and Environment Statistics*.

[17] Tirachini, Alejandro, Lijun Sun, Alexander Erath, and Artem Chakirov. 2016. 'Valuation of Sitting and Standing in Metro Trains Using Revealed Preferences'. *Transport Policy* 47:94–104. doi: 10.1016/j.tranpol.2015.12.004.

[18] Tscharaktschiew, Stefan, and Georg Hirte. 2011. 'Should Subsidies to Urban Passenger Transport Be Increased? A Spatial CGE Analysis for a German Metropolitan Area'. *Technische Universität Dresden, Faculty of Business and Economics*.

[19] Turvey, Ralph, and Herbert Mohring. 1975. 'Optimal Bus Fares'. *Journal of Transport Economics and Policy* 9(3):280–86.

[20] Vickrey, William. 1980. 'Optimal Transit Subsidy Policy'. *Transportation* 9(4):389–409. doi: 10.1007/BF00177700.

[21] Whelan, Gerard, and Jon Crockett. 2009. 'An Investigation of the Willingness to Pay to Reduce Rail Overcrowding'. Harrogate, England.

A MULTI-LEVEL PERSPECTIVE ON VELOMOBILITY IN CZECH REPUBLIC

Daniel Szabo, Mgr.[1]
Transport Research Centre
Lisenska 33a, 636 00 Brno, Czech Republic
daniel.szabo@cdv.cz

ABSTRACT

Technological and social progress in bicycle use and vehicle versatility (smart and social innovations, namely electrification, smart infrastructure, bicycle-based logistics systems or crowd-shipping) and landscape pressures, connected to GHG emissions and changing patterns of virtual and physical mobility, have the potential to radically tip the scales of the dominant socio-technical transport regimes in personal mobility. The article utilizes a mixed-theory approach, based on analyzing the data, collected through the Czech National Travel Survey (Česko v pohybu), conducted in 2019, that feed into a multi-level perspective (MLP) model of sustainability transition pathways.

The article aims to identify core regime factors of bicycle mode choice in the Czech Republic through controlling for variables such as length and time of the journeys, daily and weekly variations, activities, connected to bicycle use, trip chain structure, and socio-geographical variables, such as urban structure and density, amenity density, socio-demographic variables and others.

The results of the statistical analysis form a basis for an MLP model, that analyses current and future niche, regime and landscape factors and potential pathways towards increasing bicycle mobility, motility, universality and users' diversity, concerning systemic interplays between infrastructure, users, governments and other human and non-human actors.

Through identifying the possible development scenarios and viability of achieving the goals, defined in national strategies and urban policies and exploring the potential for hybridization of the components of the regime, that are the most sensitive to change, the article further proposes the space for new potential policies to support broader viability of velomobility for the general population.

Keywords:
Sustainable urban mobility, Velomobility, Mode choice, Multi-level perspective, Socio-technical innovations

1. INTRODUCTION

There are few singularly more important urban mobility goals than promoting cycling as a universally viable choice mode. The evidence is compelling and mounting, that the current (and accelerating) automobile-dominated mobility system is unsustainable from a wide range of criteria – environmental, social (equity) or public health – and a change in course is direly needed. In the presence of decades of strategies aiming to attain just that[8], the indicators, connected to macroscopical cycling mobility goals (accident incidence rates, modal split) in the Czech Republic are slow to progress, even receding at times – we might speak of a *velomobility inertia.* On the other hand, another significant trend has been looming on the horizon for at least a decade – so-called *peak travel, or peak car* (Millard-Ball, 2011) – even Czech motorization rate is climbing still, and is projected to do so.

Cycling mobility consists of a wide spectrum of activities and experience, where the dominant narratives – usually of *highly skilled, but unobtrusive* cyclists, and cycling as an *activity as a goal* – might serve as a gatekeepers both for cyclists and infrastructure design. Focusing on too a narrow range of cycling patterns (represented e.g. by data, collected by fitness applications, such as Strava Metro) or indicators (such as vehicle miles travelled) is doomed to be counterproductive, when assessing

[1]E.g. through the National strategy for development of cycling mobility 2013-2020.

for a softer, more equitable mobility – not a *performance* one.

The first Czech National Household Travel Survey (NHTS) (Centrum dopravního výzkumu, 2020) is a publicly available dataset on harmonized personal mobility data, that hasn't been used for a research into mobility patterns yet. The aim of the article is thus to utilize the available data to analyze the current factors, influencing modal choices, and the stabilizing and destabilizing factors of the state in order to identify possible transition pathways towards cycling culture, open and accessible for all people and all purposes.

The paper is structured as follows. **Chapter 1** (Introduction) presents the general overview of the current state of the velomobility and its confrontation with the incumbent mobility regimes in order to better outline the landscape in which the mobility choices are being made. **Chapter 2** (Methodology) concerns the methodological background on external environmental factors on the bicycle mode choice and presents the data, used for the analyses. **Chapter 3** (The binomial regression model) presents the analyses and results that form a base for the multi-level perspective model. **Chapter 4** (The multi-level perspective model) outlines the potential for urban and national policies and policymakers to adopt support for the regime change. **Chapter 5** discusses the results and conclusions in the light of the potential for a politically-driven change.

1.1. The context of cyclist infrastructure planning in Czech Republic

1.1.1. The overview

Nello-Deakin (Nello-Deakin, 2020) has recently proposed not only that practical (ie. technical or economical) aspects are not the main barrier for cycling infrastructure provision, but that the theoretical knowledge has expanded enough to a point, where there is little else to explore to further illuminate the forming policies – and multiplying research in underlying factors of cycling might be in fact counterproductive.

Indeed, the aim of this article is not primarily to uncover an previously unknown factor, that would once again "solve" the problems of geographically and societally uneven propensity to cycle. On the other hand, where the general theory of largely universally applicable cycling factors is in practice frowned upon i.a. due to the lack of data, an analysis of both the mobility patterns and the specific and varying problems, the Czech municipalities face, is not misplaced.

Moreover, regardless of the deemed lack of gaps in theoretical research, the *practical application* of proven factors of modal choice through transportation modelling – which is the core part of the process of deciding on the location and design of road network – to a large extent lacks detail and regard for soft modes.

One of the core aspects of a viable cycling system is a relatively dense, connected and safe linear infrastructure – i.e. segregated cycleways. From this standpoint, it is safe to say that segregated cycling infrastructure throughout Czech Republic is border to non-existent. There are less segregated cycleways (lengthwise) in Czech Republic, than there are highways (1.94 % of length compared as of 2020). There are still less segregated cycleways and mixed cycle-pedestrian ways *combined*, than highways (92.5 % of length compared as of 2020 (CEDA, 2020)). Given the low share of highways on the road network and magnitude of difference in construction, operational and externalities costs, the magnitude of difference is incomparable – and still growing. The rate of growth of the network is further geographically uneven, and still mostly concentrated along a few long-distance routes (such as Eurovelo), even negative at times (e.g. via recategorization). The difference is the highest in the urban areas, where protected measures are scarce, due to spatial conflicts, or outright dangerous due to a common lack of skilled cycling infrastructure planners.

The mean increase of segregated cycling and mixed cycling-pedestrian paths has been 28 meters per municipality per year for the last three years – although this is a substantially higher growth rate than for the whole road network (8.7 % per year, compared to 0.08 % per year). New cycle-only roads represent a staggering 26.75 % relative annual growth rate, which only illuminates the paradox, where even small projects represent a tremendous relative leap in the overall network capacity. The vast majority (84 %) of the municipalities haven't recorded any change in their segregated cycling infrastructure.

Reliable data on softer (non-segregated) cycling infrastructure are not yet being collected – while the OpenStreetMap data structure is exhausting, the completeness of mapping cannot be estimated (albeit it is probably fairly high, and shows a similar pattern: a dozen of cases of active municipalities with relatively higher share of cycle lanes). The support of cycling measures is highly geographically uneven and dependent on the proactive approach of the respective municipalities, public pressure and cooperation of several administrative bodies to push for individual projects' approval.

1.1.2. The municipal planning

The cycling infrastructure is planned mainly on local (municipal), regional/long-distance (NUTS 3) level, or both in coordination.

- Technical standards for cycling infrastructure more often than not deter from pursuing the safer (or any) cycling solutions, by raising the bar of costs, land-use/spatial requirements, and technical and safety conditions, that can be objected to.
- The focus on isolated cycling infrastructure projects is favoured both by funding schemes both on national and European level (given the high threshold of project preparation costs). Needless to say, the projects, that favour recreational cycling, might produce an uneven and damaging effect on various non-recreational cycling/cyclists and their presence in the vast majority of non-cycling road space.
- The focus on large, linear infrastructure projects disregards the value and qualities of surrounding space, that are paramount to a human movement, and the vast majority of daily travel patterns: the street permeability, accessibility of destinations (including parking provision and safety), or presence of amenities.
- Both cycling and non-cycling projects consist mostly of new, added corridors, rather than transformation of the existing space. If unmitigated, this adds capacity to the road network, inducing new automobile demand (such as in the case of ring roads (Drabicki, Kucharski, & Szarata, 2020)), reduces land use density and increases environmental costs (e.g. habitat fragmentation, road runoff pollution).
- Whether the cycling infrastructure stimulates cycling demand or vice versa, there is a positive correlation – even if this is city-specific, and many municipalities sustain high levels of cycling without any segregated infrastructure; and, on the other hand, segregated infrastructure (especially recreational) does not necessarily translate into a well-serving infrastructure.

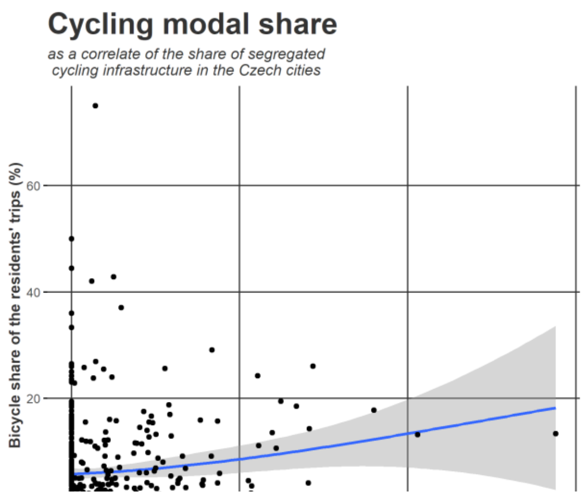


Figure 1: The correlation between modal share of cycling and the share of segregated cycling infrastructure in the Czech cities

1.1.3. The role of expertise in sustaining a locked in automobility regime

The current state constitutes a *locked-in regime of automobility* – that is a socio-technical regime (Geels F. W., 2007), sustained through shared expert practices (ie. solutions heuristics) but are also embedded in the societal and cultural practices and rules, or built infrastructure and its obduracy (Urry, 2004) (Hommels, 2005) (Hoffmann, 2017) . The dominance of the regime reverberates in different stages of transport policies creation or enactment, even if its effects are variegated as a result of groups and networks interactions. While it is outside of the scope of this article, it is worth noting that policies, that favour growth of automobility (or stymie alternatives), pervade through different sectors, mostly connected to transport and construction. The examples may draw from the legal basis for requirements for parking place provision (in different context, although to an extent applicable, famously analyzed in (Shoup, 2011)) – or in practice relaxed interpretation of technical norms, that are intended to serve active or sustainable modes of mobility.

The *automobile regime lock-in* has found ways to trickle down further to transport modelling or mode choice modelling:

1. for one, it appears as fixed automobility growth coefficients (even where the automobility growth directly contradicts the long-term transportation policies' goals) both in the rate of motorization and in the rate of traffic intensities and vehicle miles travelled. The macroscopic pressure translates into increasing the network capacity, where lack thereof is projected.
2. It is expressed in involuntarily skewed data from national travel surveys (travel diaries), where lower validity of active trips leads to their underrepresentation (e.g. there is 3.5 % higher representation and 0.6% lower representation of car/bicycle trips respectively in our valid dataset, compared to the raw dataset, see Table 1*).* Even further, the active trips are underrepresented by involuntarily lacklustre reporting of the travellers (as the less important, forgettable trips), or where they exhibit a higher degree of diffusion between the trips and the activities (resulting in lower trip rate for active modes, connected to several activities). Furthermore, in activity-based models, non-trip activities are not considered at all, inflating the significance

of the major trips – and reducing sensitivity to the shift in virtual mobility/presence.

3. The role of a technical expert obfuscates the leeway that exists when creating or interpreting the models and selecting the proposed measures. The scientific knowledge – be it expertise on human health, air pollution, transport modelling or econometrics – is shaped by the alternatives – and means for pressing towards the pre-selected alternatives are manifold – their (lack of) accuracy is therefore predictable (Flyvbjerg, Holm, & Buhl, 2007). On the other hand, where conflicting expertises accumulate and are politicized (usually in the cases of controversial variants, or exceeding uncertainty), the expertise may be *bypassed* by the decision-makers (Konopásek, Stöckelová, & Zamykalová, 2008).
4. At last (and this is the starting point of our article), the impact is no less present in the common functions for mode choice, that stem from the four step models' requirements for reducing the scope of modelled variables and spatial detail. The active modes of transport are still often omitted, or left behind, often due to lack of data collected (e.g. pedestrian counts observations), simplicity of the network (e.g. lacking paths) or reducing the complexity of the model. Not including factors, relevant for specific mode choice decisions then leads to underestimation or discarding of planning policies and strategies, that influence the cyclability (walkability) of the area. One of the seminal works in this area (Katz, 1995) highlights the practical aspects of modelling specifically for the bicycle (even if such case can be made for any mode of transportation) – quantifying mutual causal relationships between urban planning processes, individual choices (such as household-work location) and transportation mode choices raises the probability of bringing about the change, as it allows for their explicit inclusion. From the built form point of view, especially important are the microscale and urban form factors, that are not reducible to time costs, such as network design and suitability, density or directness (Rybarczyk & Changshan, 2014).

When dealing with transportation project decisions, public authorities will refer to transport modelling, typically based on multinomial logit models, derived from travel surveys data and data about the existing infrastructure, that, combined, form the landscape in which the modelled travellers' choices are being decided. Multinomial logistic models have been a preferred approach to targeting variables influencing transportation mode choice, mostly due to their explanatory power, that can be utilized for assessing future infrastructure or non-infrastructural measures, such as traffic calming.

The approach has several possible pitfalls that demonstrably influence decisions on transportation projects:

1. The transportation models are implicitly motivated to skew the results in order to favour new development. This is a function of the power imbalance between the large technical infrastructure construction sector and marginalized actors and interests (such as environmental protection), where transport models are often created as a means to substantiate new development, not as a tool for their evaluation. Road infrastructure plans in Czech republic are construed upon an idea of *perpetually delayed completeness* that pervades through different sectors, but is probably most tangibly exemplified through the core highway and motorway network, where pressures towards new development have petrified in different socio-technical conflicts (Konopásek, Stöckelová, & Zamykalová, 2008) and are present as a permanent struggle to unstring the existing legal protections and public participation representation, such as in the case of the latest ongoing Construction law reform.
2. As (Cervero, 2002) notes, the underlying mode choice models are in practice usually underfitted, lacking plenty of possible available factors, apart from generalized costs – such as built form, land use characteristics or network quality. Omitting significant factors from the models leads to a distortion of the effect of the other factors of mode choice.
3. The model's purpose ends at identifying the factors, related to discrete choices based on statistical probability. The material, or even political reality cannot be easily bent to attain specified goals or even reshaped to fit the models, as it also would have to disregard the trade-offs between the identified factors. Even on the most simple terms, the policy trade-offs are unavoidable: increasing average traffic speed to reduce travel time leads to higher accidents* rates and higher investments toward safety and the users side, which further deter new users (such is the case of mandatory helmets on bicycles or potentially mandatory, even if non-existent helmets in cars). Land use

density (or rather suburbanization processes) are a results of decades-long deregulation of land use regulation and pressures toward profit maximization (Maier, 2012). Land value and land use regulations usually decrease with increasing distance from the urban core, leaving more room for pedestrian or cycling infrastructure – in areas with low density and transit accessibility, increasing car dependency.

The shortcomings that (Cervero, 2002) has identified in mode choice modelling are very much present in the current state of the art in Czech Republic, despite the theoretical advances and even though the availability and completeness of the datasets on possible influencing factors, such as accidents rates, land use, density, bicycle facilities or urban design, has improved considerably.

But neither changes in the bicycle infrastructure provision, nor soft, persuasive policy changes, might be the leading factor, influencing the proclivity to cycle. The disruption of public transport services, drops in road usage and shifts in daily mobility patterns due to recent Covid-19 pandemics have in many metropolitan areas worldwide resulted in a sudden spur of new, often pop-up bicycle infrastructure or bicycle-aiding schemes, such as low-traffic neighbourhoods. Other major pressures – climate change etc. – may break alone, or via major actors (e.g. European union – see (Hoffmann, 2017)) into a new momentum and a more permanent regime change. Such "regime erosions" may stem from two intersecting paths: the *"1) bottom-up mobilization, which gains momentum through positive policy feedbacks, and 2) weakening policy regimes due to negative policy feedbacks. These conditions enable policymakers to defect from the eroded policy regime to the alternative, which is accompanied by major policy change."* (Roberts & Geels, F.W., 2019, p. 4)

Indeed, it is impossible to make reliable predictions about future human behaviour in tumultuous times of potentially *critical junctures* (ibid.). Yet, precisely the critical junctures in history present the highest potential for disruptive policies to gather support and create positive feedback loops – drawing in new users, providing fast, demand-driven improvements.

Yet – are there any potential paths – or breaking points forward for velomobility in Czech Republic precisely in the light of conflicting interests and potential regime instability? The aim of this article is to bring a *multi-level perspective model of socio-technical innovations* approach to bicycle infrastructure planning. The model is in the first phase informed by a regression analysis of possible macroscopical factors, involving urban environment and individual/household characteristics, influencing current individual mobility choices for the general population.

2. METHODOLOGY

2.1. Built form and Land-use factors influence on travel behaviour

The relationship between various exogenous factors – land use, urban morphology, or endogenous factors (such as risk perception or attitudes towards a transportation mode) have been studied extensively on various scales for several decades (Rybarczyk & Changshan, 2014).

There are additional studies rooted in Czech context (e.g. (Gabrhel, 2019)) that propose extending the travel surveys to include psychological or stated preference factors, based on the individual predictors for mode (bicycle) choice. This approach, while providing useful insights, relies on extending the scape of the national travel surveys' travel diaries, which already battle with low return/fill rates due to time required and lack of motivation.

On the other hand, if shown any significance (esp. on the national level), extending the existing models with additional available data could potentially aid to create more precise models without increasing the burden on either side.

The primary candidates for such (quantifiable) factors would be those, that can be aggregated and connected to the trips (typically routes or origin-destination zones), usually based on geographical proximity or zonal aggregation – factors, such as land use, built urban form, or accident rates. Identifying more detailed parameters of either the network or built environment requires increasingly unobtainable consistency and validity of the geo-spatial data.

The chosen approach allows to control the correlation of aggregated socio-demographic factors, such as automobility rates or income with urban-transport structures (e.g. car/transit accessibility), but also individual characteristics, that form the mobility decisions on the individual and trip level. The relationship between different aspects of built environment may not be linear, or transparent – due to the diversity and variance in land use and travel behaviour through a whole republic. Research in the influence of breadth of various factors onto mode choice (or vehicle kilometres travelled) usually focuses on a strictly geographically delimited area (or comparison of multiple such areas (Cervero, 2002)).

At the same time, some of the factors might not show hypothesized effect, because of their form or localization. The typical example for this effect would be localization of bicycle infrastructure, that tends to avoid busy, quoditian connections due to spatial conflicts with motorized road infrastructure, the effect being meagre, or even negative estimates of the presence of segregated infrastructure on the mode choice, as the misplaced infrastructure is not being actively used.

Controlling for other multiple variables, such as destination activity of the trip, land use and POI density along the route, may improve the predicting precision of the model. However, for the second step of the analysis, the *substantial* significance is more important, than atheoretical modelling precision.

Yet even if untangled, the possible relationship between land use characteristics and mode choice is still not easily translatable into land use or transportation plans. Even detailed data on various geographically aggregated indicators cannot easily translate into individual decision making, bar viable public policy.

As shown in the Introduction, it's not the land use factors, or land use policies, that form mode choices – the different factors, influencing mode choice are themselves a result of interplay of broader network of actors, influencing technology, infrastructure, politics or culture.

The identified predictors are henceforth not primarily used as factors, stimulating bicycle use, on the contrary: the general assumption being the bicycle is under favourable conditions usable universally, many of the factors are better in explaining *the missing gaps* in broader adoption of the bicycle.

2.2. Sampling and data

The data, used for the analyses are as follows:

1. Czech National Household Travel Survey (NHTS, so called *Czechia on the move*) (Centrum dopravního výzkumu, 2020). The data have been collected through random sampling through years 2018-2019, using standardized approach, based on European harmonization of travel surveys (Armoogum, 2014) and Austrian methodology KOMOD, that has been adapted for Czech context.
2. The Czech Global Network (GNDB) database, created by Road and Motorway Directorate of the Czech Republic.

The built environment and road network data have been aggregated on the municipal level with the following considerations:

- The uneven precision of geographical localization might misplace trips origin-destination TAZ, while this is far less probable on the level of a city.
- The urban mobility policies tend to be more or less coherent. On the other hand, there is high variability within the city both for the built environment and for the mobility patterns (e.g. urban core and periphery/residential zones), and this variability will increase with the size of the city. This may be a reasonable argument for aggregating the data on the level of urban districts, amending the analysis.

The NHTS has been conducted as a continuous survey from May 2018 to April 2019, except for the months of December and January via *probability-proportional-to-size sampling* with sample size of 40 000 households with assumed sample response rate of 25 %, resulting in the final sample of 9419 households, contacted in person, following a PAPI or CAPI survey for all the members and vehicles of the household.

Table 1: Individual (socio-demography) variables of the sample

Variable	Description	Sample (n=22,122)
Gender	0:Male 1:Female	10,647(48.13%) 11,472(51.87%)
Age	Median	45
Education	0: Unfinished primary 1: Primary 2: Vocational 3: High school 4: Vocational college 5: University	2,746(12.4%) 2,292(10.36%) 6,537 (29.55%) 7,157 (32.3%) 526 (2.38%) 2,859 (12.9%)
Work status	0: Not working 1: Working	10,602(47.9%) 11,514(52.1%)
Car availability	0: none 1: any	7,659 (34.6%) 14,463 (55.4%)
Bicycle availability	0: none 1: any	11,497 (51.9%) 10,625 (48.1%)
Driving license availability	0: none 1: any	8,628 (39%) 13,494 (61%)

The NHTS data structure is split into several dataframes, including:

- Households
- Persons
- Trips
- Cars (vehicles)

Each of the dataframes have been checked and validated for completeness, range validity, data integrity, data usability and logical consistency.

Missing values were checked and imputed or corrected, according to typical errors present.

The trips have been geolocated via identification on various levels (the most precise being house level) with the official registry of addresses, combined with API geolocation approach. The probable routing of the trip has been calculated, based on constructed network and a separate API for public transport queries. Validity of the geolocation of origin-destination points and the validity of computed/declared time and distance of the trip has been cross-checked, based on relative and absolute thresholds of difference between computed and declared time for each mode, which leads to relatively lower share of active modes.

Table 2: Proportion table for the raw/validated sample of the trips

Main mode	All trips [%]	Reliable trips [%]
Car driver	29.2	31.6
Car passenger	9.9	11
Bus	3.8	3.8
Bicycle	4.5	4.2
PT	14.6	16.1
Other	0.8	0.5
Walk	35.4	31.1
Train	1.8	1.6

Overall, the share of time, distance and location-wise validated trips is 69.86 % (15,498 invalid trips, with critically corrupted structure, out of 51,434). Since all of the variables are critical for our analyses and there is no assumption of non-standard distribution of the omitted data, we have decided further to work with the validated data only.

Main mode for the trip, where combination of modes has occurred, has been chosen, according to the general practice (train > coach > car > public transport > cycling > walking).

Table 3: Trip variables distribution in the sample

Variable	Description	Sample (n=35,936)
Distance measured	Mean	3.29 km
Main mode	Bicycle Car-driver Car-pass. Coach Public transport Train Walk Other	4.22% 31.62% 11.01% 3.84% 16.05% 16.31% 31.08% 5.23%
Destination (activity) purpose	H: Home W: Work S: Shopping, Services L: Leisure E: Education M: Personal B: Business/work trips F: Food O: Other	42.77% 19.78% 10.12% 8.34% 6.39% 7.66% 2.35% 10.38% 15.22%

3. THE BINOMIAL REGRESSION MODEL

3.1. Analyses

The framework, used for estimating the propensity of Czech residents to cycle is based on and extended binomial regression model, that tries to improve the goodness of fit when predicting bicycle mode choice, based on additional factors. The binomial regression model was fitted through pscl package (Jackman, 2008).

The model, estimating the choice of bicycle for the Czech residents has been constructed as:

$$P_n = \log\left[\frac{P(bicycle = yes)}{1 - P(bicycle = yes)}\right] = \alpha + \beta_1(T_d) + \beta_2(H_v) + \beta_3(r_x) + \beta_4(C_x) + \beta_5(P_c) + \beta_6(H_d) + \beta_7(d_{px}) + \beta_8(T_m) + \beta_9(Slope_a) + \beta_{10}(P_a) + \beta_{11}(W_m)$$

Where probability P for person n of choosing the alternative (bicycle or non-bicycle mode) is a linear combination of β_{1-p} predictors, that is a vector of parameters estimating the probability of choosing the alternative.

The following predictors/categories were considered, based on existing theory and available data:

1. **Land-use attributes, urban morphology:** the diversity, density and design of the origin/destination area are hypothesized to increase the probability to choose bicycle as a main mode of transport. Average slope of the streets of the city or mean altitude of the city may play a role in mobility choices.
2. **The urban bicycle/sustainable mode policy attributes proxies:** The density of network of either segregated bicycle network or mixed bicycle/pedestrian network in the origin/destination area. The ratio of traffic-calmed streets (less than 50 km/h) to the residential streets length may be correlated with more favourable conditions for the cyclists. The ratio of pedestrian network to residential streets length.
3. **Extended costs of the trip, where possible to identify:** steepness, altitude difference, detour factor, overall length, overall time.
4. **Character of the trip:** intra or inter-urban; destination activity (purpose) – considering utilitarian vs. non-utilitarian trips.
5. **The traveller's personal attributes:** car/bicycle and driving license availability, income, gender, age.
6. **Safety:** Areas with higher share of serious RTA (with/without pedestrian/cyclist victim) are hypothesized to reduce propensity to choose bicycle as a mode of transport.
7. **Weather**: While tracing back weather patterns for the concrete locations several years back would prove hard to obtain and imprecise due to the publicly available data structure, average annual weather/temperature variations have been assumed.

The following variables have been selected as of yet, based on available or derivable data, valid with reasonable confidence:

Table 4: Comparison of additional potential predictors for bicycle mode choice

Identified predictors	**Category**	**Operationalization**	**Availability / Source**
Traffic calming	Urban policy	Share of speed limited (< 50 km/h) streets in the city of residence (%)	GN 2018
Quality of cycling infrastructure	Urban policy	Share of cycling infrastructure per non-service roads in the city	GN 2018 / OSM
Road grading	Geography	Average slope of distinct road segments in the city	GN 2018
Safety	Risk perception	Traffic accidents severity index	Police of the Czech Republic
POI density	Urban structure	Number of non-transportation related POIs in the vicinity of origin of the trip (< 300 m)	GN 2018/OSM
Weather	Comfort	Months of July/August	-

Segregated, any or contra-flow cycling infrastructure is rare in Czech cities: the contra-flow lanes or roads don't exceed length proportion of 2.5 % in any of the cities, i.e. the policy is scarce even in the cities that adopt it (when the general recommendation of universal adoption has been debated (Methorst, 2017)). The indicator has not been included in the model, since its validity cannot be estimated due to the impossibility of obtaining accurate historical data for the sampled year.

The urban design, density and diversity variables, along with public transport accessibility, were disregarded for the time being, as different approaches to their operationalization have yielded wildly varying results.

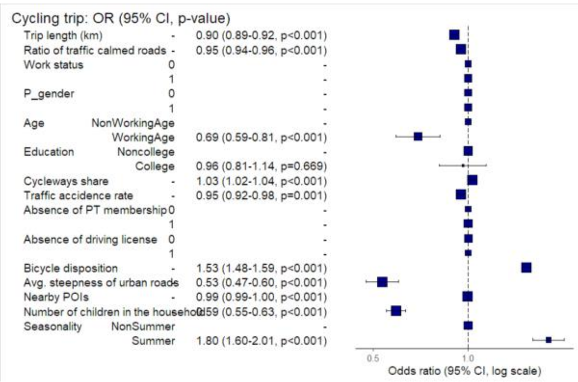


Figure 2: OR plot of factors for binary mode choice

The following table shows the results of the fitted model without beta standardization of the coefficients:

Table 5: Binomial logit model for predicting bicycle only mode choice in Czech population

	Model I	Null Model
(Intercept)	-2.46 ***	-3.38 ***
	(0.13)	(0.09)
Distance travelled	-0.10 ***	-0.10 ***
	(0.01)	(0.01)
	(0.00)	
Work status	0.38 ***	0.37 ***
	(0.09)	(0.08)
Gender	-0.24 ***	-0.23 ***
	(0.06)	(0.06)
Productive age	-0.37 ***	-0.37 ***
	(0.08)	(0.08)
College Education	-0.04	-0.16
	(0.09)	(0.09)
Absence of public transport membership disposition	0.27 ***	0.39 ***
	(0.07)	(0.06)
Bicycle availability	0.43 ***	0.44 ***
	(0.02)	(0.02)
Absence of driver's license	0.33 ***	0.33 ***
	(0.07)	(0.07)
Number of children in the household	-0.53 ***	-0.55 ***
	(0.04)	(0.04)
Weather (summer)	0.59 ***	
	(0.06)	
Share of segregated cycling infrastructure in the origin city cadaster	0.03 ***	
	(0.01)	
Traffic accident severity index of cyclists (Reinhold) per population in the origin TAZ	-0.05 ***	
	(0.02)	
The average slope of streets in the city of origin	-0.63 ***	
	(0.06)	
POI count in the origin TAZ	-0.01 ***	
Share of speed limited (< 50 km/h) streets in the city of origin	-0.05 ***	
	(0.00)	
nobs	35125	35203
null.deviance	12420.22	12427.04
AIC	10775.40	11163.55
BIC	10910.87	11248.24
df.residual	35109.00	35193.00
Goodness-of-fit (pseudo-R^2)	0.16	0.12
McFadden	0.14	0.10
χ^2	1676.82 (15)	1283.49 (9)

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

The model performs better, than the null model, albeit not significantly. This is not surprising, given the variability of cities and the individual variability that exceeds the explanatory potential of aggregated characteristics. To a large extent, cycling shares the effect of the observed factors with walking, which significantly reduces the presicion of the model, but creates a space to explore the synergy between the two modes and the possible co-dependence of their promotion.

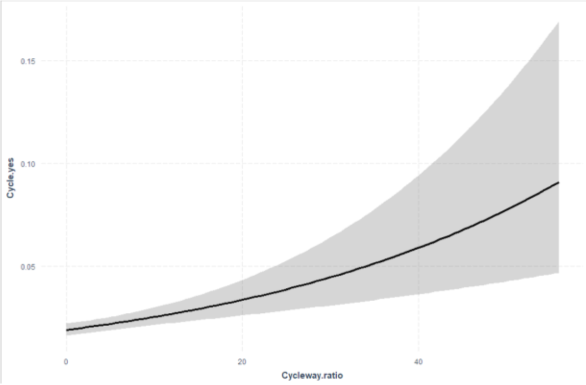


Figure 3: Modelled effect of segregated cycle infrastructure share on mode choice.

The identified infrastructural factors do not explain a significant part of the variability, ie. the macroscopical reasons, why people in some municipalities cycle more frequent and for longer trips, than others. This is not surprising, given both the unreliable and aggregated character of the data on the cycling infrastructure, and their variability on the national level, even in design, which is highly city-specific.

Additionally, there is a non-standard distribution of the used indices, such as bicycle infrastructure, or traffic calming measures ratios. These measures are typically present either in a larger scale, or, for the majority of municipalities, are non-existent, leading to significantly right-skewed distribution. This presents an additional potential for model and spatial aggregation improvement. Since multiple observations (trips) have been made for each individual, the generalized linear model that assume independence of the observations might not be the ideal solution.

There is a significant potential for the model improvement regarding several factors that may be considered in follow-up: weather (obtaining precise data for the specific trips, regarding rainfall, temperatures with high variability due to continuous design of the study); slope (obtaining estimated slopes and cumulative ascent/descent for each trip).

3.2. The potential for model and data collection improvement

3.2.1. Road network

Apart from classification problems, the road network lacks more detailed information on non-segregated cycling infrastructure. Crowdsourced projects, such as OSM, tend to improve both in the depth of detail, and in the completeness of mapping of the current state of the infrastructure, but the overall level of accuracy (completeness and classification precision) is impossible to estimate, and might vary across the country and across different categories of data.

The road network used (CEDA, 2018) is collected twice a year for the use of The Road and Motorway Directorate of the Czech Republic, responsible solely for management, security, administration, maintenance and repair of motorways and main roads, hence the lack of details on the cycling infrastructure. Respective municipalities conduct independent surveys, collect and maintain the data on the infrastructure in their own structures, and are required to inventarize their own infrastructure.

Direct demand models for bicycle use (e.g. (Fagnant, 2016)), or models based on cycling traffic level stress (such as BLOS) usually utilize more precise data on factors, such as cycle lane width or type, traffic intensities, or weather.

There is an overlap between theoretical models and increasing demand for better cycle routing (stemming from the evolution of cycling-specific navigation). This has led to a steady co-evolution of public domain data on the infrastructure (such as OpenStreetMaps), private-domain, pay-per-request data (such as HERE or Google API) and land use and their usage in applications, allowing for adding more and more cycle or walk-specific criteria to the routing choices, such as safety, elevation, avoidance criteria, amenities, or even real time traffic data.

3.2.2. Accidents

(Bílová, Bíl, & Müller, 2010), analyzing cyclists' fatalities in Czech republic have identified car speeding as the most significant cause of cyclists' fatalities, where segregated infrastructure and/or enforced traffic calming, primarily on the critical "hotspots" is identified as the main (or rather the only) solution.

The proposed model observes a relationship between (cycling) mode choice and overall, geographically aggregated accidents rate, where higher accidents rate is hypothesized to deter potential cyclists and correlate with more dangerous road conditions in general. While higher cycling rates lead to higher absolute incidence rate, it is probable, that the relative incidence rate *decreases* with higher modal share.

3.2.3. Routing

Due to various considerations, it is not practical for travel diaries to be geolocated or routed via GPS aided devices (Kouřil, 2020). The data collected via geolocating applications (such as Strava) additionally usually suffer from a lack of representativeness of the sample (e.g. bias towards certain socio-demographic groups) (Lee, 2021).

The process of geolocation of the origin-destination points and the routes of travel diaries in the NHTS is highly dependent on the completeness of the respective travel diaries, in the latter case joined by uncertainty about precise routing due to variability of the individual choices, or other factors, that are hidden from the sight and do not follow the shortest route preference (e.g. joining several activity-trips into one trip, detouring or delaying for various non-utilitary reasons). Arguably, the very principle of forcedly utilitarian route and activity choice discriminates against the active modes, that from their nature deny such classification.

3.2.4. Weather patterns

Especially active modes exhibit variable elasticity for changing weather patterns, which may be locally specific and correlate with the network quality and age (Goldmann, 2020). In case of uncontrolled weather variables, this further deepens the crevasse between the "cyclist-friendly" cities and the rest.

All in all, we should not expect the model to perform significantly well, given the identified limitations. Still, the model presents conclusions, that allow for a modestly confident inclusion into prospective urban sustainable mobility policies.

3.2.5. Urban structure

While the model incorporates a generic correlate of POI accessibility in the direct vicinity of the origin, a more detailed approach to the factors such as land-use density, amenity accessibility, road network density/block density along the routes or at the origin/destination TAZ would yield more detailed results and a better representation of the dimension.

3.3. Discussion and inclusion into the MLP model

The distance travelled has a negative, or rather cutoff effect on choosing the bicycle as a mode for transport. The means to attain higher cycling levels are in general connected to the *short-distance city* measures, ie. urban structure densification, yet pervious – connected via direct, safe routes. Technological innovation penetration, such as pedelecs, tend to further raise the threshold of willingness to commute both for the distance and slope of the route.

Contrary to the popular belief, size of the households is positively associated with bicycle use, while number of children in the household is only mildly negatively associated with bicycle use – mainly for four children or more.

While cycling is a popular mode for leisure (be it as a leisure activity itself, or as a means to it), the dominant purpose of bicycle use remains work (15.8 to 20.76 %).

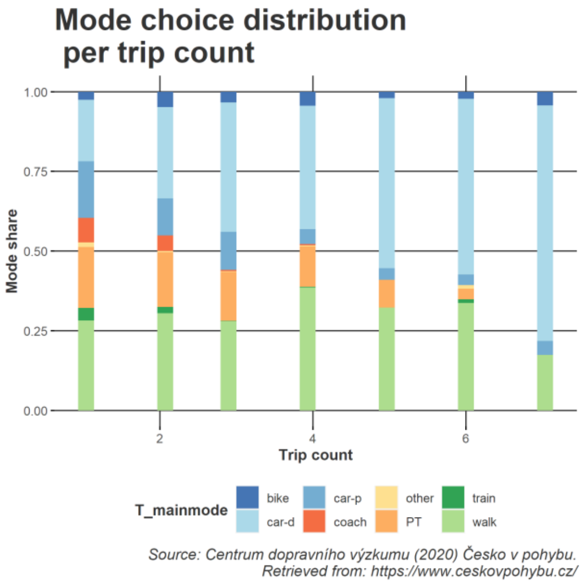


Figure 4: Mode choice distribution per trip count.

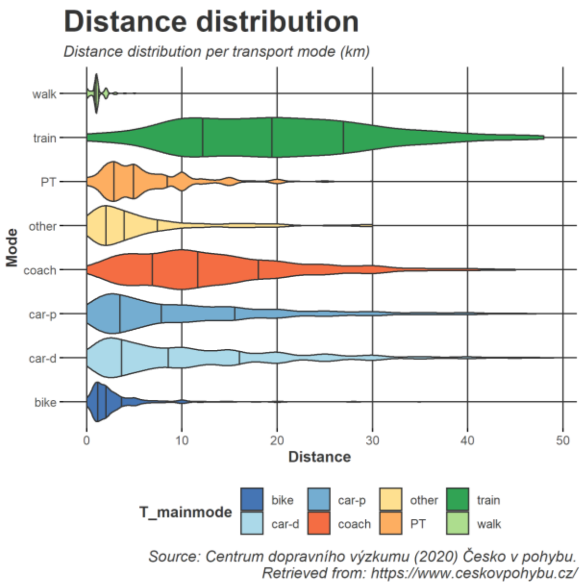


Figure 5: Measured distance distribution per transport mode

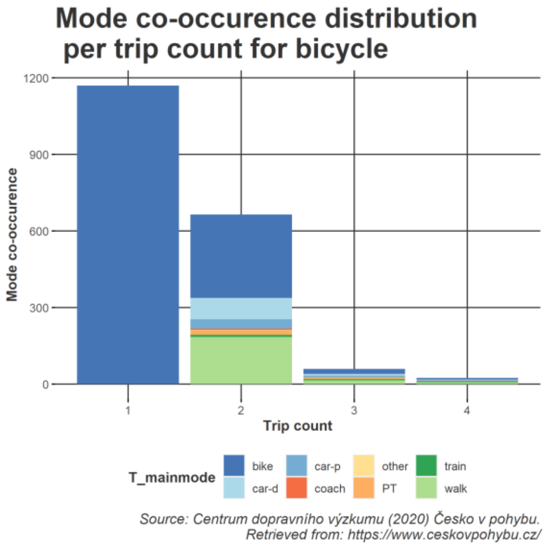


Figure 6: Mode co-occurence distribution per trip count for bicycle

There is a stronger correlation between trip count (personal mobility ratio) and mode choice (Figure 2); with the former increasing, the probability of choosing car also increases. The co-occurrence of modes within-person or within-household (i.e. choice homogeneity) has not been analyzed, mainly due to the overall low number of trips recorded per person.

For cycling persons, the prevalent co-occuring mode is walking, regardless of the number of overall trips. Along with the distribution of cycling-only persons, it appears, that cycling-active modes are *the primary mode* of transportation for the majority of those persons, who cycle – or at least on the day they cycled. This defies a general notion of bicycle as a spare time, or secondary vehicle (Figure 4).

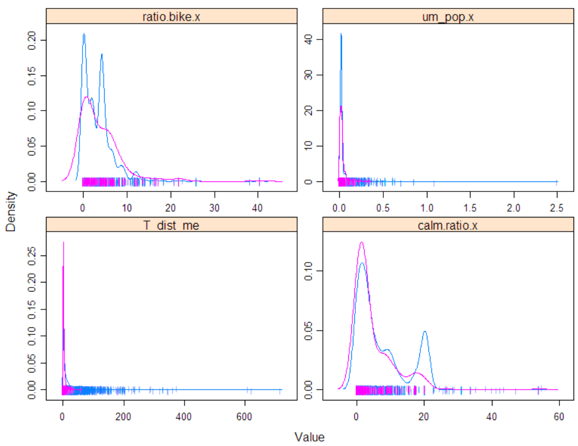


Figure 7: The density distribution of binary choice (blue=non-cycle) per a) share of cycling infrastructure b) traffic fatalities per population c) distance travelled d) traffic calming share (left top to bottom right)

As for the road infrastructure, the paradoxical relation between traffic calming measures and mode choice could be explainable by the character of significant portion of the traffic-calmed zones, such as living streets, that are usually located in the suburbian, highly car-dependent neighbourhoods (as shown in the spike in the Figure 5 often with no additional changes to the street profile. The results call for further analysis of covariance of additional urban morphology measures, in order to identify the relationship more precisely. As traffic calming is one of the preferred means to achieve both lower road traffic injuries and deaths and higher active mobility rates, a close attention should be paid to the form, placement, scope and enforcement of the calming measures.

On the contrary, segregated bicycle infrastructure has positive, if small impact on the bicycle mode choice. This is a likely result of mutual support of the mass demand and infrastructure provision. On the other hand, spatial conflicts with other land uses (such as road infrastructure) push cycling infrastructure on the fringes of effectivity (i.e., long routes, distanced and disconnected from the urban core, recreational cycling only), thus reducing the estimated effect. Alas, the cycling infrastructure in Czechia is too scarce and disconnected to estimate more detailed aspects of usability on the network level.

Number of traffic fatalities and injuries on bicycle-allowed roads has a negative impact on the bicycle mode choice. It is presumed, that segregation and traffic calming are the most effective measures for reducing both physical risk of cyclists' injuries and the perceived risk when choosing whether to cycle and where to cycle.

Availibility of a bicycle (number of bicycles in the household) is also positively associated with bicycle mode choice. While the significance of the parameter is usually understated in traffic modelling, not having the option to cycle in a first place is naturally the most probable barrier, non-cyclists will face. While growing number of policies or initiatives worldwide are targeting this either via tax deduction schemes or direct subsidies with significant results (e.g. (Caulfield & Leahy, 2011)), there has not yet been such effort on national level in Czech republic. Another means for targeting new users are bike sharing schemes, albeit these operate to a limited degree, predominantly in the larger cities.

The analyzed predictors thus can, to a degree, form the ground for inclusion into a MLP model, considering the mutually co-evolving scape of actors and processes.

A more precise analyses in the vein of national *bicycle level of service* equivalent (as is present in several nations) could yield more precise elasticities for wider range of factors, that could be considered as a predictive tool for possible variances and developments in bicycle services provision.

4. THE MULTI-LEVEL PERSPECTIVE MODEL

Yet, the bicycle services provision is firmly rooted in the long-term constellations of actors and processes, that have established their positions and have a significant influence on the public policies.

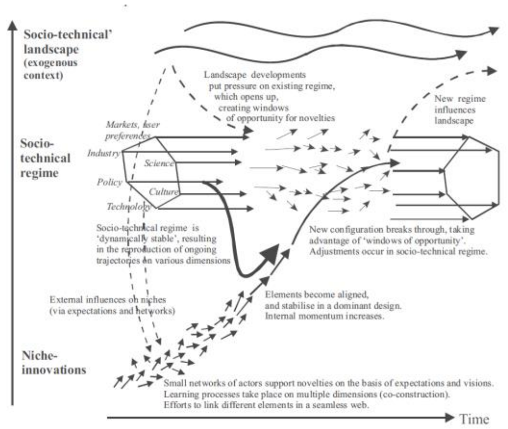


Figure 8: Conceptualizing political defection (represented by the fat arrow) in the multi-level perspective on transitions. Source: Geels and Roberts 2019:3.

The *multi-level perspective on sustainability transitions* has been developed mainly by F. W. Geels, mainly focusing on energy systems (low-carbon) transitions and mobility systems transitions (Geels F. W., 2007).

Originally and dominantly a historical analysis, the framework has evolved into several strains and approaches, and has been also used to analyze current trends and unfolding pathways in innovation-lock-in dynamics and policy implications (Geels F. W., 2019). In recent years, the framework has gained popularity in identifying the dynamics of the mobility regimes and their stabilizing and destabilizing factors, as a means to analyze potential for sustainable mobility transitions (Hopkins, 2017). Recently, the framework has been adapted for analyzing public policy (policymakers') adaptation to the system evolution and feedback via crossover with historical institutionalism analysis (Roberts & Geels, 2019).

The MLP analysis, operating on different levels (micro-niche, meso-regime and macro-landscape) is employed to conceptualize the interactions between two possible regimes – one, the incumbent automobility regime, and the second, the promising *velomobility* regime, that is represented by rapidly growing number of bicycle-centered innovations on technological-vehicle level, infrastructure and organizational level, and on practice-culture level.

4.1. Barriers and pressures for and against velomobility

4.1.1. Infrastructure development

As has been mentioned above, the development of cycling infrastructure has been stagnant both on the spatial and temporal level, with only recent and length-wise relative spikes in development. Both regional and municipal level strategies have since 2007 renewed focus on recreational cycling as the main driver of new infrastructure construction, being financed from national and European funds.

4.1.2. Policy changes

The pressures towards new infrastructure have resulted in amendments of technical standards for cycling infrastructure, that draws broadly from the Dutch Design manual for cycling traffic (CROW).

While on the municipal level, since early 2000s, several Czech cities have adopted newly conceptualized *Plans for the development of cycling infrastructure*, their contents and realization have been a subject of permanent public and political struggle for space, along the lines of spatial conflicts, traffic safety and control.

Since 2017, the first *Sustainable urban mobility plans* (SUMP) have been created in cities with population exceeding 40,000, where the plans are seen as a condition for receiving public funds for infrastructure development. While the principles of the SUMPs promote sustainable mobility first approach, there are no mechanisms present that would enforce either inclusion of cycling measures, or condition development of road infrastructure and parking provision with sustainable measures. For example, while parking regulation is more and more present in the city cores, this is a result of internal pressures between commuters and residents, short-term and long-term parking, that leads little to no results in freeing up space for alternative modes.

On the other hand, SUMPs shed some light on the soft mobility patterns on the urban level, and their potential for a growth, possibly countering the narrative of efficient urban mobility system as one, based on automobile.

4.1.3. New bicycle niches and municipal policy feedbacks

New institutional arrangement

Several new forms of velomobility have pervaded in recent years, meeting with mixed responses from the policymakers:

- Informal cycling clubs and bike sharing schemes have been formalized in many cities, mainly through entering the market. Public transport services providers and municipalities are beginning to cooperate with two types of private companies:

- **The communal bike clubs and bike sharing schemes,** whose impact tend to go beyond the most basic sharing model, as they are more dependent and responsive to the local environment. Such schemes are more active in provisioning additional services (such as bike repair workshop, lectures, critical masses), and are significant actors in their own respect in the process of planning and constructing new infrastructure. They actively participate in the planning processes, be it via consultation and expertise, helping the municipalities e.g. with project preparation – or, on the other hand, lawsuits.
- **The private companies' bike-sharing and renting schemes** draw from a market-based "predict and provide" approach. Since the rules for providing these services (such as using publicly available cycle parking) vary, these services tend to achieve early market penetration and domination also via formal or informal agreements with the local municipalities.

- The sporadically filled positions of urban *coordinator of cycling mobility* are associated with more systematic action for supporting velomobility and inclusion of its facilities into projects (such as new development or road reconstruction) on both on small and large scale. On the other hand, the added costs of an optional position are spreading the budgets of many, especially smaller cities thin; and without strictly formalizing the responsibilities, the coordinator has to battle for acceptance in the decision making – or for even being invited to the decision making at all. This has been accompanied by a growing number of plans and strategies for the development of cycling infrastructure for a growing number of cities and regions; although these also struggle with deeper inclusion into the land use and transport planning in general, mainly reserving to isolated projects (cycleways, parking provision), rather than transforming the general transportation plans approach and practice.

These arrangements have added to an increasingly stable and growing institutional substrate for applying the principles for dense and protected, connected and direct cycling infrastructure and cycling facilities provision, as stated in the technical norms.

New technological arrangements

- *E-bicycles* are rapidly gaining popularity due to increasing effectivity, range of use and availability. Given the high effect of hilliness on the probability to cycle in general, pedelecs are highly relevant in overcoming the distance and altitude effort barriers, accounting for the value of time as well.
- *Cargo bicycles* have achieved a small, but significant niche emergence via delivery and postal services, where the impetus for change stems rather from operational costs and international commitments than from local pressures, albeit there have been multiple feasibility studies conducted for urban consolidation centres. On the individual level, rapid growth in effectivity and decline in the price of electric cargo-bicycles has made both privately owned or shared cargo bicycles more broadly available, allowing for increased flexibility.
- Unclear or unevenly enforced legislation, concerning pedelecs, scooters and other types of small electric vehicles have seen the vehicles' counts rise and drop sharply where drawing pushbacks on the municipal level, or simply falling out of the hype cycle.

Regarding cycling measures, the municipalities often face internal conflicts between sectors, that represent different interests – local or national police department, building authority, land use plans, political parties or traffic safety engineers, local residents organizations and factions.

The crucial factor here is the non-existence of *standard cycling infrastructure provision* – albeit such is codified through technical norms and standards in increasing detail, it is not present or prevalent in the streets. The scarcity of actual (any) or high-quality cycling infrastructure is shaping both users' expectations and behaviour and increases capacity for both on and off-the-road conflicts – there might be no "one size fits all" for cycling promotion in general – as travelling patterns, experience and hence requirements radically differ – but there certainly is one for the core safety and usability of cycling infrastructure.

This has been recently exemplified on a national level, when a call for codified safe passing distance and other protecting measures has reverberated in both legal, technical and public debate, following such measures in neighbouring states (e.g. Germany). Nevertheless, the rationale for the safety measure – the main reason for not passing such law, is that the *impossibility* of complying with the safe passing distance is hardwired into the *current built form of the roads*, and therefore unenforceable.

This illuminates the main both strengthening and weakening factors of the velomobility niches. Cycling is engrained to a large extent as popular leisure or sport, weekend activity, that is not in contradiction with car use (except for occasional road conflicts, that can be subjectively attributed to individual behaviour) – as is after all discourse, preferred by major Czech car manufacturers. On the other hand, *utility, everyday cycling*, that enters the domain of both space and traffic rule organization and affects the motorized transport flows and capacities, is considered a structural, coordinated network, that undermines the hard-fought victories of roads provision.

4.1.4. Feedbacks of the road transport infrastructure development policies

While most of the settlements have been developed in close connection to busy roads, continuous increase in road transport intensities have been steadily pressuring local municipalities into negotiations on regional and national level to plan and construct bypasses and/or ringroads. Indeed, presence of higher-category road is a prerequisite for multiple traffic calming policies, such as Low-emission zones (LEZ), but more importantly, the delayed promise of *complete infrastructure* allows for postponing other sustainable urban mobility policies for at least several decades. Even when the municipalities are prone to support sustainable mobility measures, the system's orientation on expansion of the road network is draining their capacity, pressuring them to join forces in the coalitions in the favour of the construction of the new road. Only where the variants of the possible development lead to uneven distribution of externalities (such as noise levels or air pollution levels for different localities), the actors in conflict reach out to technical or scientific knowledge, in order to strengthen their position – and hinder the development – either way, there is no tangible contra-force to the primacy of constructing new roads as the ailment for motorized transport externalities.

This has been made tangible in the omission of velomobility (or sustainable mobility for that matter) from the Czech National recovery plan. Moreover, there has been a growth of share in the national budget present for the years 2020 and 2021, that has seen a record investment into transport infrastructure, prevailing over other sectors reforms, again lacking any substantial sustainable mobility support. At the same time, the Covid-19 pandemic has resulted in unforeseen losses in public transport, including the services provided or covered by municipalities that are continually destabilized, reducing their services frequency, coverage, accessibility and capacity.

The current crisis thus can be paradoxically identified as *rapidly accelerating* the dominance of the automobile. The crisis and subsequent recovery promise have allowed the incumbent coalition actors to rally round the non-controversial road construction, rather than to succumb to its nevertheless growing instability (due to significant landscape pressures towards net carbon neutrality).

5. DISCUSSION

The lack of detail of the regression model, based on the national level still results in several departure points, from where the current and emerging forms and determinants of velomobility can be studied.

Velomobility currently faces the main barrier of low, uneven infrastructure coverage, when compared with any other major transportation mode. This state has been sustained by perpetual spatial conflicts with both motorized and pedestrian-designated space and lack of attention to the details in their realization.

While several promising niches emerge, to an extent supported on national and transnational level through expertise sharing and financial support on European level, there appear to be several caveats, that reject the hypothesis of looming regime change or major breaks:

- The conditions for rapid growth of velomobility have been arguably the most ideal in the recent era, due to significant changes in mobility patterns. However, the new influx of travellers resulting from disruption of public transport services appears to favour car-based mobility. This is partially due to the structure of origin-destination locations (density, structure, work/home choice considerations), resulting from long-term defined and stabilized mobility patterns, to a large extent appear beyond the individual threshold for utility cycling.
- The propensity to develop new cycling infrastructure is to a large extent predicted by the urban morphology – preferring wide corridors – negatively correlating with the bicycle-friendly (i.e. dense and polyfunctional) urban morphology and cycle-friendly street profiles (narrower streets, allowing for freely dispatching and crossing the streets). This renders any more significant changes towards *reducing* the road space in favour of active modes (i.e. road diets) more unlikely.
- The opportunity to take up the newly available road space for newly designated cycling infrastructure has been largely missed and restricted to the existing plans.
- Given the low prevalence of cycling infrastructure even in the majority of the cities, where the measures are broadly supported, the form of infrastructure-

behaviour patterns has not been stabilized, resulting in conflicts in use (e.g. pedestrian-cyclist conflicts on shared-use paths) and conflicts regarding its presence (prevalence of efforts to scrap the newly established contra-flow cycling roads/lanes in several cities).

- The existing strategies on both national and municipal level to support urban cycling lack measurable targets and clear plans for their achievement, even where the means for achieving the support couldn't be more clear.

The accelerating rate and unrelenting governmental focus on road infrastructure growth, that dwarfs investments into cycling infrastructure, will have far-reaching consequences for the prospects of sustainable mobility. The projects will not be unbuilt; the resources will not be recovered; on the contrary – road construction begets road construction. We return to the Nello-Deakin's (Nello-Deakin, 2020) proposed turn to researching the traffic disappearance as the main theoretical and practical impetus for promoting the viability of cycling infrastructure planning; where even the possibility of reducing the car utilization rates is unaccounted for, the sustainable modes are silently omitted.

However, there is a sliver of hope, that even if not imminently destabilizing, the new prospects in active and hybridized velomobility (more acceptable by the regime insiders) will continue to garner popularity and shape the mobility and built environment patterns. The local initiatives are gaining momentum. More research is needed on whether – and how this power can be facilitated.

This article was produced with the financial support of the Ministry of Transport within the programme of long-term conceptual development of research institutions.

REFERENCES

[1] Armoogum, J. e. (2014). *Survey Harmonisation with New Technologies Improvement – SHANTI.* Les collections de L'INRETS.

[2] Baltus, R., & Rubbers, A. (1999). Piping reliability improvement through passive seismic supports . *Nuclear Energy in Central Europe '99* (pp. 251-258). Portorož, Slovenia: Nuclear Society of Slovenia.

[3] Bílová, M., Bíl, M., & Müller, I. (2010). Critical factors in fatal collisions of adult cyclists with automobiles. *Accident Analysis & Prevention,*(6), pp. 1632–1636. doi:https://doi.org/10.1016/j.aap.2010.04.001

[4] Caulfield, B., & Leahy, J. (2011). Learning to cycle again: Examining the benefits of providing tax-free loans to purchase new bicycles. *Research in Transportation Business & Management*, pp. 42–47.

[5] CEDA. (2018). Global network v2018v.6. Czechia.

[6] CEDA. (2020). Global Network v2020.06. Czechia.

[7] Centrum dopravního výzkumu. (2020). Česko v pohybu. Retrieved from https://www.ceskovpohybu.cz/

[8] Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D 7* , pp. 265–284.

[9] Dendy Jr., J. E., Swartz, B., & Wendroff, B. (1977). Computing travelling wave solutions of a nonlinear heat equation. In J. H. Miller (Ed.), *Topics in Numerical Analysis* (Vol. III, pp. 447-463). London: Academic Press.

[10] Drabicki, A., Kucharski, R., & Szarata, A. (2020). Suppressing the effects of induced traffic in urban road systems: impact assessment with macrosimulation tools - results from the city of Krakow (Poland). (F. S. Esteve Codina, Ed.) *Transportation Research Procedia*, pp. 31-138. doi:https://doi.org/10.1016/j.trpro.2020.03.085

[11] Fagnant, D. J. (2016). A direct-demand model for bicycle counts: The impacts of level of service and other factors. . *Environment and Planning B: Planning and Design*(1), pp. 93–107. doi:https://doi.org/10.1177/0265813515602568

[12] Flyvbjerg, B., Holm, M., & Buhl, S. (2007, 11 26). How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation. *Journal of the American Planning Association*, pp. 131-146.

[13] Gabrhel, V. (2019). Feeling like cycling? Psychological factors related to cycling as a mode choice. *Transactions on Transport Sciences*(1), pp. 19-30. doi:https://doi.org/10.5507/tots.2019.006

[14] Geels, F. W. (2007). Transformations of Large Technical Systems: A Multilevel Analysis of the Dutch Highway System (1950-2000). *Science, Technology, & Human Values*(2), pp. 123–149. doi:https://doi.org/10.1177/0162243906293883

[15] Geels, F. W. (2019). Socio-technical transitions to sustainability: A review of criticisms and elaborations of the Multi-Level Perspective. *Current Opinion in Environmental Sustainability*, pp. 187–201. doi:https://doi.org/10.1016/j.cosust.2019.06.009

[16] Goldmann, K. &. (2020). Some people feel the rain, others just get wet: An analysis of regional differences in the effects of weather on cycling. . *Research in Transportation Business & Management*. doi:https://doi.org/10.1016/j.rtbm.2020.1005

[17] Hoffmann, S. W. (2017). Discontinuation of the automobility regime? An integrated approach to multi-level governance. *Transportation Research Part A: Policy and Practice*, pp. 391–408. doi:https://doi.org/10.1016/j.tra.2017.06.016

[18] Hommels, A. (2005). Studying Obduracy in the City: Toward a Productive Fusion between Technology Studies and Urban Studies. *Science, Technology, & Human Values*(3), pp. 323–351. doi:https://doi.org/10.1177/0162243904271759

[19] Hopkins, D. (2017). Destabilizing automobility? The emergent mobilities of generation. *Ambio*(3), pp. 371–383. doi:https://doi.org/10.1007/s13280-016-0841-2

[20] Jackman, S. (2008). pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University, Stanford, California. Retrieved from http://CRAN.R-project.org/package=pscl

[21] Katz, R. (1995). Modeling Bicycle Demand as a Mainstream. *Transportation Research Record*, pp. 12-28.

[22] Konopásek, Z., Stöckelová, T., & Zamykalová, L. (2008). Making Pure Science and Pure Politics: On the Expertise of Bypass and the Bypass of Expertise. *Science, Technology & Human Values*(4), pp. 529-553. doi:10.1177/0162243907306699

[23] Kouřil, P. &. (2020). Usability of Wi-Fi fingerprint approach for place departure recognition in travel surveys. *Travel Behaviour and Society*, pp. 83–93. doi:https://doi.org/10.1016/j.tbs.2019.10.004

[24] Langhaar, H. L., & Chu, S. C. (1970). *Development in theoretical and applied mecanics.* New York: Pergamon.

[25] Lee, K. &. (2021). Strava Metro data for bicycle monitoring: A literature review. . *Transport Reviews*(1), pp. 27–47. doi:https://doi.org/10.1080/01441647.2020.1798558

[26] Maier, K. (2012). Nástroje územního plánování k regulaci suburbanizace. *Urbanismus a územní rozvoj*.

[27] Methorst, R. S. (2017). Can cycling safety be improved by opening all unidirectional cycle paths for cycle traffic in both directions? A theoretical examination of available literature and data. *Accident Analysis & Prevention*, pp. 38-43. doi:https://doi.org/10.1016/j.aap.2016.05.018

[28] Millard-Ball, A. &. (2011). Are We Reaching Peak Travel? Trends in Passenger Transport in Eight Industrialized Countries. . *Transport Reviews* (3), pp. 357–378. doi:https://doi.org/10.1080/01441647.2010.518291

[29] Nello-Deakin, S. (2020). Environmental determinants of cycling: Not seeing the forest for the trees? *Journal of Transport Geography*. doi:10.1016/j.jtrangeo.2020.102704

[30] Roberts, C., & Geels, F. (2019). Conditions for politically accelerated transitions: Historical institutionalism, the multi-level perspective, and two historical case studies in transport and agriculture. *Technological Forecasting and Social Change*, pp. 221–240. doi:https://doi.org/10.1016/j.techfore.2018.11.01

[31] Roberts, C., & Geels, F.W. (2019). Conditions for politically accelerated transitions: Historical institutionalism, the multi-level perspective, and two historical case studies in transport and agriculture. *Technological Forecasting & Social Change*, pp. 221-240.

[32] Rybarczyk, G., & Changshan, W. (2014). Examining the Impact of Urban Morphology on Bicycle Mode Choice. *Examining the Impact of Urban Morphology on Bicycle Mode Choice*(2). doi:https://doi.org/10.1068/b37133

[33] Shoup, D. (2011). *The High Cost of Free Parking.* Routledge.

[34] Urry, J. (2004). The 'System' of Automobility. *Theory, Culture & Society*(4-5), pp. 25–39. . doi:https://doi.org/10.1177/0263276404046059

[35] Vo, T. V., & Edwards, D. R. (1994). Development of In-Service Inspection Priorities for PWR. *Nucl. Technol., 106*(110), 253-259.

PLACING HOURLY VARIATION IN BICYCLE VOLUME AND CRASHES IN THE BIGGER PERSPECTIVE OF ROAD SAFETY FOR CYCLISTS IN DUTCH CITIES

Teun Uijtdewilligen, MSc
SWOV Institute for Road Safety Research
PO Box 93113, 2509 AC The Hague, The Netherlands
teun.uijtdewilligen@swov.nl

ABSTRACT

While bicycle levels in Dutch cities increased rapidly the last few years, it is still debatable what the effects of increasing bicycle volumes are on road safety. On the one hand, a safety-in-numbers effect exists, meaning that increasing cycling levels lead to a less increasing risk for cyclists. On the other hand, recent studies found a hazard-in-numbers effect, revealing that more cyclists on the road leads to a large increase in serious bicycle crashes and a slight increase in bicycle fatalities. A forthcoming study by the author contributes to this debate by investigating the effect of hourly variation in bicycle and motorised traffic volume on bicycle crashes related to type of bicycle facility. But why is hourly variaton in bicycle traffic so important to investigate road safety for cyclists in Dutch cities, the author's PhD topic? This is discussed in the present paper. Broadly speaking, most studies related to bicycle crash risk make use of the average annual daily bicyclists and the average annual daily traffic to predict bicycle crashes, barring additional factors. This is very helpful to understand the effects of bicycle volume and motorised traffic volume on road safety for cyclists. However, such analyses do not grasp the temporal variation in daily bicycle volume, daily traffic volume and bicycle crashes. To analyse varying loads of the road network at different times of the day, and how this relates to road safety for cyclists more detailed bicycle and traffic volume data must be used. The present paper discusses how this could help for bicycle safety studies in Dutch cities, by evaluating the usefulness of hourly bicycle and traffic count data from count stations in a mature cycling city in the Netherlands.

Keywords:
road safety, bicycle volume, cycling, safe cycling infrastructure, urban road network

1. INTRODUCTION

After years of progression, road safety becomes a growing problem for cyclists in the Netherlands again. Over the past decade, the number of bicycle crash related casualties increased. While there were 162 bicycle fatalities in 2010 (25.3% of all road fatalities), in 2018 there were 228 bicycle fatalities (33.6% of all road fatalities) (CBS, 2019; OECD, 2018; Weijermars et al., 2019). Furthermore, the number of seriously injured cyclists increased as well (43% of all reported serious road injuries in 2008 against 53% of all reported serious road injuries in 2017), especially caused by the increasing number of single-bicycle crashes (SWOV, 2018; Wegman et al., 2012; Weijermars et al., 2019).

At the same time, two trends happen in the four major Randstad cities in the Netherlands: Amsterdam, Rotterdam, The Hague and Utrecht. Firstly, these cities experience rapidly increasing numbers of inhabitants and, subsequently, become more densely populated (de Jong & Daalhuizen, 2014). These new inhabitants have to find their way in the city. Secondly, bicycle use in the Netherlands increases, especially in the four major Randstad cities (Harms et al., 2014; Jonkeren et al., 2019). As a result, the Dutch vehicle fleet becomes more bicycle, and in particular e-bike, oriented (BOVAG, 2019). These two processes combined lead to an intensely used road network and have unknown implications for road safety of cyclists. Despite these unknown implications, the fact is that the number of severe and fatal bicycle crashes increases (CBS, 2019; Weijermars et al., 2019). This shortly summarizes the main topic of the author's PhD research, which investigates road safety for cyclists in Dutch cities.

The effect of increasing bicycle volumes on road safety is widely known as the safety-in-numbers effect, as introduced by Jacobsen (2015). Safety in numbers can be explained by the non-linearity of risk (Elvik, 2009). This means that the number of bicycle crashes increases proportionally less than the increase in bicycle volume (Schepers et al., 2014). Thus, cyclists are safer when there are more cyclists on the road (Carlson et al., 2019). Conversely, cyclists are more vulnerable than car users, which may imply that more cyclists lead to a decrease in road safety (Wegman et al., 2012). Several Dutch studies found a large increase in severe bicycle crashes and a small increase in bicycle fatalityies after a modal shift toward more bicycle use, especially due to an increase in single-bicycel crashes (Schepers & Heinen, 2013; Stipdonk & Reurings, 2012; van Wee & Ettema, 2016). In Berlin, Lücken and Wagner (2020) even found a hazard-in-numbers effect after rescaling from a yearly scale to a monthly scale. These studies imply the cause of the abovementioned increasing seriously injured cyclists in the Netherlands.

Although a lot has been done in the area of bicycle volumes and cyclist safety, there are still some gaps in the literature. Firstly, as mentioned above, it is still debatable what the effects of increasing bicycle volumes are on road safety. Secondly, most safety-in-numbers literature aims to understand its effect at an aggregated level, such as a city-wide safety-in-numbers effect (Aldred et al., 2018; Dozza, 2017), or even a country-wide safety-in-numbers effect (Tin Tin et al., 2013). The present study aims to persuade the need of very detailed bicycle and traffic data in order to understand the safety-in-numbers effect on urban network level. In this way, the purpose of this study is to elaborate on why it is valuable to use detailed temporal variation in bicycle volume in order to carry out a proper bicycle safety study. This is especially true for the amount of bicycle traffic in Dutch cities, which fits in the scope of the author's PhD research: '*Road safety for cyclists in Dutch cities*'. The detailed traffic data used in the present study comes from permanent count stations for cyclists in Utrecht, while the hourly traffic data comes from traffic counts at traffic lights. Bicycle crash data is derived from the Dutch national database for traffic crashes: BRON.

In the remainder of this paper three arguments will be discussed about how the temporal variation in bicycle traffic and motorized traffic play a role in explaining biycle crash frequency in Dutch cities. Firstly, variaton in bicycle volume during the day and on different weekdays will be discussed. Secondly, the unique case of the four major cities in the Netherlands will be discussed in terms of bicycle volumes. Lastly, it is showed how hourly variation in bicycle and motorized traffic relates to bicycle crash frequency.

2. HOURLY VARIATION IN BICYCLE VOLUME

In the majority of bicycle road safety research, the annual average daily bicyclists (AADB) and the average annual daily traffic (AADT) are used as measures of exposue, which in most cases are the main predictors for bicycle safety (Gomes et al., 2019; Kamel et al., 2020; Mukoko & Pulugurtha, 2020; Osama & Sayed, 2016; Strauss et al., 2015). These are often present as the daily average bicycle counts or the amount of bicycle kilometres travelled, as well as the daily average counts of motorized traffic, the amount of motor vehicle kilometres travelled or the time being in traffic (Dijkstra, 2011). While this is very helpful in explaining the effect of bicycle volume and traffic volume on road safety for cyclists, it may not clarify the temporal variation in bicycle volume, traffic volume, and eventually bicycle crash risk. Therefore, it is necessary to use detailed bicycle volume data as well as detailed traffic volume data. Broadly speaking, it is assumed that detailed traffic volume data is more widely available and easier to collect compared to bicycle volume data. Therefore, in the remainder of this paper, the focus is on bicycle volume data.

First of all, one reason for why such detailed bicycle volume data is not always used is that this type of data is often unavailable or does not even exist. When this is the case, it is important to collect data in the field or to find other sources which collect such data. One way to overcome this problem is to make use of GPS data.

In the last few years, using GPS data as a source for detailed bicycle volume data became more widespread. Several studies explain how GPS data can be used to estimate bicycle volumes (Jestico et al., 2016; Pogodzinska et al., 2020; Strauss et al., 2015). Sometimes the GPS data is collected by the authors themselves, while in other studies crowd sourced data is used. Sources like the sports tracker app STRAVA are well-known sources for crowd sourced bicycle GPS data. In many cases, this data is then calibrated by using data from permanent bicycle count stations or pneumatic tubes. Another method can be to only use data from count stations and predict bicycle volumes in the surrounding areas by the use of a transport model (Beitel et al., 2017).

The result of collecting detailed bicycle volume data and its comparison to traffic volume data is shown in Figure 1. The figure illustrates an average week on an hourly scale from a permanent bicycle count station in Utrecht in the Netherlands. The pattern of the bicycle volume shows a similar pattern as the traffic volume. Although, it seems that motorized traffic has a larger afternoon peak and during midday the volume does not decrease as much as the bicycle volume. Furthermore, the pattern clearly shows the morning and the afternoon peak hours on the weekdays and an overall peak during the shopping times on the weekenddays. What can be seen in the figure as well is that bicycle volume varies during the day. This is important information because when the goal is to investigate different loads of traffic in a network, such kind of data is very useful. Especially when its effects on road safety for cyclists is studied.

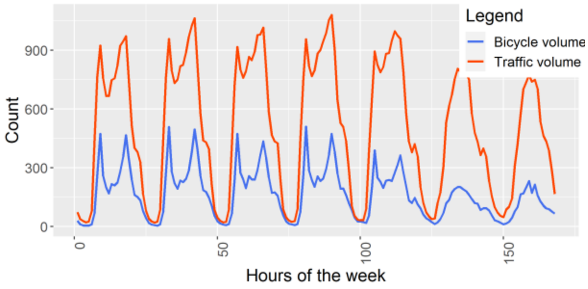


Figure 1: Average hourly bicycle volume and traffic from a permanent count station in the Venuslaan in Utrecht, the Netherlands

3. DUTCH BICYCLE LEVELS

While bicycle levels between different countries in the world may vary, the Dutch situation is something special compared to most other countries. In the most urbanised areas of the Netherlands, the Randstad with its capitals Amsterdam, Rotterdam, The Hague and Utrecht, almost 40% of the trips between 1 and 7 kilometer is traveled by bicycle in 2015 (Jonkeren et al., 2019). This is a large share compared to other modalities. On street level, this may result in a large difference between motorized traffic and bicycle traffic. Figure 2 shows the average hourly bicycle volume compared to the average hourly traffic volume in 2018 for a single permanent count location in Utrecht. The AADB for this location is 16,536 while the AADT is 10,852. It seems that bicycle volume may exceed traffic volume.

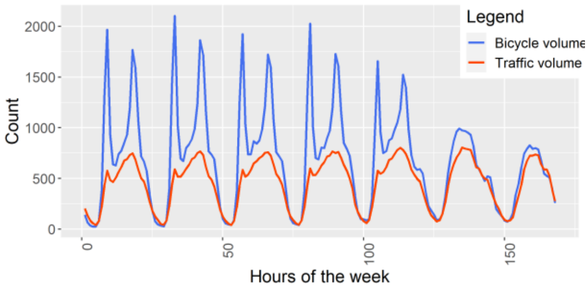


Figure 2: Difference in bicycle volume and traffic volume on street level on the Amsterdamsestraatweg in Utrecht, the Netherlands

The difference between Figure 1 and Figure 2 can be explained by the location of the permanent count station. While the location of the count station from Figure 2 is close to the city centre and Utrecht's main public transport hub, the count station from Figure 1 is located further away from the city centre and lies on an inner-city ring road. The city centre in Utrecht attracts large numbers of cyclists as a destination for cyclists itself, and because it contains the main corridor to the Utrecht Science Park for cyclists. Moreover, the the public transport hub is located next to the city centre and contains the main Central Station for trains, as well as the city its main bus terminal and main tram station. A large share of users of this public transport hub uses cycling as an access or egress mode. On the other hand, an inner-city ring road attracts more motorized traffic compared to cyclists, because these roads are well-connected to the highways and connect different neighborhoods. Therefore, traffic volume levels are higher in Figure 1 and bicycle volume levels are higher in Figure 2.

In order to compare Figure 2 to other European cities, Figure 3 shows the modal split for twenty major European cities. This figure illustrates that only Copenhagen has a larger share of cyclsits than Utrecht and most other European cities are very low in their share of cycists. While it is not hourly data, it may give an idea of how the composition of traffic in these cities looks like. This illustrates why Dutch cities are a special case compared to urban areas from other European countries. Moreover, this may clarify the large difference in bicycle traffic and motorized traffic in Figure 2.

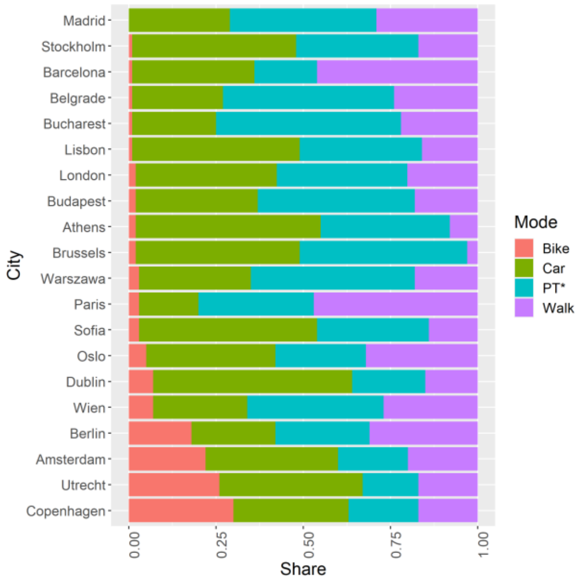


Figure 3: Modal split in percentage of trips in twenty large European cities (EPOMM, 2019). * PT: Public transport

The Netherlands is unique in terms of bicycle traffic and infrastructure. Cycling levels in most other countries are incomparable due to too low bicycle volumes compared to the Dutch bicycle demand. Moreover, the Netherlands has a well-designed bicycle network containing more bicycle facilities than any other country in the world, resulting in an unique cycling environment. Therefore, studying bicycle traffic and its safety in the Netherlands needs a different approach and more detailed bicycle and traffic data .

4. DIFFERENCE IN BICYCLE CRASH NUMBERS

As discussed above, bicycle volumes vary during the day and during different days of the week. How does this apply for bicycle crashes? Do they follow a similar distribution during the week or do they differ? A first look on aggregated hourly bicycle crashes from 2015-2019 in Utrecht creates Figure 4. The figure illustrates how bicycle crashes in Utrecht follow the distribution of cyclists and motorized traffic during the days of the week.

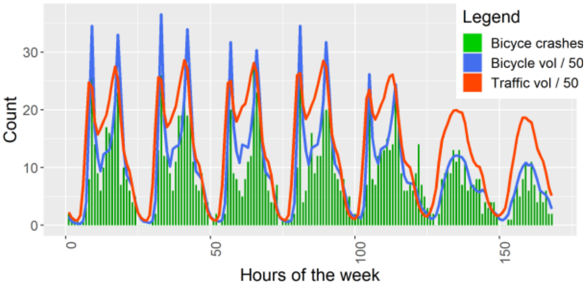


Figure 4: Average hourly bicycle volumes and traffic volumes for seven permanent count stations in Utrecht, including hourly bicycle crashes

According to this figure, it may be indicated that hourly variation in exposure plays a large role in explaining bicycle crashes in Dutch cities. On weekdays, when exposure increases during rush hours, bicycle crashes increase as well. Whereas during off-peak hours and in the evening and nights, crash frequency is much lower. At these times, especially in the evening and at night, exposure has also decreased compared to daytime hours. This is especially true when night times are compared to peak hours.

On weekenddays, a similar pattern is found, although at different times. During shopping hours, both exposure and bicycle crash frequency increase, while in the evening and during night times both decrease. One remarkable issue is the large crash number on Friday night in relation to the low bicycle and traffic volume during these hours. This is the moment where other factors than exposure start to play a role. For this particular case, alcohol may be part of the explanation as Friday nights are often the nights when people go to bars and clubs. Interestingly, similar findings come from a study by Dozza (2017). He found that during weekend nights, crash numbers cannot be explained solely by exposure anymore. However, one drawback of this study is that the author only used bicycle volume data and tries to explain bicycle-motorized vehicle crashes.

Interestingly, Figure 4 also shows that the average bicycle volume of seven permanent count stations exceeds the average traffic volume for the same locations during some time frames. This is in line with what Figure 2 revealed, that cyclists may outpace motorized traffic in Utrecht. This may be of special interest when looking to a safety-in-numbers effect for cyclists in Dutch cities. Therefore, hourly variation in bicycle traffic, motorized traffic and bicycle crashes is an important perspective to investigate road safety for cyclists in Dutch cities.

5. CONCLUSIONS

The present paper tried to argue why hourly variation in bicycle traffic, motorized traffic and bicycle crashes is important in explaining road safety for cyclists in Dutch cities. Three arguments were given. Firstly, detailed bicycle volulme data sheds light on how the amount of cyclists varies during the day and between different days of the week. This may give other results than looking to average annual daily bicycle traffic when trying to explain bicycle road safety. Secondly, compared to other countries, the Netherlands is a special case in terms of bicycle demand. The share of bicyclists in the four major Dutch cities for short trips is higher than for cars. This relates to the difference in hourly bicycle traffic and hourly motorized traffic at street level. In this way, a study about road safety for cyclists in Dutch cities needs a different approach compared to studies using conventional crash frequency models. The temporal aspect may be very important to include and this needs a different perspective. Lastly, it is made clear that for most hours in a week the distribution of bicycle crashes follows the distribution of bicycle volume and traffic volume. Therefore, hourly exposure data may help to grasp the temporal aspect in road safety for cyclists in Dutch cities. Although, other factors may play a role as well.

The results from this study are helpful for transport planners from municipalities. These insights may be an argument to even more try to reduce bicycle crashes, especially during peak hours. One solution could be to focus on disentangeling cycling infrastructure and motorized traffic infrastructure (Schepers, 2013). This leads to separating cyclists from heavy and fast motorized traffic. Another solution may be to decrease maximum speeds of motorized traffic. In the Netherlands, there is an ongoing discussion to decrease all maximum speeds in urban areas to 30 km/h. This speed is seen as a safe speed limit for vulnerable road users (Twisk et al., 2013).

The findings in this paper are useful in further research to road safety in Dutch cities, the author's PhD topic. In this project, the presented data from the present study will be input for a crash frequency model where the temporal aspect is the main interest. Instead of using daily average traffic and bicycle counts for specific road segments, hourly estimates are the guiding principle. For every road link, it is known how, for an average week, the hourly variation in cyclists and traffic looks like.

Furthermore, for the bicycle crashes it is also known in which hour and on which day of the week they occured. This information will be aggregated. In this way, the whole road network becomes one road segment with an aggregated hourly bicycle count, traffic count and bicycle crash count. The resulting data frame therefore has 168 rows, being every hour in a week (Monday 00:00-01:00 till Sunday 23:00-00:00). This will be the starting point for the modeling in the author's PhD research, where a Generalized Linear Model will be used.

In conclusion, in order to find out if more cyclists on the road leads to decreasing road safety for cyclists, it is necessary to take hourly variation in bicycle volume and traffic volume into account. Especially for cases where bicycle use exceeds car use, like in Dutch cities. The analyzes in this study make it necessary to examine correlations between hourly bicycle volume, traffic volume and bicycle crashes in more detail. This will derive statements concerning bicycle road safety in urban areas, which is part of the author's PhD thesis.

REFERENCES

[1] Aldred, R., Goodman, A., Gulliver, J., & Woodcock, J. (2018). Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. *Accident Analysis and Prevention*, *117*, 75-84, https://doi.org/10.1016/j.aap.2018.03.003

[2] Beitel, D., McNee, S., & Miranda-Moreno, L. F. (2017). Quality measure of short-duration bicycle counts. In *Transportation Research Record* (Vol. 2644, pp. 64-71).

[3] BOVAG. (2019). *Zonnig 2018 stuwt omzet fietsbranche naar record*. https://mijn.bovag.nl/actueel/nieuws/2019/maart/zonnig-2018-stuwt-omzet-fietsbranche-naar-record

[4] Carlson, K., Ermagun, A., Murphy, B., Owen, A., & Levinson, D. (2019). Safety in Numbers for Bicyclists at Urban Intersections. In *Transportation Research Record* (Vol. 2673, pp. 677-684).

[5] CBS. (2019, 18 April 2019). *Overledenen, doden door verkeersongeval in Nederland, wijze van deelname*. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/71936ned/table?ts=1574759336784

[6] de Jong, A., & Daalhuizen, F. (2014). *De Nederlandse bevolking in beeld: verlden, heden, toekomst*. The Hague: Planbureau voor de Leefomgeving.

[7] Dijkstra, A. (2011). *En route to safer roads: How road structure and road classification can affect road safety* [Doctoral thesis, University of Twente]. Leidschendam: SWOV.

[8] Dozza, M. (2017). Crash risk: How cycling flow can help explain crash data. *Accident Analysis & Prevention*, *105*, 21-29, https://doi.org/http://dx.doi.org/10.1016/j.aap.2016.04.033

[9] Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis & Prevention*, *41*(4), 849-855, https://doi.org/https://doi.org/10.1016/j.aap.2009.04.009

[10] EPOMM. (2019). *TEMS - The EPOMM Modal Split Tool*. http://tems.epomm.eu/compare_cities.phtml

[11] Gomes, M. M., Pirdavani, A., Brijs, T., & Souza Pitombo, C. (2019). Assessing the impacts of enriched information on crash prediction performance. *Accident Analysis and Prevention*, *122*, 162-171, https://doi.org/10.1016/j.aap.2018.10.004

[12] Harms, L., Bertolini, L., & te Brömmelstroet, M. (2014). Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends. *Journal of Transport and Health*, *1*(4), 232-242, https://doi.org/10.1016/j.jth.2014.09.012

[13] Jacobsen, P. L. (2015). Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention*, *21*(4), 271-275,

https://doi.org/10.1136/ip.9.3.205rep

[14] Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, *52*, 90-97, https://doi.org/https://doi.org/10.1016/j.jtrangeo.2016.03.006

[15] Jonkeren, O., Wust, H., & de Haas, M. (2019). *Mobiliteit in stedelijk Nederland*. The Hague: Kennisinstituut voor Mobiliteitsbeleid.

[16] Kamel, M. B., Sayed, T., & Bigazzi, A. (2020). A composite zonal index for biking attractiveness and safety. *Accident Analysis and Prevention*, *137*, https://doi.org/10.1016/j.aap.2020.105439

[17] Lücken, L., & Wagner, P. (2020). On the relation between bicycle volumes and individual risks for bicyclists in Berlin. *Journal of Transportation Safety & Security*, *12*(1), 210-223, https://doi.org/10.1080/19439962.2019.1597003

[18] Mukoko, K. K., & Pulugurtha, S. S. (2020). Examining the influence of network, land use, and demographic characteristics to estimate the number of bicycle-vehicle crashes on urban roads. *IATSS Research*, *44*(1), 8-16, https://doi.org/10.1016/j.iatssr.2019.04.001

[19] OECD. (2018). *Road Safety Annual Report 2018: The Netherlands*. Paris: International Transport Forum/OECD.

[20] Osama, A., & Sayed, T. (2016). Evaluating the impact of bike network indicators on cyclist safety using macro-level collision prediction models. *Accident Analysis & Prevention*, *97*, 28-37, https://doi.org/https://doi.org/10.1016/j.aap.2016.08.010

[21] Pogodzinska, S., Kiec, M., & D'Agostino, C. (2020). Bicycle Traffic Volume Estimation Based on GPS Data. *Transportation Research Procedia*, *45*, 874-881, https://doi.org/https://doi.org/10.1016/j.trpro.2020.02.081

[22] Schepers, P. (2013). *A safer road environment for cyclists* [Doctotal Thesis, Delft University of Technology]. Leidschendam: SWOV.

[23] Schepers, P., Hagenzieker, M. P., Methorst, R., van Wee, B., & Wegman, F. (2014). A conceptual framework for road safety and mobility applied to cycling safety. *Accident Analysis and Prevention*, *62*, 331-340, https://doi.org/10.1016/j.aap.2013.03.032

[24] Schepers, P., & Heinen, E. (2013). How does a modal shift from short car trips to cycling affect road safety? *Accident Analysis & Prevention*, *50*, 1118-1127, https://doi.org/http://dx.doi.org/10.1016/j.aap.2012.09.004

[25] Stipdonk, H., & Reurings, M. (2012). The Effect on Road Safety of a Modal Shift From Car to Bicycle. *Traffic Injury Prevention*, *13*(4), 412-421, https://doi.org/10.1080/15389588.2012.660661

[26] Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2015). Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. *Accident Analysis and Prevention*, *83*, 132-142, https://doi.org/10.1016/j.aap.2015.07.014

[27] SWOV. (2018). Ernstig verkeersgewonden in Nederland. In. The Hague: SWOV.

[28] Tin Tin, S., Woodward, A., & Ameratunga, S. (2013). Incidence, risk, and protective factors of bicycle crashes: Findings from a prospective cohort study in New Zealand. *Preventive Medicine*, *57*(3), 152-161, https://doi.org/https://doi.org/10.1016/j.ypmed.2013.05.001

[29] Twisk, D. A. M., Vlakveld, W. P., Dijkstra, A., Reurings, M., & Wijnen, W. (2013). *From bicycle crashes to measures*. Leidschendam: SWOV.

[30] van Wee, B., & Ettema, D. (2016). Travel behaviour and health: A conceptual model and research agenda. *Journal of Transport & Health*, *3*(3), 240-248, https://doi.org/https://doi.org/10.1016/j.jth.2016.07.003

[31] Wegman, F., Zhang, F., & Dijkstra, A. (2012). How to make more cycling good for road safety? *Accident Analysis & Prevention*, *44*(1), 19-29, https://doi.org/https://doi.org/10.1016/j.aap.2010.11.010

[32] Weijermars, W., de Goede, M., Goldenbeld, C., Decae, R. J., Schepers, P., Wijlhuizen, G. J., Mons, C., Hoekstra, A. T. G., & Dijkstra, A. (2019). *Monitor verkeersveiligheid 2019 - Achtergrondinfromatie en onderzoeksverantwoording*. The Hague: SWOV.

APPLICATION OF AUTOMATION OF PUBLIC ROAD WORKS VEHICLES IN UNSTRUCTURED ENVIRONMENTS

Alfredo Valle Barrio, Phd Student in Artificial Intelligence
Alberto Cruz Ruiz, Master in Artificial Intelligence
Victor San Telesforo García, Degree student in Computer Science
José Eugenio Naranjo Hernandez, Professor in the UPM
Felipe Jiménez Alonso, Full Professor in the UPM
INSIA, UPM, Campus Sur UPM
Carretera de Valencia km 7, 28031, Madrid, España
alfredo.valle@upm.es

ABSTRACT

Today we are used to hear news related to vehicle automation and its integration on roads in different parts of the world. These news items mainly concern the automation of highly sensorised cars for conventional roads, which must be perfectly signposted and maintained. However, there is still a long way to reach this point in commercial vehicle automation, especially in their use in unstructured environments.

The proposal reflected in this paper is the automation of an industrial vehicle, specifically a Volvo A25, to use it in unstructured environments such as the excavation of a tunnel. The design is focused on carrying out automatic debris removal tasks without need a driver or supervisor. To this goal, the essential vehicle functions have been automated, such as speed control, steering, gear change, and tilting the dump box, which allows the system to have total control over the vehicle's capabilities.

In addition to this, it has been equipped with an inertial-differential GPS that provides it with the capacity of sub metric self-location. Thanks to this GPS, the vehicle has the capacity to circulate autonomously on a pre-recorded map, maintaining its trajectory.

In the tests carried out on the INSIA track, the maneuvering capacity that has been implemented in the vehicle has been verified, obtaining as a result that it is capable of driving for an unlimited time on a pre-recorded map, carrying out the controlled braking maneuvers and resuming the route, either due to an event defined or due to the intention of an operator to control the vehicle remotely.

Keywords:
Connected and automated vehicles, autonomous vehicles, Intelligent transportation systems, Robot Operating System (ROS)

Research domain:
The work included in this paper is closely related to two of the seminar's research topics. These are firstly Intelligent transport systems ITS & Traffic management. Secondly, Transport civil and road engineering.

1. INTRODUCTION

The application of autonomous vehicles is expanding its frontiers beyond those seen in the news, passenger transport. There are many areas where they can be applied, such as agriculture, military scenarios, or public works (Jiménez, 2018). The use of autonomous machinery in such environments was strictly restricted to specific repetitive tasks requiring machine movement. This was due to the difficulty and danger of drilling walls for explosive charges or spraying concrete.

Gradually, we are starting to hear from commercial vehicle construction companies that are more frequently incorporating driver assistance systems or low-level automation. These companies' problem is that trucks or equipment of this type tend to have a relatively long service life, making upgrading to new automated vehicles difficult.

Companies such as Aurora, Caterpillar, and Virgin Hyperloop have already shown their intention to introduce industrial vehicles in unstructured environments such as mines or excavations. Some of the drawbacks of these private companies' proposals are economical, as they sell a closed system that is not very flexible and requires a significant deployment of communications infrastructure.

Due to the complexity and cost of developing this type of vehicle and technology, more and more projects are attempting to take automation outside the field of passenger transport. One of these is the TUNNELAD project, which includes the development of this article. This is a pilot project for the development of autonomous industrial vehicles for unstructured environments.

This article wants to show the possibility of automating such a commercial vehicle to a high level of automation that requires minimal interaction with personnel (Jiménez et al., 2020). This allows the vehicles to be reused for a long time while increasing safety (Pereira de Oliveira et al., 2020) in the hazardous environment in which they work. Economically, it reduces maintenance costs, personnel costs, and potential accidents. It is also highlighting noting that this type of installation does not require infrastructure deployment as all the equipment is embedded in the vehicle, which is a requirement in many public works projects.

This article will detail the architecture, both hardware and software, developed for this project (Naranjo et al., 2020) and the result of the automation and guidance tests carried out on the south campus of the Polytechnic University of Madrid at the INSIA.

2. PLATFORM AND TEST TRACK

The work developed in this paper has been implemented on an industrial vehicle, a Volvo A-25 truck, as shown in Figure 1. The automation of this vehicle has been a challenge due to its large cargo, size, and, precisely, the way it turns, since, in this case, the wheels do not behave like those of a typical vehicle. The wheels move in solidarity with the head to turn the vehicle, so it pivots on its axis.

Figure 1: Volvo A-25, vehicle to automate

This project's development and testing have been carried out at the INSIA facilities, which belong to the Polytechnic University of Madrid. These facilities are located in the south of Madrid, Spain, at the junction of the M.40 road with the A-3 motorway. Figure 2 shows an aerial image of the INSIA on whose test track can be distinguished in the image next to the M-40 road, where the tests have shown at the end of this article has been carried out.

Another of the problems encountered when working with such a large vehicle is that the track is too small, as it has a diameter of just 40 meters at its widest part, which does not leave much room for maneuvering.

Although the INSIA test track is not designed to accommodate and maneuver such large vehicles, the availability of this track on the outskirts of Madrid has allowed to carry out many tests, especially those of vehicle guidance by GPS, without having the problem that the urban canyon situation poses for the GPS. The possible implementation of this project will be carried out in open mines for the most part, so GPS is essential.

Figure 2: Bird view of INSIA

The project's natural development requires to test the system in a more realistic environment, so the vehicle's transport and testing at the project contractor's facilities in Noblejas, Toledo, Spain, has already been planned.

More demanding tests will be carried out at these facilities. They will include autonomous driving, as shown in this article, loading/unloading the vehicle, taking control by an operator during vehicle operation, and even simulating entry into a tunnel without GPS coverage, where the vehicle will have to be guided by Lidar.

In this new facility, the vehicle will face several challenges. On the one hand, speed control, having to overcome steep ascents and descents, whether the vehicle is loaded with up to 25 tons of earth. The difficulty of keeping such a large vehicle in a confined space at a considerable speed of up to 40 km per hour. All this during a 2 km journey where interaction with an operator will be minimal.

3. AUTOMATION AND HARDWARE ARCHITECTURE

For this project, complete automation of the vehicle has been carried out, consisting of the following sections:

- Steering
- Speed
  - Throttle
  - Brake
- Gears
- Dump box

The following sections will show in detail how each actuator's automation required for the vehicle's complete control has been carried out.

3.1. Steering

To control the vehicle's steering, (Jiménez et al., 2013) (Alonso et al., 2014) a device has been attached to the steering column.

On the upper side of Figure 3 shows the patented device that will perform the steering wheel turning functions. This has an electromagnet that will act as a magnetic clutch to take control of the steering. The lower side of the image shows how the device is attached to the steering rod.

A signal controls this device through a CAN BUS that will indicate the angle at which it should turn. This angle is calculated based on the reading of a permanent turn sensor attached to the steering rod, because the reading of this position is not available through the vehicle's CAN. The magnetic clutch is controlled by a digital signal obtained from the engine itself, activated or deactivated via a CAN BUS signal.

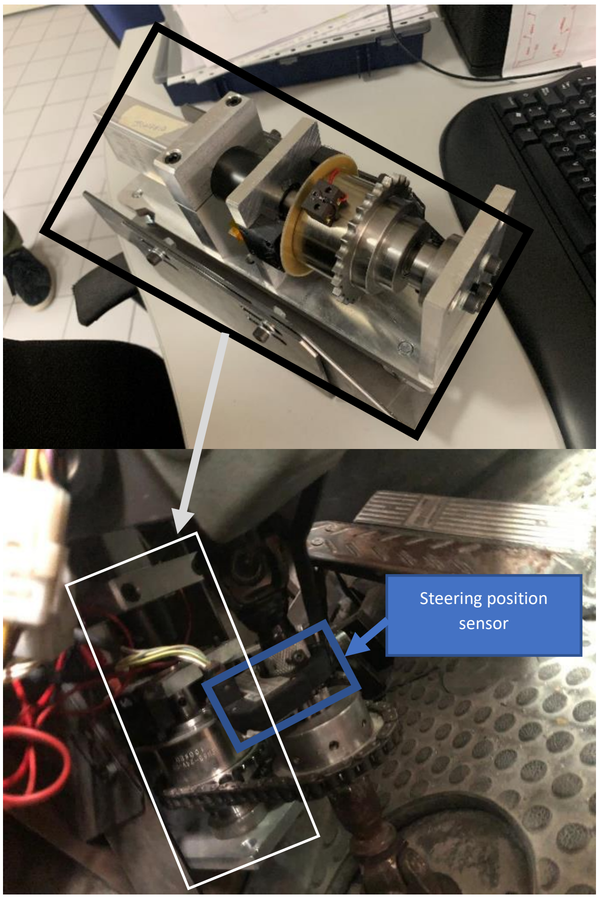


Figure 3: Detail of the steering automation assembly

3.2. Speed

The speed control of the vehicle is done by controlling two actuators. On the one hand, the brake control and, on the other hand, the throttle control. The automation of these actuators is described below.

3.2.1. Brake

The brake control is pretty similar to the steering wheel control. The same motor is used in the steering control device. The brake pedal is actuated mechanically using a chain. In contrast to the steering control device, the brake pedal's pressure does not require measuring the brake pedal position. This is because its actuation is controlled dynamically by a PID from the speed value.

Figure 4: Detail of brake assembly

3.2.2. Throttle

In this vehicle's case, the accelerator pedal is a potentiometer that generates a voltage proportional to the pedal position. To control the throttle and return control to an operator when necessary, a set of relays selects the original throttle signal's sending or the signal provided by an analogue board.

Figure 5 shows the INSIA designed board, which has two inputs corresponding to the throttle potentiometer signal and the analogue board signal, and an analogue output to be sent to the vehicle ECU.

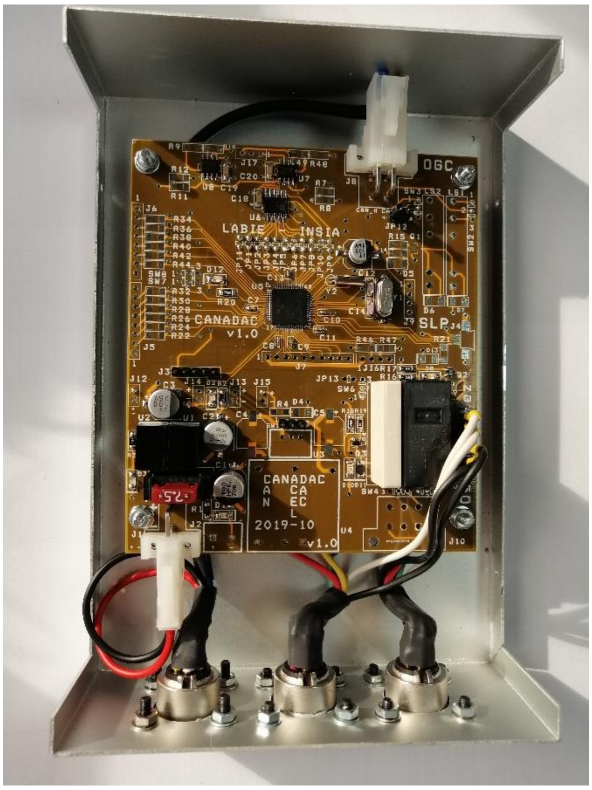


Figure 5: Throttle signal bridging board

3.3. Gears

Since the vehicle will have to perform complex manoeuvres, including changing gears to make tight turns or load/unload the truck, the gears' control has been automated.

In the case of this vehicle, gear shifting is effortless to automate. Although the gearshift lever may look like a too mechanical device, it is a digital signal selector that tells the control unit which gear to select.

Figure 6 shows the relay board's control of the gears with a tester and the final installation inside a watertight box with standard connections. The control of these relays is carried out employing CAN BUS communication that selects the relays' state of excitation.

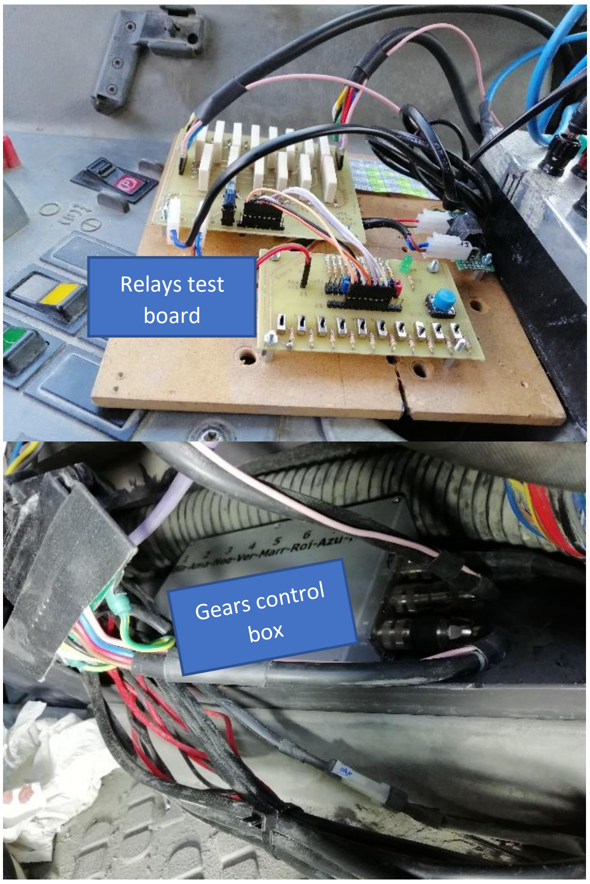


Figure 6: Installation of the gear control relays

3.4. Dump box

Dump box control has been made with a stepper motor, coupled to a device consisting of a worm screw and two metallic guides to avoid oscillations. The worm screw has an associated anchorage to the dump box control lever, making the dump box move in conjunction with the motor's rotation. An end-of-stroke sensor has also been installed to calibrate the bowl positions.

The assembly of this device can be seen in Figure 7. Observing the image, it can be seen that the anchor that joins the control lever with the worm screw is not welded to the lever but grips it in such a way that it can be unscrewed to carry out manual control of the box.

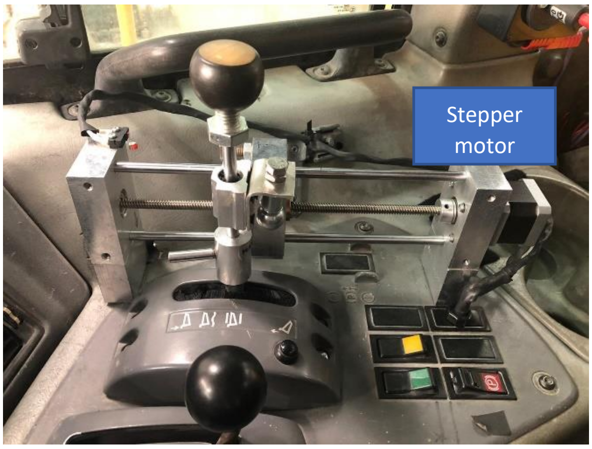


Figure 7: Dump box tooling

3.5. Control Board

To unify the control of all the devices mentioned above, they are connected via a single CAN BUS to the control board, which can be seen in Figure 8: Control suitcase.

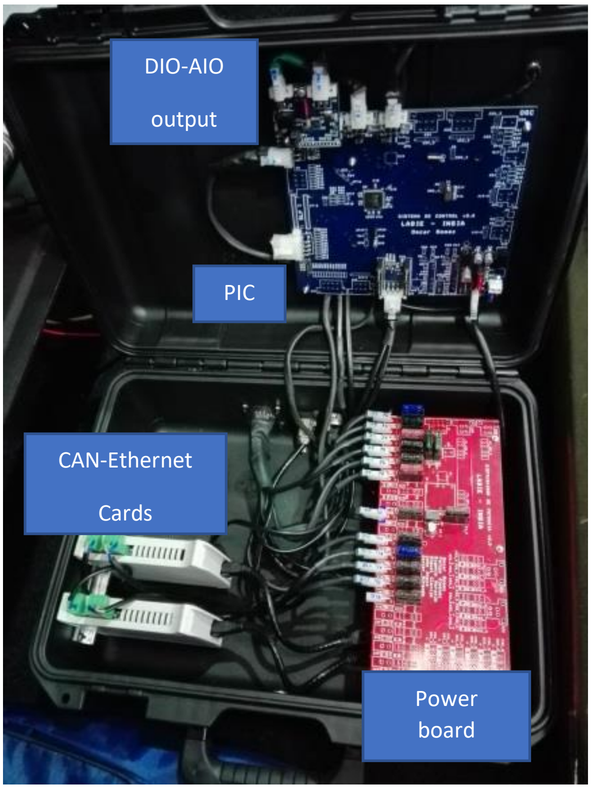


Figure 8: Control suitcase

This suitcase consists of one hand of 2 Ethernet-CAN cards. One for collecting data from the vehicle's CAN, precisely the speed, and the other for transmitting data to the devices mentioned above.

In the lower right part of the image, we can see an electronic card that functions as a power distributor. From this card, the power supplies of all the motors and the computers are controlled.

The electronic card anchored to the case's cover is a card designed at INSIA, responsible for monitoring and managing the CAN BUS control and guaranteeing a possible emergency always stop.

Finally, a fanless computer is installed on top of the suitcase, where the system's software control is installed. This oversees the necessary commands to the actuators and, in turn, executes the high-level systems such as waypoint and later the obstacle detection by Lidar. This computer is connected via Ethernet to a switch that manages the connections between the two Ethernet-CAN cards and the computer.

4. SOFTWARE ARCHITECTURE

The entire software architecture has been implemented in the ROS 2.0 framework, in its "eloquent" version. This framework has been used as the basis for the entire development because modular, scalable, and standard implementation develop will be used in all our vehicles.

ROS provides several operational advantages because of its modular functionality. On the one hand, it provides the usual features of an operating system, such as hardware abstraction and message passing between processes. Simultaneously, it abstracts from the computer's operating system, which is especially important since several computers, possibly with different operating systems, will be used to develop the project.

On the other hand, it allows having several people develop code for the exact vehicle, specializing in a specific model and only worrying about maintaining consistency in the name of the topics.

In the case of this project, we will use ROS on the Ubuntu 18.04 operating system. We have chosen this operating system because it is the one that works best with ROS and at the same time allows better management of resources and networks. It is also a much lighter operating system than Windows, which allows to use less power and consequently cheaper computers.

In the following, the implemented modules will be developed, and the data flow between them. The system consists of 5 ROS packages and seven nodes divided as follows:

- CAN Package
  - Control Node
- Decision Package
  - Decision Node
- Sensors Package

  - GPS Node
- Waypoints Package
  - Waypoints Node
  - Recorder Node
- Teleoperation Package
  - Keyboard Node
  - Tablet Node

Figure 9 shows the general system scheme and its data flow.

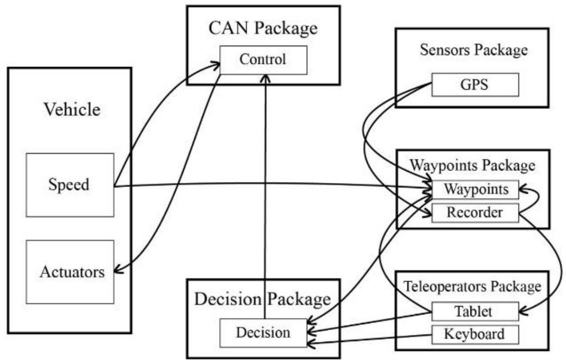


Figure 9: Overview of software architecture

4.1. CAN Package

The CAN package contains only one node, which is the control node. It must open TCP connections to the Ethernet-CAN cards to transmit messages or receive messages from the vehicle. It is also responsible for converting the speed and direction commands to values that the actuators can understand.

The low-level control performs, as the following:

To send steering commands, it takes the encoder's value coupled to the steering rod. It generates a CAN packet encapsulated in a TCP packet, which contains the instructions for a relative rotation of the motor concerning the current position. Enabling/disabling the steering control is done by the same CAN packets that enable/disable the digital signals controlling the electromagnet coupled to the steering link.

To send speed commands, it has a PID controller, which controls the throttle and brake. In the throttle case, a packet is sent to the analogue card that controls the signal sent to the ECU. In the brake case, a packet similar to that of the steering motor is sent, but in this case, in absolute format. In the brake, disabling the speed control means moving the brake to the zero position, not pressed. In the throttle case, it is rerouting the signal from the accelerator pedal back to the ECU.

The gear commands are sent via a CAN packet to the board controlling the relay board that selects the gear engaged. To enable/disable the gear control, a CAN packet is sent to tell the gear control board to use the reader on the lever and not on the relay board.

Finally, in the case of the tank, it sends a CAN packet to the arduino board that controls the stepper motor's movement coupled to the tank movement lever. Unlike the previous devices, the dump box cannot be enabled and disabled by software. The lever is mechanically associated with the motor, so if you wanted to disable the vat control, you would have to disassemble it manually.

The steering, throttle, and brake commands are controlled at 20 Hz, while the gears and dump box operate at 10 Hz. The maximum frequency at which CAN packets are sent from all devices together has been calculated. This does not coincide with the frequencies mentioned above because these frequencies only control the devices, not the number of packets required to carry out this control. The frequency at which packets should be sent to the Ethernet-CAN card has been defined as 300 Hz.

4.2. Decision Package

The decision package contains only the decision node. It has two main functions.

Firstly, it is responsible for collecting the speed, direction, and gear setpoints from those nodes that publish these parameters. Secondly, it is responsible for sending a single speed/steering/gear setpoint to the vehicle's low-level control at a constant frequency through a predefined priority system.

This frequency has been empirically adjusted on the vehicle in such a way as to allow rapid action on the vehicle control without saturating the control line components (ethernet-can cards, motors, and low-level computer).

4.3. Sensors Package

Currently, only the GPS node is implemented in this package, including sensors such as Lidar, Mobileye camera, or stereo camera later.

The GPS node connects via a TCP client to even differential sub-metric GPS, which continuously provides us with the vehicle's position at a configurable rate of between 10 and 20 Hz. Besides, it provides various signal quality parameters to identify whether we are entering a tunnel or safe to continue autonomous driving.

4.4. Waypoints Package

The waypoints package is composed of 2 nodes, the recorder node, and the waypoint node.

4.4.1. Recorder Node

The recorder node has as input the GPS position of the vehicle and a processing trigger flag.

The node collects all messages received from the GPS node and checks if the received position is within a minimum distance from the last saved position. This distance is manually defined in the node configuration file.

Once the trigger flag is enabled, it continuously saves the vehicle's GPS position and its speed in a text file. The file follows its own structure defined in the code, although it is easily configurable.

Finally, the recording has two modes of operation. Static speed recording means that the recording will record a fixed speed in the entire route's file. Recording with dynamic speed means that the map will be recorded with the same speed as it is travelled during the recording. It is respecting the margins defined in the configuration file.

4.5. Waypoints Node

The Waypoints node is in charge of loading and replaying the route recorded by the recorder node. It also performs the path tracking calculation to keep the vehicle on the previously recorded path.

The operating cycle of this node is the following:

Firstly, the selected map is loaded using a topic. In case that a map has not been selected, a warning is emitted, and a default map is loaded.

Next, a self-localization on the map is carried out. To do this, the closest point on the map to the current position is searched for. Two situations can occur at this point.

If the Waypoints has just been started and there is no previous position of the vehicle to calculate its trajectory, the closest point on the map to the vehicle's current position will be searched. If we do have information of the trajectory followed by the vehicle, the nearest point on the map will be searched within the following n points from the last nearest point found. If none of these points meets the validation range as the closest point, the closest point on the map will be chosen, even if it does not meet the validation range.

The next step is to calculate the target point to which the vehicle is to be directed. To do this, the target distance or look ahead is calculated, which is defined by the following formula:

$$look_{ahead} = (m * current_{speed}) + n \quad (1)$$

Where m and n are manually defined parameters in the configuration file and change based on the vehicle. Moreover, current_speed is the current vehicle speed collected from the BUS CAN.

Once the target distance is calculated, those points between the vehicle's current position and this one are collected, always following the trajectory of the vehicle. From these points, a Bézier curve is inferred, which will provide us with the target point at which we have to focus the vehicle. With this Bézier curve, we get the vehicle to make the turns more smoothly and continuously.

To finish calculating the target point, this is obtained by segmenting the Bézier curve obtained in 4 segments and defining the target point as the third point of the curve.

The vehicle trajectory's error angle with respect to the target point is calculated to finalize the path tracking. This angle is obtained by trigonometric calculations corresponding to the following formula:

$$\alpha = asin\left(\frac{distance\_to\_target\_point}{distance(target\_point, current\_point)}\right) \quad (2)$$

In addition, the Waypoint node has the task of generating the acceleration ramp at the start of the journey and the deceleration ramp at the end of the journey.

4.6. Tele-Operation Package

The teleoperation package contains two nodes. On the one hand, the tablet node allows the vehicle to be controlled from an external application not implemented in Ros, hosted on a Windows tablet. On the other hand, a keyboard node allows us to test the systems' operation.

4.6.1. Tablet Node

The tablet link node is used to open a series of UDP connections to an external application, not implemented in ROS, hosted on a Windows 10 tablet.

This node allows high-level control of the vehicle's operation. Its capabilities include:

- Enable/disable autonomous steering wheel control.
- Enable/disable autonomous speed control
- Emergency braking
- Remote control of speed and steering wheel rotation
- Select the route to follow
- Activate autonomous guidance
- Make new map recordings

Once the system is launched on the low-level computer, the Tablet has total vehicle control, and it is a vehicle information interface. The connection with the ROS node is made through a wifi access point installed in the vehicle so that the Tablet can connect to any automated vehicle within its wifi signal range.

There are two scenarios if wifi is disconnected. In case that the vehicle is working in autonomous mode, nothing happens. The vehicle will continue driving until the end of the route and it will stop automatically there of if it detects any obstacle in the route. In case that the wifi connection is broken during while the vehicle is manually controlled, through the tablet, it will stop until the connection is fixed.

In dynamic tests carried out in Noblejas, it has been possible to control the vehicle at around 50 m if there is direct visibility between the vehicle and the Tablet.

4.6.2. Keyboard Node

Keyboard node contains a minimal implementation of the controls necessary to operate the vehicle. It allows to control the speed, the steering wheel rotation, the control of the gears, and the dump box's operation with the keyboard arrows and some alphanumeric keys.

5. TEST AND RESULTS

In this section, the results of the tests carried out at INSIA will be shown. These tests have allowed us to check the system's good functioning during the initial development phase, and later we have used them to check the system's accuracy and performance.

5.1. Test track

The route followed for the tests reflected in this paper is as shown in the Figure 10.

Figure 10: Recorded route

This route allows has been used to check the operation and accuracy of:

- Speed control
- Steering control
- Gear shift

The speed control should behave as follows. It must accelerate progressively at the start, maintain a constant speed throughout the journey, and brake gently at the end of the journey.

The steering control is judged on its ability to follow the line marked by the route. It includes two sharp left turns, but the real challenge is to follow straight lines with little margin for lateral error.

Finally, the gear changing system demonstrates how it works: at the start of the journey, the vehicle starts in neutral gear, automatically shifts into direct gear, and at the end of the journey, it must be shifted back into neutral gear.

The dump box operation is not tested because it is always carried out at a standstill and under an operator's supervision outside the vehicle.

5.2. Results

The results obtained are shown below, which show the following route superimposed on the recorded route, the speed profile followed to check the correct functioning of the throttle and brake control and finally the steering wheel turning profile.

The following route is shown in Figure 11. The vehicle starts at a different point from the beginning of the route, it calculates its trajectory to return to the route and follows it until the end of the route where it stops a few meters before the end point.

The vehicle makes two left turns, which, being very tight for the size of it, it tends to take them a little on the inside, although within the permitted margins. In the image is a change of colour in the route followed showing the change of speed.

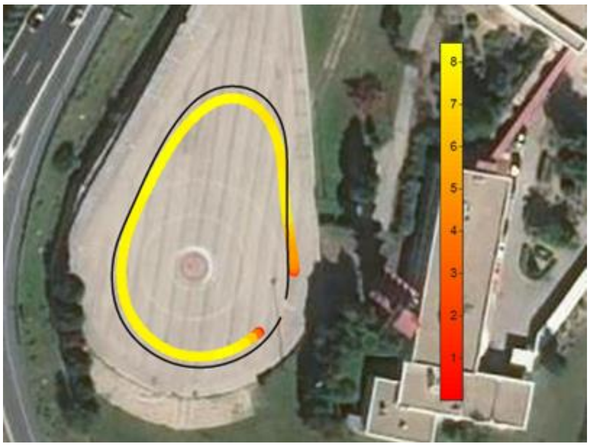


Figure 11: Followed route

Figure 12 shows the speed setpoint together with the actual speed of the vehicle. It can be seen that during the first few seconds it takes a little while to react, this is due to the fact that during this time the vehicle is changing gear, so it keeps its speed at zero until it does so.

It then makes a speed ramp up to 8 km per hour, which is the target speed for this route. At the end we can see how the vehicle slows down. The speed setpoint decreases sharply from 8 km/h to 0 km/h, because the jump is very small but the vehicle brakes smoothly, thanks to the control of the brake motor.

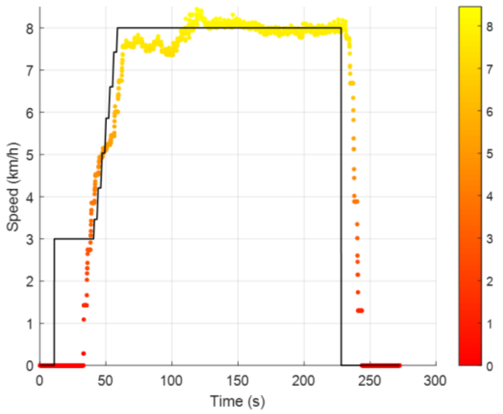


Figure 12: Speed profile

Finally, Figure 13 Figure 13: Steering profileshows the steering profile. This shows how at the beginning of the route it performs a little rougher manoeuvres due to the lack of GPS data, but during the route both the setpoint and the actual steering wheel angle move smoothly and very evenly.

At the end of the trip there is a significant disturbance in the speed control, which is due to having to calculate the target point extremely close and having the vehicle at a very low speed the GPS error increases. Even so, although the setpoint tends to shoot up, the steering wheel remains within a small range.

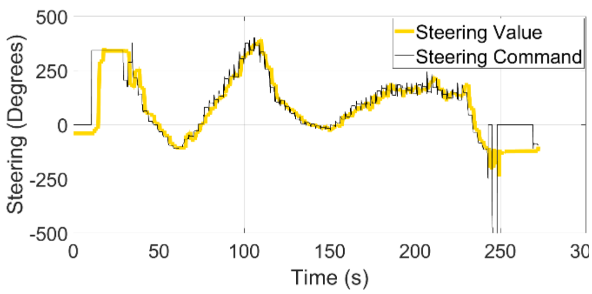


Figure 13: Steering profile

6. CONCLUSIONS

This paper has presented the automation, software architecture and INSIA testing and results of an industrial vehicle for unstructured environments. This vehicle attempts to take a step forward in the development of autonomous and driver assistance systems for industrial environments. It is ahead of the development of industrial vehicle production companies that are starting to market vehicles with driver assistance systems.

The development of this project has focused on the automation of this vehicle with a view to its future application in mines and unstructured environments with highly repetitive work. The need for continuous transport of material along a defined route is common to many civil engineering works, so its application would be immediate.

With this development, the aim is to be able to take advantage of vehicles, so that after a process of automation, they are capable of carrying out this task, reducing costs, maintenance and increasing safety.

The INSIA tests reflected in this article show the system's capacity to perform the task for which it has been developed. Apart from the visualisation of the route followed, data from the steering wheel and speed controllers have been included, demonstrating its good response even in a vehicle of these characteristics.

The project is now fully developed and ready for its planned field tests at the contractor's facilities in Noblejas, Toledo, Spain. The final validation tests will be carried out there and the vehicle's capabilities will be extended with a Lidar sensor.

Aknowledgement

This research has been partially funded by the Spanish Ministerio de Ciencia, Innovación y Universidades (TUNNELAD porject, reference RTC-2017- 6382-4), Ministerio de Ciencia e Innovación (CCAD project: reference PID2019-104793RB-C33) and Comunidad de Madrid (SEGVAUTO 4.0: reference S2018/EMT-4362)

REFERENCES

[1] Alonso, F. J., Hernández, J. E. N., & Casado, Ó. G. (2014). Dispositivo universal para el control automatico de la direccion de un vehiculo. *DYNA*, *89*(4), 398–404.

[2] Jiménez, F. (2018). Aplicaciones especiales de la conducción autónoma. *Revista de Obras Públicas: Organo Profesional de Los Ingenieros de Caminos, Canales y Puertos 3604*, 76--82. https://dialnet.unirioja.es/servlet/articulo?codigo=6765303

[3] Jiménez, F., Martín, M., Anguera, A., Gómez, O., Sánchez, S., Cruz, A., Valle, A., & Naranjo, J. E. (2020). Operación autónoma coordinada de vehículos de obras públicas en entornos complejos. *ITS España*.

[4] Jiménez, F., Naranjo, J. E., González, M., & Gómez, O. (2013). Equipo para controlar automáticamente la dirección de un vehículo. Patente ES2516568.

[5] Naranjo, J. E., Jimenez, F., Anguita, M., & Rivera, J. L. (2020). Automation Kit for Dual-Mode Military Unmanned Ground Vehicle for Surveillance Missions. *IEEE Intelligent Transportation Systems Magazine*, *12*(4), 125–137. https://doi.org/10.1109/MITS.2018.2880274

[6] Pereira de Oliveira, L., Jiménez Alonso, F., Vieira da Silva, M. A., Tostes de Gomes Garcia, B., & Messias Lopes, D. M. (2020). Analysis of the Influence of Training and Feedback Based on Event Data Recorder Information to Improve Safety, Operational and Economic Performance of Road Freight Transport in Brazil. In *Sustainability* (Vol. 12, Issue 19). https://doi.org/10.3390/su12198139

SUITABILITY OF HUMAN MONITORED AUTONOMOUS LOCKERS FOR URBAN DELIVERIES: A COMPARISON WITH THE CURRENT SCENARIO THROUGH FIELD OBSERVATIONS

Kartik Varma
Université Gustave Eiffel, Laboratoire SPLOTT
Groupe RENAULT, Département Innovation et Recherche

François Combes
Université Gustave Eiffel, Laboratoire SPLOTT

Pierre Eykerman
Groupe RENAULT, Département Innovation et Recherche

ABSTRACT

The aim of this paper is twofold. Firstly, it aims to deepen the understanding of the process of urban deliveries through lessons from field observations of the current scenario, and secondly, to use this knowledge to envision and examine a scenario for urban deliveries involving autonomous vehicles, remotely monitored by a human operator, from both a practical and financial perspective. In the introduction, urban logistics, its importance, and trends are briefly described. Questions of desirability of the use of autonomous vehicles, from the perspective of both customers and logistic suppliers, are inspected. Through field observations areas of crucial human intervention are identified. A scenario accomplishing similar delivery characteristics, but utilizing autonomous vehicles instead is developed. Questions and calculations of practicality and cost along with and recommendations for future research are discussed.

Keywords:
Autonomous Vehicles, Urban Logistics, Field Observations

1. INTRODUCTION – AECS VEHICLES AND URBAN LOGISTICS

The commonplace ‘car’ may be at the cusp of change; actors across the automobile industry are aware that the next generation of vehicles – due to a combination of factors – might make the act of driving a ‘leisure’ activity rather than a necessity. This may disrupt both the automobile and mobility industry. Billions of dollars have been allocated to the development of Autonomous Vehicles (Varma & Combes, 2020b). McKinsey and Company refers to these disruptive factors as ‘diverse mobility, autonomous driving, electrification, and connectivity’(*Disruptive Trends That Will Transform the Auto Industry | McKinsey*, n.d.), While Infosys refers to them as the ‘electric, autonomous, connected (EAC) vehicle’. Drawing from both academic and non-academic literature, these factors can be summed up under the acronym ‘AECS’ which refer to the Automation, Electrification, Connectivity, and Sharing possibilities of the next generation of vehicles. These factors, taken together, hint at a paradigm shift characterized by, among other changes, different patterns of human-vehicle interaction (i.e. the non-necessity of a human driver), stricter emission norms, changed ownership scenarios (i.e. sharing instead of owning) and a higher ratio of time utilized per day of vehicles.

This transition to AECS will affect vehicle classes differently†; AECS cars, buses, trucks and Light Commercial Vehicles (LCV’s) can each be expected to be used in a singular way, different than currently used, and used differently than each other. This difference of usage will stem from not only the transition to ACES, but also the context of usage of the vehicle (ex. Commute/mobility or freight transport.)

1.1. Urban Logistics

1.1.1. What is Urban Logistics?

According to Laetitia Dablanc, Urban or City Logistics *"can be defined as any service provision contributing to efficiently managing the movements of goods in cities and providing innovative responses to customer demands. The objective of city logistics is to make deliveries to urban residents, visitors, and businesses possible at the highest economic, social, and environmental standards. City logistics includes physical operations such as order preparation and packaging, transportation and deliveries (including homedeliveries), short-term storage of goods, management of drop-off/pick-up boxes for parcels, return of goods, waste management including the management of recycled goods, and the management of empty pallets and packages. City logistics makes use of sophisticated information and communication technologies to increase coordination and efficiency and ensure adequate enforcement of urban delivery operations."* (Dablanc, 2019)

Trends like increasing urbanization (*68% of the World Population Projected to Live in Urban Areas by 2050, Says UN*, 2018), increasing e-commerce, negative externalities linked to Urban Logistics (congestion, pollution, etc.), coupled with the inherent importance and problematic nature of Urban Logistics have generated many studies addressing various issues. (Holguín-Veras et al., 2018a) and (Holguín-Veras et al., 2018b) through a review of the public-sector initiatives that could be used to improve freight activity in metropolitan areas produce a ranking of suggested initiatives. (Combes, 2019) presents a structural microeconomic model of the choice of warehouse location in urban logistics, which focus on operational constraints, their diversity, and their influence on costs to arrive at an understanding of current trends and recommendation for land-use planning. (Quak et al., 2014) discusses the business aspects that need to be considered when implementing or scaling up city logistics solutions by using the business model canvas, which is accompanied by data on field demonstrations which show improvements on operational efficiency, flexibility and environmental impact. (Gatta et al., 2019) make an exploratory analysis which aims at understanding and evaluating the environmental and economic impacts of a crowd shipping platform in urban areas through modelling on the city of Rome. The literature review above aims, through mentioning a small number of papers, to show the diverse and eclectic nature of this subfield of transportation research.

1.2. Autonomous, Electric, Connected and Shared Vehicles

While autonomous vehicles have had a rich history (ex. The EUREKA funded PROMETHEUS project), due to a combination of factors including, but not limited to, advances in telecommunications, widespread GPS availability, increasingly reliable and fast Internet connectivity, increased processing power at reduced costs, and increased sensitivity and desire for sustainability, this trend has re-entered importance. Various large OEM's have invested in this technology – and start-ups have raised money – both to the tunes of billions of dollars (Varma & Combes, 2020b).

However, it is important to consider that the next paradigm shift may be a combination of 4 factors; the removal of the human component, the electrification of the drive train, the connection of vehicles with each other and other technologies, and the changing ownership models – the possibility of sharing instead of owning - brought about due to various connected technologies. These factors are considered together as they are complementary disruptors; the potential disruption is 'multiplied' when these factors are considered together. In more concrete terms, if a vehicle is considered 'autonomous' in the strictest sense, the only value addition it brings is the removal of a human driver – however, adding the possibility to be connected to the internet enables a host of economically lucrative possibilities (ex. Robo taxis). However, this understanding is not always reflected in the literature. Out of various articles consulted for this paper, only 1 refers to all 4 aspects together. (Another component not mentioned is the changing shape of the vehicles, however it is out of the scope of this article and has not much literature on it)

Based on the available literature, it is difficult to argue that autonomous vehicles may not be both useful or used in logistics. (Bucsky, 2018) inspects the potential of autonomous vehicles across different freight modes (road, rail and urban logistics) and concludes that in urban logistics, the connection between the transport of individuals and freight transport can offer efficiency gains, though regulation and safety can be a bigger challenge, however market push effects of the rise of e-commerce will help developments in this field. (Meldert & Boeck, 2016) study vehicle automation technology, liability and legislative challenges, along with the ethics & human factors challenges applied to the usage and potential consequences of AVs for the logistics industry and conclude adoption of AVs holds the promise of innovating the way in which mobility and transportation logistics are dealt with. Studies inspecting the impact on business models have also emerged (ex. (Fritschy & Spinler, 2019)) indicating interest across the ecosystem.

Apart from companies such as Waymo which are working on automation of personnel transport, other companies are working specially on delivering freight. For example, Nuro a start-up based in Arizona recently raised almost a billion dollars for their local goods delivery vehicle. (*Nuro — Product*, n.d.) Mercedes has successfully demonstrated a service which integrates a drone and an autonomous Light Commercial Vehcle (LCV) to deliver small parcels. (*Mercedes-Benz*, n.d.) Renault, under the aegis of CityPod has been working on a modular autonomous vehicle, with similar loading volumes as an LCV. These vehicles can be broadly classified into 3 types (Figure 3); Multiple delivery vehicles (Ex. Nuro), Single delivery pods (ex. Amazon Scout), and non-standard delivery bots (ex. Ford's Digit).

Image 1: Nuro, Scout and Digit (L-R)

1.3. Research Questions

In light of recent trends linked to urbanization, urban logistics and AECS vehicles (mentioned above), through this paper, we examine the question of the suitability of AECS vehicles for urban deliveries, notably Autonomous Multiple Delivery Vehicles.

Using, as the foundation of our understanding, data from field observations, we closely inspect the process of urban deliveries to identify the suitability of substitution of conventional vehicles by AECS vehicles. We examine with attention the role of the driver of the vehicle in the delivery process in order to better envision autonomous deliveries.

2. METHODOLOGY/RESULTS

In this section, first the field observations, and engendered questions – linked to the application of autonomous vehicles in urban logistics – are described. Areas of necessary human intervention during the delivery process are delineated. Finally, financial aspects of different scenarios, including when the delivery process is replaced with an autonomous vehicle with the aspects of human interaction handled remotely, are examined.

2.1. Field observations

The information collected for this paper was gathered during various field visits. Most notably 2 days were spent delivering parcels with a freight company in the outskirts of Paris, France.

In this organization, at 7 am each working day, the drivers report to a centralized office, where each driver is allocated their round for the day. The roles of the 'drivers' and 'managers' are clearly demarcated; the managers handle the planning of the delivery (creation of delivery round, distribution among drivers) in the office. The drivers handle the execution of the delivery in the field. Then, the driver is handed their round for the day in the form of slips of paper. Each of these slips represents an individual delivery and contains the address and contact details for the respective client. Also, these slips are presented to the driver as per a computer specified order (an optimal round). However, at this particular logistic hub, the opening/closing time of the enterprises are not considered by the computer at the time of generating the round. Using this (time of opening hours) and other factors, the drivers rearrange their round. In the image above, a driver (Abdel) is reorganizing his round. In general, a driver receives roughly 50 slips per round in this facility.
At the end of this reclassification, the drivers start loading their vehicles in the warehouse. While conditional on the number of doors on the vehicles, the objects are generally loaded 'first in last out'. This process takes roughly 60-90 minutes. Sometimes there are instances when the objects, according to the computer specified round, do not fit in the vehicle. In these cases, the driver does not load these objects, and leaves out the corresponding slips from the delivery round. When the vehicle is loaded, the driver recontacts the manager; the reordering (along with the final number of deliveries (in case some objects are not loaded) is validated by the manager. At the end of this step, the driver is ready to start the delivery process.

The delivery process had several peculiarities. Firstly, in both the observation cases (with different drivers and different vehicle types), the drivers were intimately familiar with the area of delivery and the addresses. In the case of the larger vehicle, since most of the deliveries were to business, the driver, based on his knowledge of opening times of various enterprises did not have to make as many individual calls as the driver of the smaller vehicle. The driver of the smaller vehicle, as he was dealing primarily with B2C deliveries made calls for almost every delivery.

Thus, the order of activity is as follows. Once the driver has mounted the vehicle, s/he will contact the first client, if the first client is not a business. If the client responds, based on whether s/he is ready, the driver will drive to the specific client. The driver did this, in almost all instances while *actively engaged* in the act of driving the vehicle. Once the driver arrives at the location s/he will locate the specific parcel in the storage and make the physical delivery (in return for a signature). At the end of a delivery, the driver reenters the vehicle and repeats the cycle of contacting and delivering. At the end of all deliveries, s/he makes his/her way back to the warehouse. In these two observations, only 1 delivery (out of approximately 90) was reattempted on the same day[1]. The drivers did their best to achieve delivery on first attempt.

In this organization, every working day, delivery rounds were generated by software. Then, these delivery rounds were allocated to drivers. The vehicles are loaded based on the round, following which the driver starts the delivery process. The delivery activities are carried out in the following order. Clients are contacted and if they are available, the driver drives to them to complete either the delivery or the return. In almost all instances, the driver contacts the clients simultaneously while driving. Once at the location, the driver locates the specific parcel in the storage and make the physical delivery (in return for a signature).

A rough, qualitative estimate of the time spent per activity in the delivery process is presented in Figure 1.

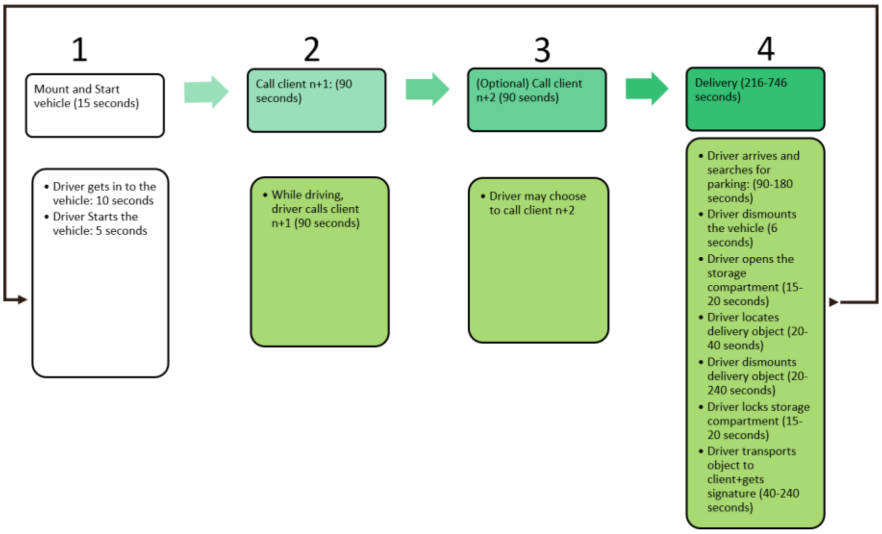


Figure 1: Time spent per activity

2.2. Deliveries using an autonomous vehicle without physical human assistance

This section discusses the possibility of using an autonomous vehicle in the process sequence described above.

2.2.1. Type of autonomous vehicles suitable for urban deliveries

According to the SAE, a level 5 autonomous vehicle can accomplish 'the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver'. In other words, no human

[1] This was a special scenario concerning delivery to a supermarket located in a shopping mall. Owing to the uniqueness of this location, the entry to the logistic bay was often crowded and could imply larger wait times, leading to over an hour spent on one single delivery, disrupting the rest of the round. In this instance, once the driver attempted delivery as per the order of the round, and once after finishing some orders, when generally the bay is less crowded. However, as it was still crowded (this can be visually identified based on the number of vehicles queueing to enter), the reattempted delivery was abandoned. It should be noted that the product delivery was not urgent, and that the driver makes multiple trips per week to this same supermarket, thus, the marginal cost of transferring this delivery to another day was small. These were factors that the driver was aware of and could include as parameters while making decisions in real time.

intervention for driving is necessary(Varma & Combes, 2020a).

To strategically envision the use of an autonomous vehicle, the delivery process was examined based on this definition. Attention is placed on the part after the loaded vehicle along with driver is leaving the warehouse.

We imagine a scenario where a level 5 autonomous vehicle, equipped with an appropriately filled mobile locker is deployed. We base our analysis on an autonomous vehicle equipped with a mobile locker as it enables the separation of human intervention. Our choice is based on these three reasons:

- It can accomplish multiple deliveries on the same round (i.e. without returning to the warehouse)
- For each delivery, a unique 'password' (using a smart device ex. Smart phone, computer, etc.) can be issued to the end customer, which enables opening a predetermined locker on the autonomous vehicle. The merchandise located in the appropriate locker is then collected. This takes the place of the 'signature' in traditional deliveries.
- It can accommodate parcels with different physical characteristics.

Any other vehicle type (under development or test, to the best of our knowledge) does not fulfill these criteria. Indeed, in a related article (Figliozzi and Jennings, 2020) come to the same conclusion.

Iamge 2: Some autonomous mobile lockers under test and development

The following information should be noted.

- Single delivery bots (ex. Scout, Figure 1), Cargo Delivery bikes, Drones, On Foot Deliveries, Autonomous Vehicles with a human on board are not considered in our analysis. While these might be theoretical responses to urban deliveries (refer to Figure 2), they are exempted for one or more of the following reasons.
  - A central focus of this paper is the role of the driver in the urban delivery scenario. Our aim is to elucidate the practical and financial aspects of deliveries without a human driver. Thus, any other alternative with a human operator are not considered.
  - To make a relevant comparison with the current scenario in which a vehicle leaves a warehouse, generally situated not close to the sites of delivery, and makes multiple deliveries in one round, we do not consider single delivery vehicles.

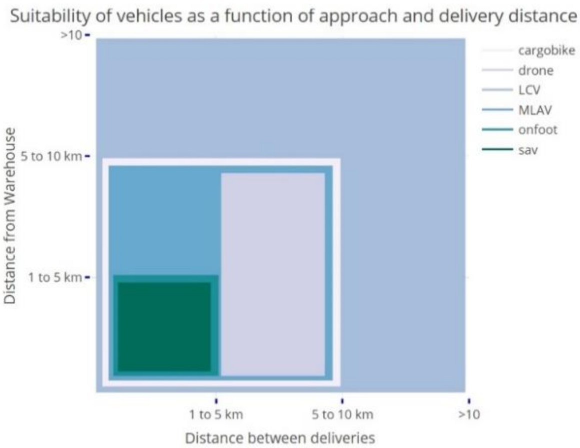


Figure 2: Suitability of various vehicles as a function of distance from warehouse and between deliveries

2.2.2. *Exchanging information with customers*

While this vehicle can drive to various customers, the question on how these customers might be notified about the imminent arrival is raised. The fact that the customer is aware and consents to the time and date of delivery or return is important for at least two reasons:

1. To avoid a delivery/return failure; in case the customer is not ready to accept the delivery of or return some merchandise, it must either be returned to the warehouse

and accommodated on another round in the coming days or collected another day. This is economically costly.

2. To avoid excessive time spent on a delivery/return; having an unprepared customer will increase the time spent per delivery/return. As is, from the supply side (logistic companies), urban logistics is an operation with very narrow profit margins. These margins heavily depend on the number of parcels delivered per day.

It should be noted that in the field observations of conventional deliveries, the driver of the vehicle is accomplishing the task of calling the customers and making a real time decision on whether to drive to them parallel while driving.

Table 1: Shift of tasks between conventional and autonomous vehicles

S.No	Task	Handled by in Conventional Delivery	Handled by in Autonomous Delivery
1.	Navigating	Driver	Vehicle
2.	Calling and notifying customer of arrival	Driver	Call Center Operator
3.	Locating merchandise in storage	Driver	Customer
4.	Unloading merchandise	Driver	Customer
5.	Delivering merchandise to end customer	Driver	Customer (customer collects it him/herself)
6.	Getting validation of successful transaction (delivery) from customer	Driver	Vehicle (registers opening/closing of door)
7.	***Latent knowledge*** - *Creating personal relationships* - *Looking for parking* - *GPS overriding* - *Memorized entry codes* - *Real time rerouting*	*Driver*	*Call Center Operator?*

To handle this part of the delivery process, our scenario assumes the presence of a **remote management system supervised by a human**, who apart from remote monitoring and guidance, also handles the calling and letting customers know about the vehicle arrival and confirms they are ready for the delivery. For the rest of the document, the term ***autonomous delivery*** is used to represent a *delivery undertaken by an autonomous vehicle with remote human supervision (no physical human assistance)*, and the term ***conventional delivery*** is used to represent *a delivery undertaken by a non-autonomous vehicle with a human driver*. Table 1 summarizes the shift of tasks from the scenario of a standard delivery to an autonomous delivery scenario.

2.2.3. The role of Latent Knowledge

Often, logistics companies allocate their drivers to specific 'delivery rounds' within the same geographic area or patch over a long period of time. By maintaining them on the same rounds, carriers attempt to build up the driver's familiarity in a round and surrounding area, helping them learn the most efficient routes, build personal relationships with customers and maintain knowledge of a round. Owing to repeatedly deliveries in the same area, the driver develops an extremely useful, practical, understanding of the area and its inhabitants. A non-exhaustive list of these understandings and their applications are reflected in Table 2.

The effect of this latent knowledge on delivery effectiveness parameters, like time spent on average per delivery, first time delivery success rates, time spent of average for parking, etc. is, very poorly tackled in the literature, and to the best of the knowledge of the authors, not reflected in any study doing economic cost considerations. However, the effect of this knowledge can be extremely significant on some performance parameters. For example, according to one of the rare studies which explicitly consider this "*the difference between two of our drivers (D22 and D24) with similar round sizes and parcel volumes shows a considerable variation in effective- ness, with D22 driving 44% less distance, spending 35% less time per parcel, 29% less driving time per parcel, and 39% less parking time per parcel. The variation in effectiveness of our drivers relates to better route planning, exploitation of accumulated knowledge of the round, personal relation- ships with other stakeholders, the amount of time spent at the curbside and the influence of walking. These statistics show that more effective drivers achieve higher rate of delivery of parcels per minute while spending less time driving and parking in the van*".(Bates et al., 2018) The relevant matter of the transfer of these significant efficiency gains in the context of autonomous deliveries is considered in the discussion section.

Table 2: Latent knowledge and its applications

S.No	Skill/Knowledge	Application
1.	Personal relationships with various inhabitants	- Alternate delivery possibilities (ex. If customer is not available, driver delivers to neighbour, based on previous agreements) - Accommodating non prepared return on other rounds (Especially true for business deliveries, If a return form a customer is not prepared, s/he can call the driver and inform them, making the round more efficient) - Concierge
2.	Knowledge of Parking spaces	- Reduction of time spent to look for a parking space - Knowledge of parking time restrictions - Alternate parking spots
3.	Knowledge of traffic conditions and trends	Ex. Higher traffic in a specific repeated delivery address at a particular time (commercial center) - Reordering delivery order to achieve faster overall delivery
4.	Knowledge of geographical quirks	Ex. Access codes, GPS Map Failures : Dead Ends - Driver saves a lot of time by already knowing the access code, or the requirement of it. - Driver aware of GPS failures in certain specific scenarios, ex. dead ends, and avoids them.

2.3. Cost competitiveness and financial considerations

To have any incentive to switch to autonomous deliveries, we assume that some value must be added in the delivery process. This added value could be for the end customer, or for the organization fulfilling the delivery.

For the customer, this added value could take the form of Reduced delivery prices and greater flexibility regarding delivery times. For the logistics provider, this value could arise from reduced delivery costs arising from reduced first time failure rates and reduced delivery times (as deliveries can be scheduled), or from more deliveries per round, higher energy efficiency of vehicles, owing to flexible hours (autonomous vehicles can deliver till the limits of their autonomy). It could also arise from better delivery management since the vehicles are connected and can provide real time fulfilment data along with better scheduling and higher customer satisfaction (if the customers can, for example, schedule the delivery more flexibly).

In our analysis, we develop an economic model focused on the delivery costs. We assume the demand is fixed. The cost[2] is an extremely important decisional parameter for players in the domain of logistics, ex. for fleet managers the TCO is the most important consideration.(Nesbitt & Sperling, 2001).

We first make the distinction between a conventional delivery (i.e. a non-autonomous vehicle operated by a human) which is denoted by S and an autonomous vehicle remotely operated by a human operator denoted by AV. We then incite the following concepts:

The vehicle capital cost is denoted by C_k(€/day), the operating costs by C_o(€/km) and the driver costs by C_w(€/h)

The approach distance is denoted by d_a and the distance between deliveries is denoted by d_z. The distance per round then becomes $D = 2.d_a + n.d_z$

The number of rounds per day is denoted by n_r, the number of deliveries per round denoted by n. We thus arrive at the number of deliveries per day, denoted by $N = n_r \times n$

The time spent for delivery is denoted by h. We denote the delivery time window by H. Some logistic companies make multiple deliveries rounds per day, thus H takes the form of 24 hours. In some other cases the logistic company might only make a single round. The round duration[3] thus take the form of $H_r = 2\frac{d_a}{v_a} + n\left(h + \frac{d_z}{v_z}\right)$. By extension, the number of rounds $n_r = \frac{H}{H_r}$.

The user cost per delivery is denoted by u. it stands for the fact that a) the customer may have to participate to the realisation of the delivery and b) the level of service does not fit their preferences perfectly. Without loss of generality $u^s = 0$.
The following operating constraints are assumed.

[2] Greater flexibility regarding delivery times for customers, better delivery management and higher customer satisfaction that may be offered by autonomous vehicles, for example, are also relevant parameters for the logistics provider. However, at this stage of development of autonomous vehicles the magnitude and direction of the changes associated with these parameters due to the advent of autonomous vehicles remain nebulous and thus difficult to quantify.

[3] This definition of the round duration does not include the charging or the loading and unloading time. The complete definition thus becomes $H_r = 2\frac{d_a}{v_a} + n\left(h + \frac{d_z}{v_z}\right) + (charging\ time) + (loading\ time) + (unloading\ time)$.

- In the case of a conventional delivery S, the delivery time window and round duration (delivery time window duration $H(h)$)
- In the case of autonomous vehicles, AV, the operating constraints take the form of the max number of shipments carried by the vehicle when it leaves from the warehouse. This is denoted by n^{AV}
- We arrive at the Total Cost of Delivery (TCD) which is denoted as $TCD = C_k + C_w.h + n_r.D.C_o$
- And finally, we denote the Cost per Delivery $C = \frac{TCD}{N}$

2.3.1. *Case 1: Conventional Delivery*

In the case of conventional deliveries S the TCD takes the form of

$$TCD^S = C_k^S + C_w^S.h^s + D.C_o^S$$

In the European Union, a Light Commercial Vehicle used in an urban context, has an effective life span of 10.7 years. Total cost of ownership (TCO) of a light commercial vehicle used in an urban context is calculated over 10.7 years. (Lebeau et al., 2019). We found the work of the MOBI laboratory in Belgium to be most adapted for the extrapolation of figures for this article as they use a consistent, comprehensive methodology over various papers applied to vans in urban logistics. Notably (Lebeau et al., 2019) considers explicitly the case of electrical vans in urban deliveries. Electric vans for urban deliveries are currently the 'conventional state of the art', further, as described above, autonomous vehicles may probably be electric. Thus, figures from this publication are used.

They note that the average TCO (in eurocents/km) across 18 vehicles is 36.82 €c/km. Further, the average kms travelled per year are found to be 16000, with 260 operating days. Thus, the average kms travelled per day are roughly 61.54km/day. Thus, we note $C_k^S + D.C_o^S = 22.66€/day$. Through rough salarial estimates obtained during our field observations, we note that the drivers earn roughly 9-12€/hour. We also note that the working hours can vary between 6-10 hours/day. We assume the hourly wage as 10€/hour and the hours worked per day as 8. Thus, $C_w^S.h^s = 80€/day$

$$TCD^S = C_k^S + C_w^S.h^s + D.C_o^S = 102.66\ €/day$$

We further assume, based on subjective estimates from field observations, n^Sas 50 and n_r^S as 1. Thus, $N^S = 50$. We arrive at $C^S = \frac{102.66}{50} = 2.0532\ €/delivery$

2.3.2. *Case 2: Autonomous Delivery*

The cost of autonomous vehicles is difficult to precise. Sensors and technology required would probably imply a higher purchase cost for autonomous vehicles. Though, this cost might be offset by the fact that it can operate continuously (apart from refueling/charging time). However, these savings cost might also be offset by the mechanical wear and tear of operating under these conditions. This might manifest itself as a shorter life of the vehicle and a reduced, if any, resale value. In essence, there are various factors which may contribute to either an increased or decreased total cost of ownership of an autonomous vehicle. Literature on these is currently sparse, however there appears to be a general consensus that the purchase cost would be higher, at least initially, due to the cost of technology.(Bösch et al., 2018) Some of these seemingly contradictory forces are delineated in table 3. For our analysis, however, instead of speculating the direction and magnitude of these forces, we consider the TCD^{AV} and N^{AV} to vary from half to three times that of conventional vehicles. Thus, the total cost of delivery for autonomous vehicles is represented as;

$$TCD^{AV} = C_k^{AV} + C_w^{AV}.h^{AV} + n_r^{AV}D.C_o^S + u^{AV}$$

In our autonomous delivery scenario, the deliveries are remotely supervised, and calls made by a human. We assumer a similar cost structure as assumed in our calculations regarding conventional deliveries. Thus, $C_w^{AV}.h^{AV} = C_w^S.h^S = 80€/day$. It is possible that the level of service with autonomous deliveries may not be the same as conventional deliveries, with $u^{AV} \neq 0$. While the implications of this (and the difficultly in quantifying it) are discussed in the next section, for our calculations, we assume that $u^{AV} = u^S = 0$.

$$C^{AV(Single\ Vehicle)} = \frac{TCD^{AV}}{N^{AV}}$$

There is sparse literature on TCO of autonomous vehicles, thus it makes it difficult to have numerical values, and instead we address the ratios.

Let's consider the diagram of the division of time (Figure 1.). We note that the driver is engaged in calling customers parallelly while driving for an average of 180 seconds out of the 691 seconds or for an average of 26% of the time. Assuming contacting the customer takes a similar amount of time in the scenario of an autonomous delivery, the question of how the human operator uses the rest of (74%) of the time becomes. We suppose that the economic incentive lies in that the operator uses this time to handle other vehicles. Taking P to be the number of vehicles handled simultaneously, we arrive at the following cost per autonomous delivery

$$C^{AV(multiple\ vehicles)} = \frac{P.(C_k^{AV} + n_r^{AV}D.C_o^{AV}) + C_w^{AV}.h^{AV}}{P.N^{AV}}$$

2.3.3. Comparison

For delivery by autonomous vehicles to be cost competitive,

$$C^{AV(single\ or\ multiple\ vehicles)} \leq C^S$$

We take the following ratios

- $(C_k^{AV} + n_r^{AV}D.C_o^S) = a.(C_k^S + D.C_o^S)$
- $N^{AV(single\ or\ multiple\ vehicles)} = b.N^S$

Thus,

$$C^{AV(multiple\ vehicles)} = \frac{P.(C_k^{AV} + n_r^{AV}D.C_o^{AV}) + C_w^{A}.h^{AV}}{P.N^{AV}}$$

$$C^{AV(multiple\ vehicles)} = \frac{P.a.(C_k^S + D.C_o^S) + C_w^S.h^S}{P.b.N^S}$$

$$C^{AV(multiple\ vehicles)} = \frac{P.a.(22.66) + 80}{P.b.50}$$

We take that P varies from 1 to 4 (I.e. 1 human operator can remotely handle 1-4 vehicles), while a varies from 0.5 to 2 (I.e. $(C_k^{AV} + n_r^{AV}D.C_o^S)$ is 0.5 – 2 times $(C_k^S + D.C_o^S)$) and finally b varies from 0.25 to 1 (I.e. $N^{AV(single\ or\ multiple\ vehicles)}$ is 0.25 – 1 times N^S).

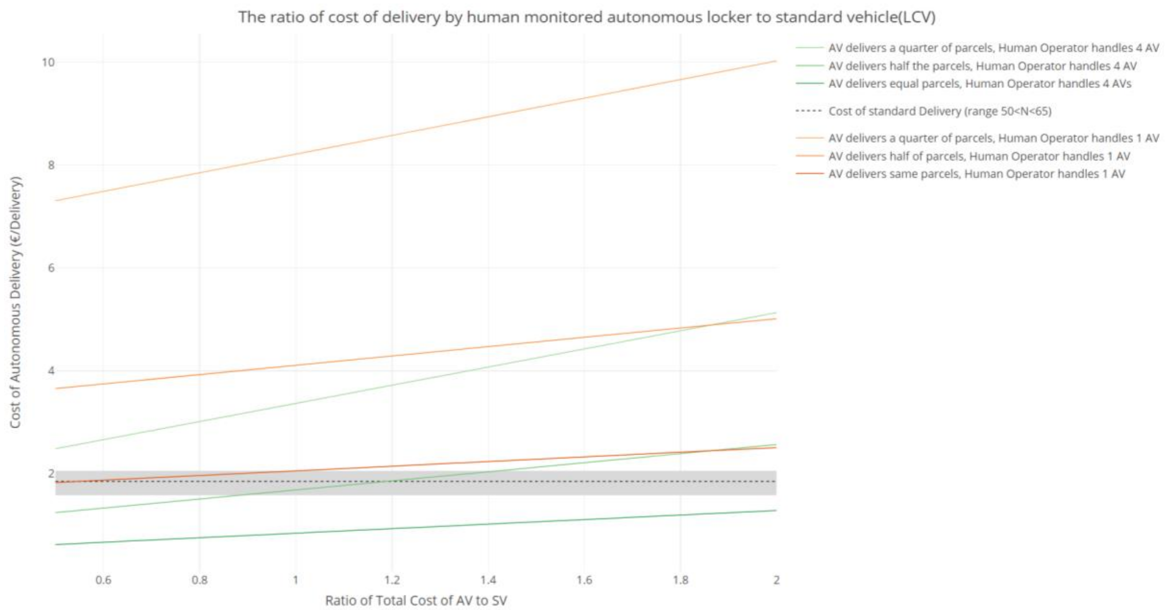


Figure 3: The ratio of Cost of Autonomous Delivery to the Cost of Conventional Delivery (with P=1 or P=4)

In the graph above we consider that the human operator remotely handles either 1 vehicle or 4 vehicles. We choose the upper limit as 4 as in our analysis of the time spent per activity of the delivery process, the driver is involved in making phone calls 26% of the time. If a human operator is concerned with exactly this part of the delivery process then, for 26% of the time he is involved in calling, but free the other 74% of the time. Thus, there is economic incentive for him to handle more vehicles parallelly. Thus, if it takes on average 26% of the time, on average he can handle (close to) 4 vehicles in parallel.

We then vary the number of deliveries that can be accomplished in the same round by an autonomous vehicle. The average number of deliveries in our observations were close to 50. Though there exist other estimates for the average number of deliveries (ex. (Beziat, 2017) states that this number is close to 30). However, based on various sources, it seems that an autonomous multiple locker would make fewer deliveries, thus we assume that it makes between 12 to 50 deliveries. We not that there are a few cases where autonomous delivery is cost competitive and might even be cheaper. This generally occurs since the human operator handles multiple autonomous vehicles. In other words, conditional on the fact that autonomous vehicles cost the same (unlikely), and can deliver similar number of vehicles in the same round (also unlikely), only the fact that a human handles multiple autonomous vehicles will lead to an economic incentive for autonomous deliveries. This is assuming no loss of efficiency due to the experience of the driver (which is also unlikely).

3. DISCUSSION

In the results section, we present a mathematical model which compares the cost of delivery in an autonomous scenario to the cost of delivery in the conventional scenario. This economic model aims to find a combination of parameters at which, theoretically, fleet managers/logisticians are indifferent between using autonomous or conventional vehicles for urban deliveries. There are various caveats to this form of modelling, however, especially if this issue is addressed practically rather than theoretically. These are discussed below.

3.1. Financial calculations

Firstly, we discuss some caveats and trends associated with the parameters used in the model. Autonomous vehicle technology is changing at a rapid pace. This has implications for the vehicle capital cost C_k, the operating costs $C_o$and N^{AV}parameters used in our calculations. Notably, both parameters are, in turn, dependent on various other parameters. These other parameters can be analytically reasoned to evolve, financially speaking, in opposite directions. A non-exhaustive list of these parameters and their evolutions, and the reason for these evolutions, is presented in Table 3 below.

Table 3: Evolution of parameters linked to Cost per Autonomous Delivery

Parameter in TCD^{AV}	Expected to	Due to
Vehicle capital cost C_k	1. Increase 2. Increase 3. Decrease	1. Higher purchase cost due to expensive sensors (ex. Lidar) 2. Higher maintenance cost due to TKD increase (more hours of operation) 3. Lower maintenance cost due to fewer moving parts due to electric
Number of Deliveries (N^{AV})	1. Increase 2. Decrease 3. Decrease 4. Increase/Decrease	1. Hours of operation 2. lower operating speed 3. smaller volume (ex. Nuro) 4. Potentially longer (if accomplished by humans) or shorter (if accomplished by sophisticated machines) loading time

Thus, it becomes hard to predict the overall direction of movement about the Total cost of Ownership of Autonomous Vehicles, and more so in the context of urban logistics.

(Figliozzi & Jennings, 2020) construct a similar financial model considering various financially relevant parameters. Our work, while modelling similarly, is idiosyncratic in proposing a solution that considers real world constraints and fits in the existing urban logistic ecosystem.

3.2. Level of service and customer willingness to pay

In a scenario where deliveries are accomplished by an autonomous mobile locker (AML), it is reasonable to consider that there might be a loss of service from the customers perspective. While traditionally the customer could expect to be delivered at his/her doorstep, with AML's, s/he would be responsible for the last few meters of the last mile. This is represented as u^{AV} or u^{S} in our delivery cost model.

This purported reduction in the level of service raises pertinent economic questions. If, on the supply side, the switch to autonomous vehicles is engendered by a financial incentive, then, a part of this incentive could be offered to the customer in compensation for the reduced level of service. However, how much does the customer value the service provided in the last few meters of the last mile? In other words, if the customer could choose between a conventional delivery (which covers the entire last mile), and delivery by an AML (which covers all but the last few meters), what would be his/her *willingness to pay* for the former? If a monetary value, albeit subjective (as the last few meters will have different importance to individuals, for example, with limited mobility compared to agile individuals) can be assigned to it, what if this number exceeds the total monetary value created by the adoption of autonomous deliveries? While simplistic, the argument serves to stress the intricate nature of urban logistics, highlighting that to effectively contribute to an understanding of the possibility of adoption of autonomous vehicles for urban logistics, examining adoption scenarios on a case to case basis is probably more fruitful; what may work for transport of hotel linen from the laundry service may not work for delivery of parcels to dispersed apartment blocks. This general research recommendation will also be reflected in our further research.

3.3. Latent knowledge of the driver

Another relevant aspect, which has been briefly discussed above, are the efficiency gains linked to an experienced driver. From a psychological perspective, physically interacting with the environment, the driver passively gathers and stores various cues like social relationships (a non-exhaustive list has been presented in table 2). These understandings are unique to physical interaction; the driver would use, for example his/her spatial memory to keep a mental repertoire of parking spaces, which a human operator remotely controlling an autonomous vehicle would be unable to. Further, various studies (ex. (Kolb & Kolb, 2005)) have found that humans encode cues in both long term and short-term memory when various, rather than a singular, sensory modality is engaged in a task. Thus, human remote operators, may be generally overall less efficient at deliveries.

As mentioned flowing figure 3, it appears unlikely that the economic incentive to switch to autonomous modes of deliveries would be due to the removal of the human component. Rather, it appears that in extremely specific cases, for example for dispersed e-commerce orders with a truly short lead time from a warehouse close to the final delivery locale (ex. a supermarket), a premium delivery fee could be charged to realize a delivery with a single delivery bot. We consider this as a theme for future research.
Evolution of Conventional vehicles (and urban deliveries with them)
While we have painted a generous picture of the potential impact of autonomous vehicles and deliveries on urban logistics, this analysis would be incomplete without considering the alternative: conventional vehicles.

How are conventional vehicles evolving? And where will they be in the next few years? Higher fuel efficiency can be expected. Parallelly, battery technology has been improving exponentially. While studies have already shown (ex.(Figenbaum, 2018)), that battery electric vehicles are already suitable for a number of deployment scenarios in urban logistics, increased battery efficiency may render them extremely cost competitive, maybe even to the point of surpassing convention fuel vehicles. Simultaneously, developments in computer science, communication, and information technology will continue to yield information flows (ex. inventory management, optimizing order of deliveries, scheduling orders) more efficient, further reducing the cost of deliveries. Continued rise in e-commerce may render centrally situated warehouses cost competitive, leading to reduced delivery times and costs. The general direction of these trends is comparatively convergent and point in the direction of reduced delivery costs for conventional vehicles.

Another aspect consider is the development of artificial intelligence. The development, deployment, and utility of autonomous vehicles is symbiotic with that of AI, especially concerned with (but not limited to) the software. For example, advancements in AI could, theoretically, eliminate the need of a LIDAR through advances in image processing software. Advancements in AI could perhaps also address the efficiency gains linked to the latent knowledge of a driver. The development of AI, and its application to AV's is well beyond the scope of this article. However, we acknowledge both the critical place it occupies, and the difficulty in prediction of its developmental milestones, linked to its application to autonomous vehicles, and subsequently their use in urban logistics.

4. CONCLUSION

In this paper, we begin by briefly discussing urban logistics, its contemporary issues and the evolution of autonomous vehicles. Considering the case of urban deliveries, we present data from field observations. Based on which, we develop a scenario of urban deliveries involving a remotely monitored autonomous mobile locker. We mention associated challenges and present financial calculations based on an economic model. This model and the methodology addressing the subject of cost competitiveness between autonomous and conventional deliveries, and its development based on real world observations are the primary contribution of this paper. We further examine the results, their implications and limitations the discussion section, concluding that there are various, uncertain factors shaping the adoption and development of autonomous vehicles, and highlight the prudence in pondering the issue of adoption on a specific, case to case basis, rather than general calculations, which we also present as a general recommendation to the field.

REFERENCES

[1] *68% of the world population projected to live in urban areas by 2050, says UN*. (2018, May 16). UN DESA | United Nations Department of Economic and Social Affairs. https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

[2] Bates, O., Friday, A., Allen, J., Cherrett, T., McLeod, F., Bektas, T., Nguyen, T., Piecyk, M., Piotrowska, M., Wise, S., & Davies, N. (2018, April 20). *Transforming last-mile logistics: Opportunities for more sustainable deliveries* [Proceedings paper]. Conference on Human Factors in Computing Systems - Proceedings; ACM. http://doi.org/10.1145/3173574.3174100

[3] Beziat, Adrien. (2017). Approche des liens entre transport de marchandises en ville, formes urbaines et congestion : Le cas de l'Ile-de-France.

[4] Bösch, P. M., Becker, F., Becker, H., & Axhausen, K. W. (2018). Cost-based analysis of autonomous mobility services. *Transport Policy*, *64*, 76–91. https://doi.org/10.1016/j.tranpol.2017.09.005

[5] Bucsky, P. (2018). Autonomous vehicles and freight traffic: Towards better efficiency of road, rail or urban logistics? *Urban Development Issues*, *58*(1), 41–52. https://doi.org/10.2478/udi-2018-0022

[6] Combes, F. (2019). Equilibrium and Optimal Location of Warehouses in Urban Areas: A Theoretical Analysis with Implications for Urban Logistics. *Transportation Research Record*, 0361198119838859. https://doi.org/10.1177/0361198119838859

[7] Dablanc, L. (2019). City Logistics. In *International Encyclopedia of Geography* (pp. 1–8). American Cancer Society. https://doi.org/10.1002/9781118786352.wbieg0137.pub2

[8] *Disruptive trends that will transform the auto industry | McKinsey*. (n.d.). Retrieved March 5, 2019, from https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/disruptive-trends-that-will-transform-the-auto-industry

[9] Figenbaum, E. (2018). Can battery electric light commercial vehicles work for craftsmen and service enterprises? *Energy Policy*, *120*, 58–72. https://doi.org/10.1016/j.enpol.2018.04.076

[10] Figliozzi, M., & Jennings, D. (2020). Autonomous delivery robots and their potential impacts on urban freight energy consumption and emissions. *Transportation Research Procedia*, *46*, 21–28. https://doi.org/10.1016/j.trpro.2020.03.159

[11] Fritschy, C., & Spinler, S. (2019). The impact of autonomous trucks on business models in the automotive and logistics industry–a Delphi-based scenario study. *Technological Forecasting and Social Change*, *148*, 119736. https://doi.org/10.1016/j.techfore.2019.119736

[12] Gatta, V., Marcucci, E., Nigro, M., Patella, S. M., & Serafini, S. (2019). Public Transport-Based Crowdshipping for Sustainable City Logistics: Assessing Economic and Environmental Impacts. *Sustainability*, *11*(1), 145. https://doi.org/10.3390/su11010145

[13] Holguín-Veras, J., Amaya Leal, J., Sánchez-Diaz, I., Browne, M., & Wojtowicz, J. (2018). State of the art and practice of urban freight management: Part I: Infrastructure, vehicle-related, and traffic operations. *Transportation Research Part A: Policy and Practice*. https://doi.org/10.1016/j.tra.2018.10.037

[14] Holguín-Veras, J., Amaya Leal, J., Sanchez-Diaz, I., Browne, M., & Wojtowicz, J. (2018). State of the art and practice of urban freight management Part II: Financial approaches, logistics, and demand management. *Transportation Research Part A: Policy and Practice*. https://doi.org/10.1016/j.tra.2018.10.036

[15] Kolb, A. Y., & Kolb, D. A. (2005). Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education. *Academy of Management Learning & Education*, *4*(2), 193–212. https://doi.org/10.5465/amle.2005.17268566

[16] Lebeau, P., Macharis, C., & Van Mierlo, J. (2019). How to Improve the Total Cost of Ownership of Electric Vehicles: An Analysis of the Light Commercial Vehicle Segment. *World Electric Vehicle Journal*, *10*(4), 90. https://doi.org/10.3390/wevj10040090

[17] Meldert, B. V., & Boeck, L. D. (2016). *Introducing autonomous vehicles in logistics: A review from a broad perspective* (No. 543558; Working Papers Department of Decision Sciences and Information Management). KU Leuven, Faculty of Economics and Business, Department of Decision Sciences and Information Management. https://ideas.repec.org/p/ete/kbiper/543558.html

[18] *Mercedes-Benz: Vans & Drones*. (n.d.). Retrieved February 12, 2019, from /content/com/en/mercedes-benz/vehicles/transporter/vans-drones-in-zurich.html

[19] Nesbitt, K., & Sperling, D. (2001). Fleet purchase behavior: Decision processes and implications for new vehicle technologies and fuels. *Transportation Research Part C: Emerging Technologies*, *9*(5), 297–318. https://doi.org/10.1016/S0968-090X(00)00035-8

[20] *Nuro—Product*. (n.d.). Nuro. Retrieved February 12, 2019, from https://nuro.ai/product/

[21] Quak, H., Balm, S., & Posthumus, B. (2014). Evaluation of City Logistics Solutions with Business Model Analysis. *Procedia - Social and Behavioral Sciences*, *125*, 111–124. https://doi.org/10.1016/j.sbspro.2014.01.1460

[22] Varma, K., & Combes, F. (2020a, March 12). *LCV's in Urban Logistics: Are AECS vehicles conceivable?*

[23] Varma, K., & Combes, F. (2020b, December 7). *What's the Buzz About? Analyzing acquisitions by automotive companies*. Gerpisa colloquium, Paris (2020), Paris. http://gerpisa.org/node/6002

DEMAND-DRIVEN OPTIMIZATION METHOD FOR SHARED MOBILITY SERVICES

Cyril Veve
Nicolas Chiabaut
Ecole Nationale des Travaux Publics de l'Etat
3 Rue Maurice Audin, 69120 Vaulx-en-Velin, France
cyril.veve@entpe.fr; nicolas.chiabaut@entpe.fr

ABSTRACT

Shared mobility services are announced as a game-changer in transportation and a promising solution to reduce congestion and improve the performance of urban mobility. They could prefigure the arrival of autonomous vehicles. Modeling of these new services is a real challenge, especially because existing approaches are mainly an adaptation of methods devoted to classic transportation services. Consequently, this paper introduces a new data-driven optimization method fully devoted to shared mobility service. First, the proposed approach decomposes the recurrent demand based on its spatio-temporal features to overcome the drawbacks of the existing methods. Notably, it makes it possible to consider larger instances and to build robust solutions. Thus, recurrent demand patterns are identified to capture the potential demand of shared mobility services using a tailored clustering process. Second, a variant of Dial-a-Ride Problem is implemented to design robust lines to serve this demand. Such a hybrid method makes it possible to define relatively massive transport lines while maintaining spatial and temporal proximity to users' real demand. The method is then tested with an open-source dataset released by the New York City Taxi and Limousine Commission.

Keywords:
Clustering, Mobility pattern, Dial-a-Ride Problem, Similarity, Ride-sharing

1. INTRODUCTION

Shared mobility services are announced as a game-changer in transportation and a promising solution to reduce congestion and improve the performance of urban mobility. Moreover, it appears that recent studies on shared mobility, particularly regarding the real-time satisfaction of the demand, could prefigure the arrival of autonomous vehicles. Shared mobility consists in the shared use of a vehicle (car, motorcycle, scooter, bicycle, or other travel modes). Modeling of these new services is a real challenge, especially because existing approaches are mainly adaptation of methods devoted to classic transportation services. These methods can be classified into two main categories: the conventional methods and the dynamic methods.

The conventional methods are used to design rapid transportation lines such as subway, streetcar, or bus lines [1,2,3]. These approaches can be qualified as long term methods because the transportation supply is defined according to both urban planning purpose and transportation demand. Consequently, the goal is motivated by serving an existing demand but also to modify the travel behavior at a long time scale. These approaches are well known in the literature for many years. However such methods involve long term demand changes. Indeed, the deployment of such lines affects the socio-economic development around the lines stops, and so affects the demand of mobility. Thus, unlike dynamic methods, conventional methods are not adapted to an instantaneous mobility demand. These methods aim to respond to a demand for global mobility; the lines designed follow mobility corridors with a high concentration of departure and arrival points. Generally, the demand is estimated using online questionnaires, surveys, or historical moving data. However, these methods have some limits; the calculated flows do not take into account each trip's specificities but only a rough estimation of the movements of a large number of users. This spatial and temporal aggregation has led to the design of significant lines for which the stops are located relatively far from the real desired departure and arrival points for users. It has been shown in numerous studies that this problem of the last mile is one of the major brakes which prevents users of personal vehicles from deporting to shared modes of transport. It is to overcome this problem of the last mile that for the past ten years, taking advantage of the emergence of smartphones and underlying geolocation technologies, research has turned towards a new approach so-called dynamic.

Contrary to conventional approaches, dynamic methods aim at adapting the service supply to the real-time demand characteristics. One of the main benefits is that such approaches provide users with short-term access to a travel mode on an as-needed basis. These transport services may take different forms: station-based roundtrip services, station-based one-way services, free-floating services, etc. Similarly, many economic models exist to meet diverse user needs: public or private, membership-based, peer-to-peer (P2P), for-hire, or public transit system. Moreover, sharing can include either sequential sharing (i.e., different users sharing the same vehicle one after the other), or simultaneous sharing (i.e., sharing the same vehicle with multiple users for the same trip). Simultaneous sharing is a particular challenge that many services try to tackle: transportation network companies (TNCs) now offer ride-sourcing services (including shared taxi, shuttle, etc.); ridesharing (including carpooling, vanpooling, etc.) is becoming more and more popular. The main objective of real-time methods (or highly dynamic) is to match a maximum number of requests while minimizing objective functions, such as the users' waiting time or the total travel time. [4, 5] provide a list of objective functions and matching policies well known in the literature. The dynamic method is well adapted to large fleets of vehicles. The dynamic matching between users and vehicles is efficient when the number of vehicles is significant. It allows reducing the waiting time and detours to pick up or drop off a user. The dynamic approach is considered as a bridge technology that will be replaced by autonomous vehicles when the technology will be mature [6,7,8]. Despite their many advantages, the dynamic methods also have certain limitations. The real-time matching between travelers and vehicles is complex, and it can not be performed on large instances. The number of passengers served by a vehicle is often low. That is why the dynamic approach is not adapted to design massive customized lines of transport.

This paper aims to propose a new hybrid approach that allows the design of massive and robust lines of transport adapted to the daily demand. The main interest that is driven by the method, is to detect a large number of similar and recurrent trips to estimate the potential demand of shared mobility, then to design transport lines allowing pickups and deposits as close as possible to the real demand of the users. The approach is defined as hybrid because it take into account the regularity of trips over time (as in conventional methods); however, the distances to be covered and the delays observed for users remain relatively short.

This paper presents a method for estimating and servicing the potential demand for shared mobility. To do this, a similarity function is defined to evaluate the similarities between two different trips. Once this preprocessing has been carried out, a clustering method is used to obtain groups of trips with similar spatio-temporal characteristics. This method's variant is then used to detect groups of similar and recurrent trips over time (meta-clusters). Finally, a vehicle route optimization method is used to serve our displacement clusters' departure and arrival points.

The rest of the paper is organized as follows. Section 1 presents the data set and introduces the methodology used to estimate demand patterns. Then, Section 2 is dedicated to the design of customized transport lines to satisfy the demand. Section 3 focuses on the analysis of the results of the demand estimation and the planning of new lines. The last part is devoted to a final discussion.

2. ESTIMATION OF THE DEMAND

This section's main objective is to show how the demand can be decomposed into spatio-temporal areas containing a significant number of similar trips. The study focus on the demand of shared mobility in Midtown and Upper East Side. The objective is to present the method used to obtain clusters of similar and recurrent trips over time (meta-clusters). These clusters will be used to define the instances of the optimization problem presented in Section 3. The methodology is based on three steps: (i) definition of a similarity function to estimate the likeness between two trips; (ii) implementation of a clustering method to create clusters of similar trips; (iii) development of a method to detect recurrent clusters over time. However, it is essential to underline the fact that the demand is analyzed from a transportation point of view even if many other aspects could be taken into account: economic, social or behaviorial. Our method determines an upper bound of the potential of shared mobility.

We use an open-source dataset released by the New York City Taxi and Limousine Commission (data source: https://www1.nyc.gov/site/tlc/index.page). Although these data are not fully representative of human mobility since they only correspond to taxi trips, such a dataset provides an attractive proxy for studying the individuals' routes within a city. The study focuses on morning peak hours from 8h to 11h of June 2011. The area studied is a well known high-density area in terms of mobility in New-York City: Midtown and Upper East Side [9,10]. For each trip i, the dataset gives access to the following information: departure time t_i^{PU} and location $p_i^{PU} = (x_i^{PU}, y_i^{PU})$ of the pick-up of the passenger(s); arrival time t_i^{DO} and location $p_i^{DO} = (x_i^{DO}, y_i^{DO})$ of the drop-off.

First, it has been shown that the similarity function is used to quantify the likeness between two trips. Then the method used to detect groups of similar trips in different spatio-temporal areas is presented. Finally, we investigate if commonalities exist between the clusters of successive studied days.

2.1. Modeling similarity between individual trips

Firstly it is essential to define a similarity function to estimate the likeness between trips. The goal of such a function is to quantify the similarity between two trips. We use the similarity function presented in [11] because it was shown that it provides excellent results to estimate the similarity for the trips defined by a pair origin-destination. The similarity is calculated according to the spatio-temporal commonalities between the trips. Let $S(i,j)$ the similarity function between trips *i* and *j*. From a physical point of view, the intuition is that two (or more) travelers may have an interest to share their trip if they start in the same neighborhood and at the same moment, and want to go to the same destination. The function S must encompass these different spatio-temporal attributes of the trips. We proposed the following function:

$$\overline{S(i,j)} = \sum_{l \in [PU,DO]} \alpha_l e^{|f^l(i,j)|} \quad (1)$$

where $f^l(i,j)$ is the feasibility function and α_l is a coefficient. Function f describes the service's potential to operate the shared trips, i.e. the ability to pick up (or drop off) the two travelers before both of their desired departure times:

$$f^l(i,j) = |t_i^l - t_j^l| - \gamma d(p_i^l, p_j^l) \quad (2)$$

where γ is the average pace to connect travelers that want to share a trip. This parameter is a general and synthetic formula to describe the operation of the service and the way in which this service gathers two demand requests into the same vehicle: defining a meeting point, successive pick-ups, etc. For example, if the first traveler must walk to the second traveler's pick-up point, then γ is equal to the inverse of the walking speed. If this distance is traveled by car, meaning that the service offers door-to-door service, then γ is equal to the inverse of the vehicle speed. Consequently, f is positive if the match can be realized before the two desired departure times t_i^l and t_j^l, whereas f is negative if travelers have to experience delay to make the match possible. Moreover, α_l is equal to $\frac{1}{2}$ if $f^l(i,j) > 0$ and to $\frac{3}{2}$ otherwise because it is more disadvantageous to be delayed. In addition to this first index of similarity $\overline{S(i,j)}$, excessive distances/durations for rendezvous are penalized. Thus, penalties $\theta_x^l$and θ_t^l are added when, respectively, the distances between origin (or destination) locations and departure (or arrival) times of trips i and j exceed, respectively, specific thresholds, δ_x^l and δ_t^l :

$$\theta_x^l = e^{d(p_i^l, p_j^l) - \delta_x^l} \quad \forall l / d(p_i^l, p_j^l) > \delta_x^l \quad (3)$$

$$\theta_t^l = e^{|t_i^l - t_j^l| \cdot \frac{\delta_x^l}{\delta_t^l} - \delta_t^l} \quad \forall l / |t_i^l, t_j^l| > \delta_t^l \quad (4)$$

Otherwise, these penalties are null. In this manner, $S(i,j) = \overline{S(i,j)} + \theta_x^l + \theta_t^l$ defines a sharp function that enhances the differences between trips and facilitates identification of similar travelers in the dataset. Notice that S is minimal (and equal to 1) when the two trips are exactly identical.

2.2. Detection of similar trips for individual days

The method presented here allows us to detect spatio-temporal areas where there are a significant number of similar trips. The function of similarity exposed above only estimates the likeness between two trips but does not detect clusters of similar trips. In order to detect such groups, a clustering algorithm is used. The function of similarity allows us to compute the similarity matrix requested by a clustering algorithm. A variant of a well-known clustering density-based method DB-SCAN [12] is applied for each day to detect groups of similar trips. The classic DB-SCAN requires only two parameters: a threshold ϵ and a minimum number of points $MinPts$, which have to be in a radius ϵ so that the studied point is considered as an element of the cluster, see [13] for more details. The parameter ϵ is the maximal distance between trips, i.e., the maximal value of S, allowed to consider them as similar and group them into the same cluster. However, this method must be slightly adapted to detect groups of different density. Thus, a successive DB-SCAN clustering is performed, i.e. **Algorithm**

Algorithm 1 Iterative DBSCAN - *itdbscan*

```
function itdbscan(data, similarityMatrix, minPtsMin, minPtsMax, epsMin, epsMax, Q_max)
    mainClustering ← ∅
    id ← 0
    test ← 0
    for minPts in {minPtsMax,minPtsMin} by step of -1 do
        for eps in {epsMin,epsMax} by step of 1 do
            clustering, K = dbscan(data,similarityMatrix,eps,minPts)
            for k in (0,K) by step of 1 do
                if Q(k) ≤ Q_max then
                    Save k in mainClustering with the number id
                    id ← id+1
                    Delete travels of cluster k from data
                    test ← 1
                end if
            end for
            if test == 1 then
                if data == ∅ then
                    break
                end if
                Delete travels of cluster k from similarityMatrix
                test = 0
            end if
        end for
    end for
    return mainClustering
end function
```

1 *itdbscan* ([14] for more information), using the similarity function S as the distance, while updating iteratively the values of the parameters.

Starting with a large value of $MinPts$ (equals to $minPtsMax$) and a small value of ϵ (equals to $epsMin$), it makes it possible to identify large groups of travelers in the initial data set of trip. In other terms, we first detect large and high-density clusters. Then, the DB-SCAN method is applied on the remaining non-clustered trips to detect groups of size $MinPts - 1$. This process is repeated until $MinPts = minPtsMin$ and $\epsilon = epsMax$.

A cluster can be considered satisfactory if locations of origin / destination and arrival / departure times of the trips within the cluster are relatively close. To this end, the function S is extended to consider sets of trips. In other terms, $|t_i^l - t_j^l|$and are respectively replaced by the mean distances, i.e. $\frac{1}{n_k}\sum_{i=1}^{n_k}|t_i^l - t_j^l|$and $\frac{1}{n_k}\sum_{i=1}^{n_k} d(p_i^l, p_j^l)$ where $n_k$is the number of trips inside the cluster k. We then normalized these values because the acceptable delays are strongly related to the length of the trips. Consequently, the quality index of cluster k, $Q(k)$, is the function S applied to the set of clustered trips divided by the average length of the trips within cluster k.Notice that we aim at minimizing the quality index, i.e. best clusters present values of Q close to zero. Indeed, $Q(k)$ is low when (i) the spatio-temporal distance between origins is low, (ii) the spatio-temporal distance between destinations is low, and (iii) the mean travel distance is large. It comes that a cluster k is selected if and only if $Q(k)$ is below a specific threshold Q_{max}. In the remaining of the study, we set Q_{max} at 3. It corresponds to a restrictive matching policy.

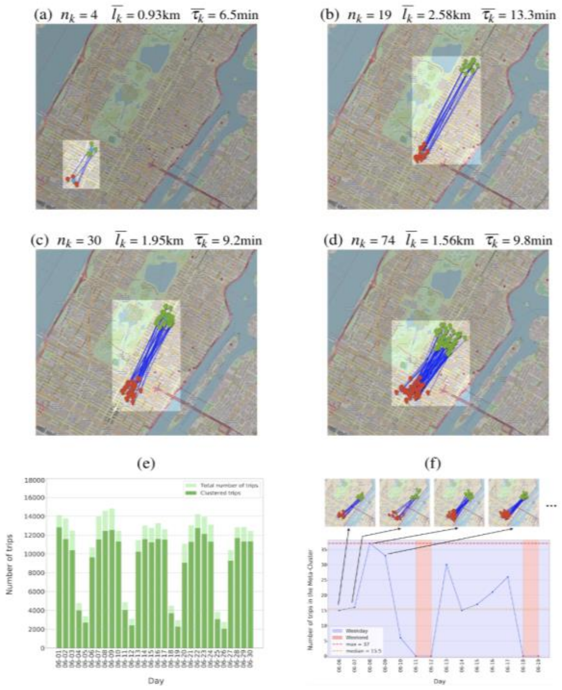


Figure 1: (a),(b),(c),(d) clusters with different characteristics, the pick-up are depicted in green and drop off in red. n_k denotes the number of trips in the cluster k, $\overline{l_k}$ denotes the average length of trips in $k$and $\overline{\tau_k}$ denotes the average duration of trips in k. (e) Ratio of clustered trips per day in Midtown and Upper East Side from 8h to 11h. (f) Example of demand graph for a randomly selected meta-cluster.

Clusters detected have different sizes, from 2 trips to 74 trips gathered into the same group. It brings to light that the shared mobility demand may take many aspects requiring different forms of transportation services to be optimally satisfied. Figure 1 depicts four clusters with different sizes and characteristics. The average travel length $\overline{l_k}$ is directly the arithmetic average of the length of n_k trips within the cluster k, whereas the average travel time $\overline{\tau_k}$ is the arithmetic average of the duration of the n_k trips. Figure 1.e shows the number of clustered trips and the total number of trips per day. The developed method detects almost 85%of similar trips per day on average in the studied zone.

2.3. Identification of regular demand pattern for multiple days

Once this daily analysis is done, we investigate if commonalities in the clusters can be identified. Many approaches exist to derive the most representative partition from a group of partitions, such as meta-clustering or consensus learning [15]. Here, we use the same clustering method to maintain consistency when scaling-up. In the following, to reduce the computational time, we focus the study on the 14 days of the dataset for which the ratio of clustered trips is the highest: June 6 to 19, 2011. The objective is now to find out if there are similar trips (relatively close departure and arrival locations and times) made several times during the studied period. These recurrent spatio-temporal areas are called meta-clusters. For that purpose, each cluster previously found is considered as a new trip, formed by the centroïd of its pick-up and the centroïd of its drop off. Centroïds correspond to the mean origin/destination locations and mean departure/arrival times of the clustered trips. This information can be useful to design the transportation supply because centroïds can be the locations of common meeting points of the standby areas of shared vehicles. A second clustering is then performed, it returns clusters with similar characteristics (without taking into account the initial day when the trips were made). In other words, two clusters are in the same meta-cluster, if their centroïds have close departure and arrival locations and times. Interestingly, we observe that more than 94% of the daily clusters are recurrent from one day to another.

The representation of a meta-cluster on a 2D map is difficult to analyze because the time dimension is not accounted for. Consequently, we prefer to focus on the evolution of the daily clusters' size and the localization of the related origin/destination whereabouts. Each meta-cluster can be depicted as a graph of the demand. Figure 1.f shows the graph of the demand for a randomly selected meta-cluster. This figure shows that in the same spatio-temporal area, similar trips can be seen every day, except on weekends. Each meta-cluster provides precise information about its location, its estimated departure and arrival times and the total number of trips performed per day. It is important to note that different individuals perform these trips from one day to another. However, global human mobility is remarkably regular; this is a valuable insight to tune transportation services and favor shared mobility efficiently.

3. CUSTOMIZED SUPPLY DESIGN

As mentioned in the previous section, the spatio-temporal areas containing similar and regular trips (meta-clusters) are detected. The study's next objective is to find a way to serve the pick-up and drop off points in each of these meta-cluster while respecting a set of constraints: time windows, vehicle capacity, size of the fleet, etc. A minimalist example of the developed method is depicted in Figure 2. Figure 2.a shows a set of 3 meta-clusters; each of them contains several clusters of similar trips. A green marker and a red marker linked by a blue line depict a cluster containing several similar trips. The green and red markers designate respectively the points of pick-up and drop off of a cluster. A meta-cluster is depicted by an aggregation of clusters in the same spatial area. On average, each cluster contains 6.22 similar trips. Moreover, a meta-cluster contains, on average 8.64 clusters. In other words, each meta-clusters contains on average 53trips with very similar characteristics (see Section 3). The method aims to design a line of transport serving a set of meta-clusters with characteristics compatible (size, time-windows, etc.). Depending on the meta-clusters chosen, the number and size of vehicles required will be different. To quantify the potential demand of a tour, we plot the total number of trips per day served by a tour going through these meta-clusters. Figure 2.c depicts the total number of trips served per day for the set of meta-clusters depicted in Figure 2.a

(DARP)

$$Minimize \sum_{k\in K}\sum_{i\in V}\sum_{j\in V} c_{ij}^k x_{ij}^k \quad (5)$$

subject to

$$\sum_{k\in K}\sum_{j\in V} x_{ij}^k = 1 \qquad (i\in P), \quad (6)$$

$$\sum_{i\in V} x_{0i}^k = \sum_{i\in V} x_{i,2_{n+1}}^k = 1 \qquad (k\in K), \quad (7)$$

$$\sum_{j\in V} x_{ij}^k - \sum_{j\in V} x_{n+i,j}^k = 0 \qquad (i\in P, k\in K), \quad (8)$$

$$\sum_{j\in V} x_{ji}^k - \sum_{j\in V} x_{ij}^k = 0 \qquad (i\in P\cup D, k\in K), \quad (9)$$

$$u_j^k \geq (u_i^k + d_i + t_{ij})x_{ij}^k \qquad (i,j\in V, k\in K), \quad (10)$$

$$w_j^k \geq (w_i^k + q_j)x_{ij}^k \qquad (i,j\in V, k\in K), \quad (11)$$

$$r_i^k \geq u_{n+i}^k - (u_i^k + d_i) \qquad (i\in P, k\in K), \quad (12)$$

$$u_{2n+1}^k - u_0^k \leq T_k \qquad (k\in K), \quad (13)$$

$$e_i \leq u_i^k \leq l_i \qquad (i\in V, k\in K), \quad (14)$$

$$t_{i,n+i} \leq r_i^k \leq L \qquad (i\in P, k\in K), \quad (15)$$

$$\max(0,q_i) \leq w_i^k \leq \max(Q_k, Q_k + q_i) \qquad (i\in V, k\in K), \quad (16)$$

$$x_{ij}^k = 0 \; or \; 1 \qquad (i,j\in V, k\in K), \quad (17)$$

Then, the selection of a set of meta-clusters to find a potential tour of vehicles is performed. We use the median number of trips per day in a meta-cluster as an indicator of its size. For each studied period, a minimal median value required is defined; we filter the meta-clusters with a median inferior to this value. This method allows us to obtain a reasonable number of meta-clusters to solve a relatively small instance of the optimization problem. However, it should be noted that the solution is optimal for each period on which we solve the problem, but an optimal result is not guaranteed for a set of periods. The meta clusters found are used to define a Dial-a-Ride Problem (DARP) instance. Given the large number of meta-clusters, it is impossible to directly model vehicle tours from the whole set of meta-clusters. That is why we focus this study only on high capacity vehicle tours. This method allows us to serve a large number of passengers by solving smaller problem instances. The main advantage of this method is that the calculation time depends on the number of meta-clusters served and not directly of the number of passengers. That is why it is crucial to find the largest possible meta-clusters. If we solve the problem for a set of large meta-clusters, the number of passengers effectively served will be significantly higher.

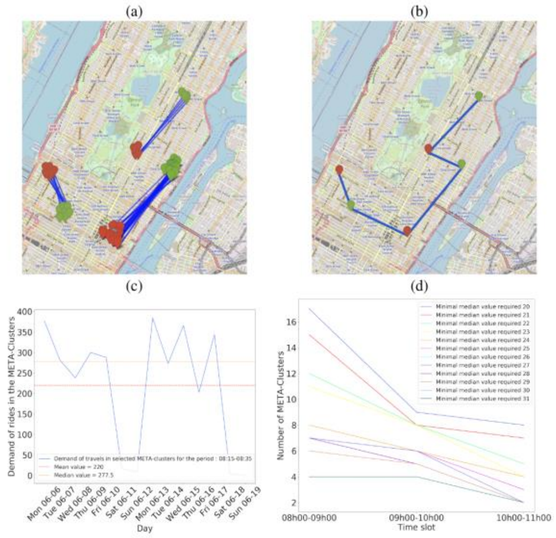


Figure 2: Green markers depict the pick-up and red markers the drop off (a) 3 meta-clusters randomly chosen between 08h15 and 08h35. (b) example of line designed, serving the centroids of pick-up and drop off of each meta-cluster. (c) Total number of trips per day served on the set of 3 meta-clusters. (d) Shows the number of meta-clusters for each time slot in function of the minimal median value of trips per day required.

There are many variants for the DARP problem, [16, 17] give a list of them based on different objective functions. In our case, we use a variant presented in [18]. This model is depicted below Eq.5 to Eq.17. The model is based on a three index formulation. Let $G = (V, A)$ a directed graph. The set of vertices V is partitioned as follow : the first and the last element are two copies of the depot, elements from index 1 to n are pick-up and elements from index $n + 1$ to $2n$ are drop off. P denotes the set of pick-up and D the set of drop off. A request is a couple $(i, n + i)$, where $i \in P$ and $n + i \in D$. The load of each vertex is defined as q_i, with $q_0 = q_{2n+1} = 0$, $q_i \geq 0$ for i in $\{1, \dots, n\}$ and $q_i = -q_{i-n}$ for i in $\{n + 1, \dots, 2n\}$. A service duration $d_i \geq 0$ with $d_0 = d_{2n+1} = 0$. K denotes the set of vehicles. The capacity of a vehicle $k \in K$ is v, and v denotes the maximal duration of a route for a vehicle k. The arc set is defined as: $A = \{(i, j) | i = 0, j \in P$ or $i, j \in P \cup D, i \neq j$ and $i \neq n + j$, or $i \in D, j = 2n + 1\}$ the cost of traversing an arc (i, j) with a vehicle k is c_{ij}^k, and the travel time between two nodes i and j is t_{ij}. L denotes the maximal ride time and the time window of a vertex i is $[e_i, l_i]$. x_{ij}^k is a binary variable equal to 1 if and only if (i, j) is traversed by a vehicle $k \in K$. Let u_i^k the time at which a vehicle k starts servicing a vertex i the load of vehicle k leaving vertex i, and r_i^k the ride time of user i.

This model presents several interesting aspects: multiple vehicles, time-windows for Pick-up, or Drop Off. The main objective of this method is minimizing the total route length. However, several other constraints can be added, such as vehicle capacity, maximum route duration, or maximum ride time for users. Nevertheless, it is essential to note that the meta-clusters previously found are independent of the method chosen to serve them and vice versa. Indeed depending on the objective searched, an approach may be more interesting that another. For example, it could be interesting to use a method to minimize the total route length for a Transportation Network Company. From a user point of view, using a technique to reduce the waiting time could be more interesting than another. Several methods aim to satisfy an objective function depicted as a combination of constraints such as transportation time, ride time, excess of maximum ride time, waiting time, time windows violations, etc. [19]. A comparison with these sophisticated methods will be studied in a future study.

4. RESULTS

This Section is devoted to the results of the proposed method for the case of NYC. First, the meta-clusters are presented and analyzed. Secondly, based on this demand decomposition, the optimization method is tested and evaluated.

4.1. Selection of the spatio-temporal areas

First of all, it is interesting to analyze the characteristics of the meta-clusters found. As mentioned in Section Method - Estimation of the demand, in the studied area between 08h00 and 11h00, almost 85% of trips can be considered similar. Moreover, more than 94% of trips are recurrent, i.e., these trips can be observed almost every day. 2136 spatio-temporal areas are detected as zones where there is a recurrent potential demand of shared mobility. On average, each meta-cluster contains 53 trips. Once again, it is important to notice that different users surely perform these trips. In the following, it is considered that the users' meeting point is defined as the centroïds of pick-up (respectively drop off) of a meta cluster. Thereby, it is interesting to know the spatial and temporal difference between the centroïds and the points of pick up and drop off. Table 1 shows that the spatial distances are close to 200m. The average temporal shifts are nearly 6 minutes which is entirely acceptable. It shows that the meta-clusters found are relatively close to the initial clusters estimated from the real rides of users. The average travel distance and time in the meta-clusters are respectively 1.71km and 11.1min. Although these data are not fully representative of human mobility since they only correspond to taxi trips, such a dataset provides an attractive proxy for studying the individuals' routes within a city.

Table 1: Average spatial and temporal distances between pick-up, drop off and the centroïds of the meta-clusters.

Variable	Result
Average spatial distance Pick-up / centroïd of Pick-up	0.21 km
Average spatial distance Drop off / centroïd of Drop off	0.21 km
Average shift between departure times / centroïd of departure times	6.16 min
Average shift between arrival times / centroïd of arrival times	6.27 min

In a first time, it is necessary to select a reasonable number of meta-clusters to solve a relatively low instance of the optimization problem. As said in Section 2, the median number of trips per day in a meta-cluster is used as an indicator of its size. Figure 2.d shows us for each one hour period the number of meta-clusters depending on the chosen minimal median value. In other words, a meta-clusters is counted if and only if its median value of trips per day is greater or equal to the chosen value. It is important to note that the number of points effectively treated in the DARP will be for each period $2 * numberofclusters$, because a vehicle serve a pick-up and a drop-off for each meta-cluster. According to Figure 2.d, the minimal median value 24 has been selected to find potential routes with a large number of users with very short execution times.

4.2. Demand-driven route optimization

According to the results showed in Section 3.1, three periods of one hour for which the meta-clusters contain at least a median of 24 trips per day are selected. Figure 3.a shows the potential number of trips per day that can be served on the period 08h00-11h00. This result illustrates one of the method's interests: the 18 meta-clusters selected represent actually 614 trips per day on average. The median number of trips served per day for this set of meta-clusters is 756. Also, we note that the demand is extremely regular every day of the week (except weekends) for the two weeks of analysis. In the case of an effective implementation of optimized lines, it would be interesting for the service to be operated from Monday to Friday.

The parameters used for DARP are adjusted for each period of one hour. Each centroïd of pick-up and drop-off of the meta-clusters are inserted in the sets P and D. The number of vehicles V for each period is depicted in Table 2. For each vehicle, its capacity $Q_k = 80$, which corresponds to the average capacity of a bus. For each arc (i, j), the cost c_{ij} is defined as the spatial distance between i and j. For each node i, we set the service duration $d_i = 2$ minutes. The load of each pick-up q_i is defined as the number of users to serve. The load of each drop off is defined as $-q_i$. A time window of 20 minutes is defined to serve the different points. This value is not representative of the real difference between the desired service times and the effective times of service, but it provides an upper and lower limit that should not be exceeded. If this value is not enough, the constraint will often be violated then; no solution will be found. The travel time between two nodes v and j is estimated according to the results presented is [10]. We set the average speed for a vehicle to $9.65 km/h$. Figure 3.a .b and .c show for each period the a map of the designed routes. Each color designates a specific transport line. Table 2 indicates the result of the DARP. For the three periods, routes allowing to serve all the selected meta-clusters in less than 9 seconds are found. This result proves that the method is a good way to design lines serving many users (more than 600 trips per day on average). Besides, for each period, we calculate the average delays and time advances for each point served by the optimized tour. This value is estimated by the difference between the wished times of departure and arrival (given by the centroïds of the meta-clusters) and the hour of service given by the solving of DARP. These values show that the developed method relatively little impact on demand. Indeed, on average, the delay is 12 minutes and the advance is 10.6 minutes, which is acceptable since the number of users served is high.

To the best of our knowledge, there are no classical optimization methods to find round serving such a quantity of similar and recurrent trips in such a tight timeframe. The theoretical studies on DARP [17] show that the exact method used in this paper can solve instance until 36 points. It would not be possible to solve instances with so many passengers without using an aggregation method in meta-clusters.

The existing dynamic methods such as [20,21,22] obtain trips delay between 2 and 6 minutes; however, these services work with large fleets of vehicles with limited capacities (between 2 and 10). Moreover, these methods work only on networks with a limited number of nodes.

Table 2 : Result of the search of rounds for the 3 time periods from 08h00 to 11h00.

	Period 1	Period 2	Period 3	**Total**
Time	08h00 - 09h00	09h00 - 10h00	10h00 - 11h00	08h00 - 11h00
Number of meta-clusters served	8	6	4	18
Number of vehicles required	3	2	1	6
Average delay (min)	16	11	9	12
Average advance (min)	9	12	11	10.6
Total travel distance	36.15 km	19.2 km	12.43 km	67.78 km
Computation Time	6.85 s	1.37 s	0.05 s	8.27 s

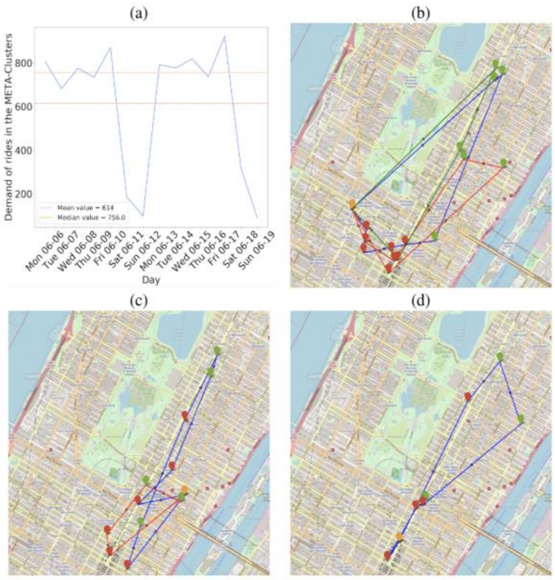


Figure 3: (a) depicts the total number of users effectively served in the set of meta-clusters selected in Section 3.1. (b),(c),(d) shows for each period presented in Table 2, the customized lines found.

5. CONCLUSION

This article presents an optimization method based on a decomposition of the demand and a resolution of DARP on a reduced instance. This data-driven method allows in a first time to identify clusters of similar and regular trips over time (meta-clusters). Then these meta-clusters are considered as points to serve in an instance of DARP. This method's main interest is to design tours of vehicles to serve a large number of potential users. As shown in Section 3, the main advantage of this method is that the execution time of the optimization problem does not depend on the number of users served, but only on the number of meta-clusters served.

As part of this study, the analysis of pattern recurrence was carried out over two weeks. However, the proposed method makes it possible to detect regular patterns over much more extended periods. This can be particularly useful in the case of effective implementation of transport lines. Besides, it is possible to integrate many other parameters into the objective function of DARP to design lines as close as possible to actual user demand.

The results obtained show us that rounds of large-capacity vehicles (80 peoples) can be identified. Making it possible to serve on average more than 600 trips per day with calculation times lower than 9 seconds. This hybrid method between classical and dynamic approaches allows to design high capacity lines based on the real demand of mobility. Moreover, it allows to obtain lines with restricted spatio-temporal deviations from the demand described by the centroïds of the meta-clusters. Setting up massive lines close to the initial demand of users provides a partial response to the last mile problem, which is one of the main obstacle to shifting users of private vehicles to shared modes of transport. As we mentioned in the introduction, this study is part of the adaptation of current methods to new requirements for the deployment of autonomous vehicles. The method presented makes it possible to overcome the incompatibilities between current methods and new approaches adapted to the needs of services based on the use of autonomous vehicles.

Several ways are studied in order to complete the current method. The first is to take real-time aspects into account in the method. This can be done in several ways, either with instant classification or by taking into account the results obtained to anticipate future demand. Another interesting aspect is the design of more or less dynamic lines according to the demand in a studied area. For example, depending on the number of users and the required responsiveness of the service, different solutions can be implemented: classic or dynamic bus lines, taxi, etc. Finally, taking into account the dynamic aspects of the network to choose routes according to the network's particular events: congestion, roadworks, etc. seems to be an excellent way to improve the current method. Finally, the method's scalability will be widely studied to maximize the number of data processed and thus the veracity of the results obtained.

REFERENCES

[1] Vuchic, V. R. and A. Musso, Theory and practice of metro network design. Public transport international, Vol. 3, 1991, p. 298.

[2] Carrese, S. and S. Gori, An urban bus network design procedure. In Transportation planning, Springer, 2002, pp. 177–195.

[3] Kocur, G. and C. Hendrickson, Design of local bus service with demand equilibration.Transportation Science, Vol. 16, No. 2, 1982, pp. 149–170.

[4] Agatz, N., A. Erera, M. Savelsbergh, and X. Wang, Optimization for dynamic ride-sharing: A review. European Journal of Operational Research, Vol. 223, No. 2, 2012, pp. 295–303.

[5] Najmi, A., D. Rey, and T. H. Rashidi, Novel dynamic formulations for real-time ride-sharing systems. Transportation Research Part E: Logistics and Transportation Review, Vol. 108, 2017, pp. 122–140.

[6] Zmud, J. P. and I. N. Sener, Towards an understanding of the travel behavior impact of autonomous vehicles. Transportation research procedia, Vol. 25, 2017, pp. 2500–2519.

[7] Fagnant, D. J. and K. M. Kockelman, Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. Transportation, Vol. 45, No. 1, 2018, pp.143–158.

[8] Bongiovanni, C., M. Kaspi, and N. Geroliminis, The electric autonomous dial-a-ride problem. Transportation Research Part B: Methodological, Vol. 122, 2019, pp. 436–456.

[9] Ferreira, N., J. Poco, H. T. Vo, J. Freire, and C. T. Silva, Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. IEEE transactions on visualization and computer graphics, Vol. 19, No. 12, 2013, pp. 2149–2158.

[10] NYC Department of Transportation, New York City Mobility Report. https://www1.nyc.gov/html/dot/downloads/pdf/mobility-report-singlepage-2019.pdf, 2019.

[11] Chiabaut, N. and C. Veve, Identifying Twin Travelers Using Ridesourcing Trip Data. Transport Findings, 2019.

[12] Ester, M., H.-P. Kriegel, and X. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, 1996, p. 6.

[13] Ester, M., H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd, 1996, Vol. 96, pp. 226–231.

[14] Veve, C. and N. Chiabaut, Estimation of the potential demand of shared mobility services through clustering of similar individual trips. Transportation Research Board 99th Annual meeting. No. 20-00168, 2020.

[15] Lopez, C., L. Leclercq, P. Krishnakumari, N. Chiabaut, and H. Van Lint, Revealing the day-to-day regularity of urban congestion patterns with 3D speed maps. Scientific Reports, Vol. 7, No. 1, 2017, p. 14029.

[16] Ho, S. C., W. Szeto, Y.-H. Kuo, J. M. Leung, M. Petering, and T. W. Tou, A survey of dial-a-ride problems: Literature review and recent developments. Transportation Research Part B: Methodological, Vol. 111, 2018, pp. 395–421.

[17] Cordeau, J.-F. and G. Laporte, The dial-a-ride problem: models and algorithms. Annals of operations research, Vol. 153, No. 1, 2007, pp. 29–46.

[18] Cordeau, J.-F., A branch-and-cut algorithm for the dial-a-ride problem. Operations Research, Vol. 54, No. 3, 2006, pp. 573–586.

[19] Jorgensen, R. M., J. Larsen, and K. B. Bergvinsdottir, Solving the dial-a-ride problem using genetic algorithms. Journal of the operational research society, Vol. 58, No. 10, 2007, pp.1321–1331.

[20] Santi, P., G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti, Quantifying the benefits of vehicle pooling with shareability networks. Proceedings of the National Academy of Sciences, Vol. 111, No. 37, 2014, pp. 13290–13294.

[21] van Engelen, M., O. Cats, H. Post, and K. Aardal, Enhancing flexible transport services with demand-anticipatory insertion heuristics. Transportation Research Part E: Logistics and Transportation Review, Vol. 110, 2018, pp. 110–121.

[22] Alonso-Mora, J., S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus, On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proceedings of the National Academy of Sciences, Vol. 114, No. 3, 2017, pp. 462–467.

CYCLIC LOAD TESTS OF RAILWAY SLEEPER-BALLAST INTERACTION

Laura Vovčko, MSc (Geol.)
Assist. Prof. Dr. Stanislav Lenart (Civ. Eng.)
Slovenian National Building and Civil Engineering Institute
Dimičeva ulica 12, Ljubljana, Slovenia
laura.vovcko@zag.si

ABSTRACT

To establish a sustainable mobility, railway transport infrastructure plays very important role. By implementing the environmental management system, railways are for the upholding the highest standards in environmental care. The demands for a clean and healthy environment are included in all aspects of different railway operations as well as in construction and maintenance activities at railway infrastructure. Durability of the railways is highly related to the type of railway track and the fragility of used material. This encourages us to find the optimum railway sleeper material which suits to technical requirements and provide the highest resistance to abrasion of ballast material in contact with railway sleeper.

This research presents the effect of material type of railway sleepers upon the degradation of ballast material during cyclic loading. Seven different compositions of railway sleeper and one type of ballast material were used in the research. Sleeper materials include wood and concrete, while the latter was improved by replacing coarse aggregates with rubber aggregates. Furthermore, under sleeper pads of geotextile and rubber layers were added. Each type of railway sleeper resting on ballast layer was loaded by one million load cycles of a maximum amplitude equal to the maximum rail traffic load. Besides deterioration of ballast material and damage of interected sleeper material, also permanent and resilient vertical displacements were observed. Structural (hysteresis) damping and stiffness characteristics vs. number of load cycles were evaluated. Their changes during different life-cycles of ballast track were observed.

Keywords:
Ballasted railway track, Railway sleeper, Cyclic loading, Ballast material, Deformation characteristics

1. INTRODUCTION

Durability of the railway track is highly related to the fragility of ballast material which is placed beneath the railway sleepers to keep them in place during the train passage. Ballast layer presents a significant part of the railway structure, its primary objective is to distribute the load through the sleepers to the ground beneath and to prevent their over stressing. To achieve a sufficient load capacity and durability of the railway structure, changes in ballast material properties during cyclic loading must be anticipated. Cyclic loading can namely cause deformation and degradation of ballast material which can lead to premature destruction of railway track quality and consequently more frequent maintenance performance to keep track functional. The predicitve maintenance can be significantly reduced with previous choosing the appropriate ballast material and the type of railway sleeper based on their behaviour under cyclic loading.

The most appropriate ballast materials generally consist of coarse-grained hard minerals (Raymond, 1985). Although the sedimentary and metamorphic rock types are in general weaker than igneous rock type, the properties cannot be unambiguously determined based only on the rock type. The formation mechanism of rocks namely presents a significant impact on the rock, respectively ballast material properties. According to Watters et al. (1987) the hardness of a rock is not depended only on the hardness of the component minerals but also on mineral cleavage fracture, textural attributes and in case of clastic sedimentary rocks on the degree of cementation or consolidation. Furthermore, since rocks are aggregates of many mineral grains or crystals, their properties depend of their various contained minerals. These properties are in general determined by averaging the relative properties and sometimes also orientations of the various grains or crystals. Many properties also depend on grain or crystal size, shape, packing arrangement, the temperature and pressure, etc. (Klein & Carmichael, 2020).

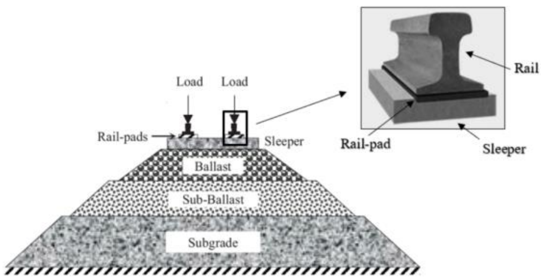
Because of the large range of factors that affect the mechanical properties of rock, the mechanical tests are considered as the most appropriate indicators for evaluating the individual ballast material properties and their deformation and degradation behaviour. In most countries the requirments for ballast materials are defined by limit values for the Los Angeles abrasion test and the micro-Deval test. Several researchers (Wright 1983; Selig & Boucher 1990) have shown that these conventional mechanical tests for ballast material do not reflect the actual field conditions during train loading. In Europe some studies of mechanical properties of ballast material and its behaviour in railway track under cyclic loading have been executed from 1970 to 1975. Since then, various types of experiments have been developed to evaluate the ballast properties. Regarding to Sadeghi et al. (2016) the characteristic of ballast material to insure better overall performance of the railway structure can be divided into mechanical, environmental, physical and geometry of ballast profile category. These categories are defined with establishing an estimation of the resistance to in track instability and degradation under loading. During loading the voids between larger aggregates can fill with finer particles resulting from the degradation of ballast grain fouling. The foulding process causes various changes in the gradation of ballast. According to Selig & Walters (1994) and Meeker (1990), the fouling can be caused by mechanical particle degradation during construction and maintenance work and traffic loading, with the chemical weathering related from environmental changes, migration of fine particles from the surface and the underlaying layers, the spillage from freight traffic and degradation of sleepers. In general, the awareness of changes that can lead to degradation of ballast material must be taken into consideration to avoid further maintenance difficulties.

While the cyclic loading can lead into the damage of railway sleepers or even the breakage of ballast layer, also the vibration response of sleepers under loading must be taken into the consideration in analysis and the railway structure design procedure. The vibration response to cyclic loading depends of a type of railway sleeper. Nowadays, composite railway sleepers present a significant role in the railway development. Several composite sleeper technologies have been developed to provide the alternative types of sleepers with ensuring sustainability from an environmental perspective. Based on the amount, length and orientation of fibres composite railway sleepers consist of three classifications (Ferdous et al., 2015). Researchers have shown that increase in rubber content in concrete railway sleeper leads to reduction in compressive strength. In general, the concrete sleepers have low tolerance for impact loading conditions (Remennikov & Kaewunruen, 2007). The reduction of compressive strength appeared at even very small consumption of rubber content in concrete. By adding the rubber into the concrete the ability to absorb the vibration energy of concrete can be improved (Meesit, 2017).

By increasing the demand to heavier traffic load and running speed a better insigne into the dynamic characteristic of railway sleepers and ballast material is crucial. The typical loading frequency of traffic loading in the track is normally around 8-10 Hz for a normal train and may reach 30 Hz for a high-speed train (Aursudkij et al., 2009). In this study, the effect of material composition of railway sleeper upon the degradation of the ballast material has been investigated under cyclic loading. Each type of railway sleeper installed on ballast layer was loaded by one million load cycles of maximum amplitude equal to the maximum railway traffic load. Beside deterioration of ballast material and damage of interacted ballast particles, also permanent and resilient vertical displacements were observed. Structural (hysteresis) damping and stiffness characteristics vs. number of load cycles were evaluated and their changes during different life-cycles of ballast track were observed.

2. RAILWAY STRUCTURE

A railway structure (Fig. 1) consists of rails resting on rail-pads which are laid on railway sleepers that are placed on top of a railway bed – ballast layer. The loads from the trains are applied by the vehicles onto the rails and the rail-pads are reducing their vibrations. Railway sleepers ensure certain amount of flexibility and elasticity to quickly return back to its original shape after the train passing. During the train passage the railway sleepers transfer the pressure from rails to railway bed which consists of ballast and sub-ballast layer above a subgrade.



Source: (Kumawat et al., 2019).

Figure 1: Railway structure.

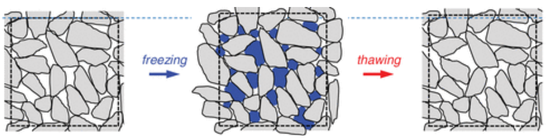
While the ballast layer is subject to deformation and degradation due to train passing, the ballast particles tend to breakdown or to develop internal cracking which leads to weakening of the railway bed. According to Selig and Waters (1994) and Ionescu (2004) one of the responsible factors for decrease of ballast efficiency is the contamination of ballast with finer particles.

Therefore, beside an appropriate design of railway structure (e.g., thickness of layers, size and spacing of the sleepers …), the properties of ballast material and the effect of pressure transfer from railway sleepers must be taken into consideration. Generally, physical properties of ballast material can be divided into physical properties of individual particle and physical properties of the whole ballast layer. Regarding to Indraratna et al. (2006) the first group includes petrological test, particle shape and surface examinations and durability tests, while the second group considers the permeability, void ratio, bulk density and specific gravity.

2.1. Ballast layer

2.1.1. Ballast material

Generally, ballast material is a free-draining granular material (Fig. 2) used as load-bearing material. It needs to be tough enough to resist breakage under external impact, hard enough to resist abrasion due to inter particle contact and dense enough to resist their lateral deformation. Furthermore, the ballast material must be also freeze-thaw resistant; therefore, voids which could be filled with water must be avoided. Voids filled with water can namely lead to inter particle volume change during freeze-thaw process. The mechanical degradation of ballast material produces finer particles that reduces the draining capacity of the railway track and can lead in internal pore water pressure development which can cause significant deterioration of railway tracks (Chanda & Krishna, 2003; Indraratna & Khabbaz et al. 2003, Ionescu, 2004).



Source: (Sang Yeob et al., 2021).

Figure 2: Freeze-thaw process.

The draining and compaction ability of ballast layer depends of their void ratio which is related by the particle size and gradation (Sadeghi et al., 2016). To reinsure appropriate void ratio, ballast material is generally composed of medium to coarse-sized particles (10-60 mm) with only small percentage of smaller particles (particles less than 20 mm). Ballast material should contain angular particles with high specific gravity, high shear strength characteristics, high toughness and hardness, high resistance to weathering, rough surface and minimum hairline cracks (Indraratna et al., 2007). Particle should consist of a hard and durable rock. The particles must sustain durable at external impact (e.g., atmospheric conditions, loading, …) which can cause their breakage. The characteristics of durability can be obtaining with aggregate crushing and wet attrition, as well as with wet strength and wet/dry strength variation and Los Angeles test. The ballast particle size distributions are very uniform, some well-graded distributions could namely reduce drainage capacity and increased fouling risk (Indraratna et al., 2007). Ballast material with well-graded distributions can be compacted to higher densities than uniform material. Indraratna et al. (2003) has shown the effect of particle size distribution on axial and volumetric strain behaviour during cyclic loading. They found out that the very uniform and uniform material exhibit higher axial and volumetric strain. Most of ballast initial properties changes under dynamic loading due to breakage, deformation and fouling. According to Indraratna et al. (2005) the aggregate degradation can be represented by the ballast breakage index (BBI) which employs the change in the fraction passing a range of sieve size and it is estimated from incorporating a linear particle size axis. The BBI can be define with equation 1, where p.s.d.c presents the particle size distribution curve and a.b. the arbitrary boundary.

$$BI = \frac{Area\ between\ the\ initial\ and\ final\ p.s.d.c.}{Area\ between\ the\ initial\ p.s.d.c.\ and\ the\ a.b.} \tag{1}$$

2.2. Railway sleeper

Railway sleepers (Fig. 3) were first made form timber; later the steel and concrete types of sleepers became more and more popular (Remennikov & Kaewunruen, 2006). During the last few decades, the railway industry has focused on a cement-based concrete rather than timber and steel sleepers. Mono-block prestressed concrete sleepers were first applied in 1943 and are now used in heavy haul and high-speed railway track constructions (Kaewunruen, 2010). In recent years, composite sleepers have been produced using new technologies. Different kinds of railway sleeper can be applied in different area and every type of railway sleeper has its own advantages and disadvantages.

Wooden railway sleeper is a traditional type of railway sleeper suitable for low-speed railway tracks. According to the types of wood, they can be divided into hardwood sleepers (oak, jarrah, etc.) and softwood sleepers (douglas, etc.). Wooden sleeper is light and therefore easy to transport, install and maintain. On the other hand, it can be easy affected by humidity and it is therefore is not durable.

Concrete railway sleeper is mostly made from pre-stressed concrete and usually used in the high-speed railway. Due to the prestressing, the damage form external pressure is reduced and therefore the prestressed sleeper has higher barring capacity. In comparison to wooden sleeper is has longer service life and it needs less maintenance. It also offers better stability due to its weight, but on the other hand the later presents also some difficulties regarding handling it. Due to its weight, it has also some limitation in application (e.g., bridges and crossing). Concrete sleepers are defined as mono-block or twin-block. In general, the concrete sleepers have damping ratio around 0,1 – 2,0 %, depending of support conditions and level of prestressing (Remennikov & Kaewunruen, 2006).

Steel railway sleeper is formed by pressed steel and with a trough-shaped section. It is stronger than wood and cheaper than concrete. Steel sleeper is easy to install and it is also easy to handle. It has a long service life and it is recyclable. The problem presents a chemical corrosion and high maintenance cost.

Composite railway sleeper is known also as plastic sleeper and it presents a mixture of plastic and used waste rubber. These type of sleepers combines the pliability of wooden and durability of concrete sleepers. Composite sleepers are good in reducing the vibration. They are cuttalbe and therefore easy to handle. Due their recyclable ability they can be reused into new sleepers. They have a long service life; the disadvantage presents their cost. According to Ferdous et al. (2015) the composite sleepers can be divided into following types:

Type-1 Composite railway sleepers can consist of recycled plastics (plastic bags, etc.) or also bitumen with fillers (gravel, sand, etc.). Their structure behaviour is mainly polymer driven and they offer a range of benefits including ease of drill and cut, good durability, consumption of waste materials, etc. However due low strength and stiffness they suffer from limited designs flexibility.

Type-2 Sleepers with long fibre reinforcement in the longitudinal direction and no or very short random fibre in the transverse direction. These sleepers are suitable for ballast railway track where the stresses in sleepers are governed by flexural and shear forces. These types of sleepers are easy to drill and cut, they have good durability, and flexural strength and modulus of elasticity are the advantages. Low shear strength and shear modules, limited design flexibility, marginal fire resistance and high price are some of the challenging issues associated with this type of category.

Type-3 Sleepers with fibre reinforcement in longitudinal and transverse directions. The disadvantages can present a non-ductile behaviour of glass fibre reinforced polymer sleeper and can present an issue when the sleepers are installed in bridge. They have excellent design flexibility, good flexural and shear strength, easy drilling and good fire performance. The production of these sleepers is quite slow and can increase manufacturing cost.

Source: (http://www.railway-fasteners.com/news/comparison-of-railway-sleepers.html)

Figure 3: Different types of railway sleepers; wooden railway sleeper (a), concrete railway sleeper (b), steel railway sleeper and composite railway sleeper (d).

3. TESTING METHOD

3.1. Ballast material

The ballast material used in this study was limestone, a prevalent Slovenian sediment rock. Material obtained from the quarry was firstly cleaned with water to remove dust and fine-grains adhering to the aggregates and then dried before the sieve analyse. Initial ballast particle distribution according to SIST EN 933-1:2012 and SIST EN ISO 17892-4:2017 has shown a uniform coarse-grained material composition. Particles had sharp angular edges.

3.2. Railway sleepers

In study, seven different types of railway sleeper were used (Table 1). Sleeper material included wood and concrete, while the latter was also improved by replacing coarse aggregates with rubber aggregates. Furthermore, under sleeper pads of geotextile and rubber layers were added (Fig. 4).

Table 1: Types of railway sleepers used in study.

Test no.	Type of railway sleeper		Test no.	Type of railway sleeper	
1	Wood				
2	Concrete		5	Concrete (10 % rubber)	
3	Concrete with rubber pad		6	Concrete (10 % rubber) with rubber pad	
4	Concrete with geotextile pad		7	Concrete (10 % rubber) with geotextile pad	

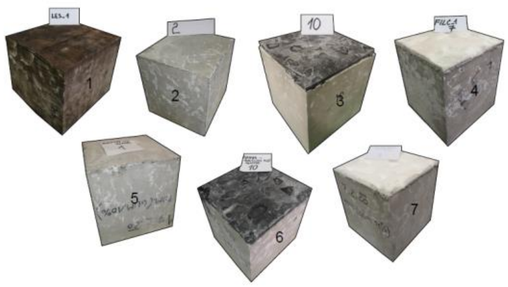


Figure 4: Types of railway sleepers used in study.

4. TEST PERFORMANCE

The effect of material composition of the railway sleepers upon the ballast material during cyclic loading has been evaluated (Fig. 5). Therefore, the initial ballast material was placed into the steel chamber (50 cm x 50 cm x 30 cm) in a combination with individual type of sleeper. For the test each 10-cm ballast layer was at first weighted and then dropped in the chamber box. The whole box was filled with 145 kg of dry ballast material; the dry density of 1933 kN/m3.

The cyclic load was applied to the rail-sleeper by hydraulic actuator through a cylindrical steel ram. Each type of railway sleeper resting on ballast layer was loaded by one million load cycles of maximum amplitude equal to the maximum railway traffic axle load of 250 kN. Beside deterioration of ballast material also vertical displacements were measured. Structural (hysteresis) damping and stiffness characteristics vs. number of load cycles were evaluated and their changes during different life-cycles of ballast track were evaluated. After the cyclic loading, the ballast was removed and the ballast particles were sieved again according to SIST EN 933-1:2012 and SIST EN ISO 17892-4:2017. For each individual type of sleeper three tests were performed.

Figure 5: Ballast material during cyclic loading.

5. RESULTS

5.1. Grain size distribution

According to SIST EN 933-1:2012 and SIST EN ISO 17892-4:2017 the sieve analyses was made for initial ballast material (IBM) before loading and for the same material after the cyclic loading (BM) for all types of sleepers. Particle size distribution curves presents relation between cumulative percentages passing vs. sieve size. The coefficients of uniformity (C_u) and the ballast breakage index according to Indraratna et al., (2005) were quantified. The breakage potential is considered to be the area between the particle size distribution and an arbitrary reference line connecting the point between the intersection of d_{95} of the largest sieve size and the minimum particle size of 2.40 mm. Obtained characterisitcs are summarized and presented in Table 2 and Fig. 6.

Table 2: Characterisics of initial ballast material (IBM) and ballast material upon each individual sleeper type after the cyclic loading (BM).

No.	Composition	C_u [-]	BBI [-]
0	Initial ballast material = IBM	1.54	-
1	IBM + Wood = BM $_{Wood}$	1.55	0.1595
2	IBM + Concrete = BM $_{Concrete}$	1.60	0.2317
3	IBM + Concrete with rubber pad = BM $_{Concrete + rubber pad}$	1.57	0.1754
4	IBM + Concrete with geotextile pad = BM $_{Concrete + geotextile pad}$	1.57	0.1469
5	IBM + Concrete (10 % rubber) = BM $_{Concrete (10 \% rubber)}$	1.56	0.1279
6	IBM + Concrete (10 % rubber) with rubber pad = BM $_{Concrete (10 \% rubber) + rubber pad}$	1.56	0.1164
7	IBM + Concrete (10 % rubber) with geotextile pad = BM $_{Concrete (10 \% rubber) + geotextile pad}$	1.56	0.1192

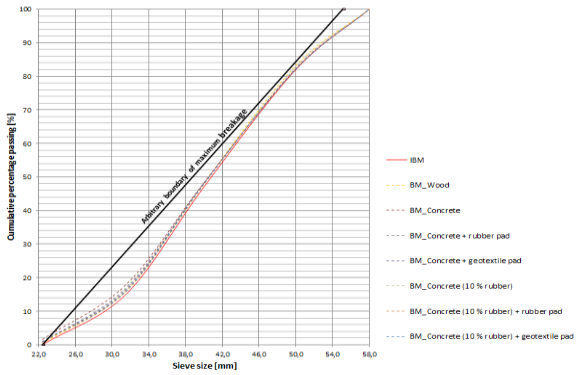


Figure 6: Particle size distirbution curves for intact ballast material and ballast material for each individual sleeper material after the cyclic loading.

5.2. Cyclic loading

Each type of railway sleeper resting on ballast layer was loaded by one million load cycles of maximum amplitude equal to the maximum railway traffic load, corresponding to axel load of 250 kN. Regarding the area of testing part of a sleeper, a sinusoidal cyclic load was applied with an amplitude of 14 kN. During the test a vertical displacement was measured and the relations between applied force and vertical displacement, they are presented in Fig. 7 and 8.

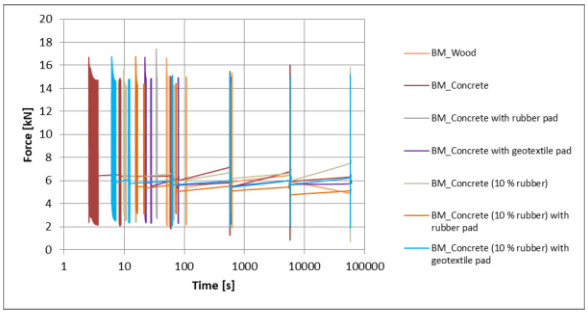


Figure 7: Force-time curves.

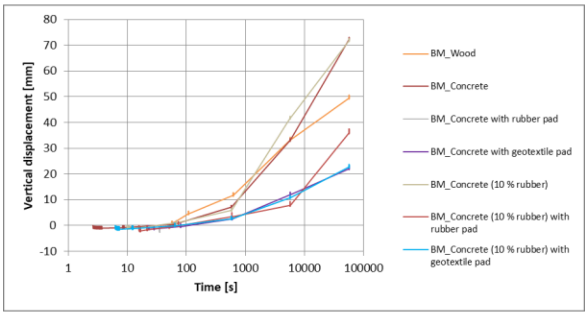


Figure 8: Vertical displacement-time curves.

6. ANALYSIS OF RESULTS

According to SIST EN 933-1:2012 and SIST EN ISO 17892-4:2017, the initial ballast material was classified as uniform coarse-grained material with a maximum particle size of 58 mm. It is evident, that the particle distribution curves of ballast material tested upon different types of railway sleepers lies above the distribution curve of initial ballast material (Fig. 6). Cyclic loading namely causes an increase of particle breakages and thus the shift of particle size distribution line. Changes in particle size distribution are evident in coefficients of uniformity (C_u) and ballast breakage index (BBI) as presented in Fig. 9.

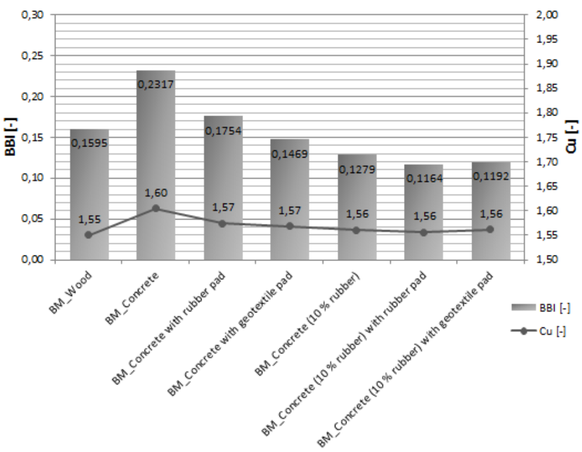


Figure 9: Coefficients of uniformity (C_u) and ballast breakage index (BBI) of ballast material upon different type of railway sleeper after the cyclic loading.

Structural (hysteresis) damping and stiffness characteristics vs. number of load cycles were evaluated and their changes during different stages of service life of ballast track, namely after different number of applied load cycles were analysed. According to Sas and Gluchowski (2012), the resilient module (M_r) is evaluated by dividing the axial stress (σ_d) by the resilient strain (ε_a). The axial stress was obtained by applied force and cross-sections area of the sample, while the resilient strain was obtained by resilient displacements in one cycle with the initial height of the sample. The relation between axial stress versus resilient strain and obtained resilient modulus at each cyclic number are presented in Fig. 10 and 11.

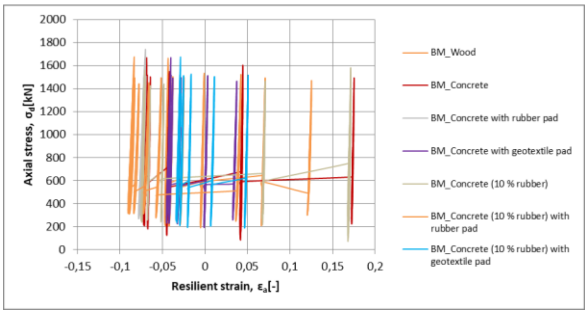


Figure 10: Cyclic loading curves.

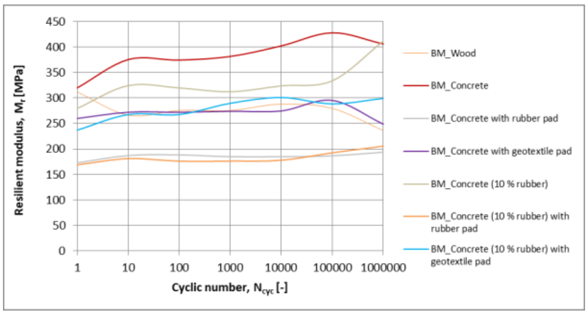


Figure 11: Resilient modulus-cyclic number curves.

Generally, a slight increase in modulus with an increase of load cycles number is observed (Fig. 11). At the beginning of each stress level, the resilient modulus becomes stable in 10 loading cycles. Afterward, the cyclic loading lids to produce a resilient vibration that results in a stable deformation. The resilient behaviour is mainly governed by the contacts of ballast particles. The highest resilent modules is due its low ability to absorb vibration energy evident at concrete railway sleeper an it lowers by adding pads of geotextile or rubber under it. Within 100 and 100 000 loading cycles some small increases and decreases of resilient modules are noticeable at test were concrete railway sleepers with added rubber aggregates and concrete with pads of rubber were used. Small changes of their resilient modules were probably caused by some negligible reorientation and particle brakeage which led to more stable reorientation that was highly resistant even at additional 1000 0000 loading cycles. At the high degree of cyclic loading the particle degradation can be developed again. The latter is evident as a decrease in resilient module at 1000 0000 loading cycles at test that included concrete, concrete with added pads of geotextile and at wooden railway sleepers. Small material degradation when using the concrete railway sleepers with added rubber aggregates indicated also low BBI value in Fig. 9, while at tests with other types of railway sleeper have higher BBI values.

7. CONCLUSION

The performance of ballast track under various types of sleepers was evaluated with this study. Laboratory test results indicate that the particle breakage depends on the type of railway sleeper. According to the particle breakage indexes (BBI), the addition of rubber aggregates to the concrete railway sleeper lowers the ballast material degradation. Particle breakage index also indicates that adding pads of geotextile or rubber under the concrete railway sleeper without rubber aggregates within concrete lowers the ballast degradation, while pads under the sleeper made of concrete with added rubber aggregates has no significant impact upon the ballast degradation.

The cyclic loading tests conducted in the present study could help to understand the long-term response of a railway track under different types of railway sleepers. We can conclude that the resilient modulus became stable after the certain number of loading cycles. The largest deformations are therefore evident at an early stage of cyclic load, in our study within 10 loading cycles. At this stage the granular material is probably influenced by compaction. Resilient modules are the highest by using the concrete railway sleeper and get lower by adding pads of geotextile or rubber under it. Small material degradation by using the concrete railway sleepers with added rubber aggregates due its ability in compere with other types of railway sleepers are also evident from BBI values.

REFERENCES

[1] Aursudkij, B., McDowell, G. R., & Collop, A. (2009). Cyclic loading of railway ballast under triaxial conditions and in a railway test facility. Granular Matter, 11(6), 391-401.

[2] Chanda, M., & Krishna, R. (2003). Selection and testing of ballast stones for underground railway tracks. African Journal of Science and Technology, 4(2), 42-50.

[3] Ferdous, W., Manalo, A., Van Erp, G., Aravinthan, T., Kaewunruen, S., & Remennikov, A. M. (2015). Composite Railway Sleepers – Recent developments, challenges and future prospects. Composite Structures, 134, 158-168.

[4] Indraratna, B. H., Khabbaz, H., & Lackenby, J. (2003). Behaviour of railway ballast under dynamic loads based on large-scale triaxial testing. AusRAIL PLUS (pp. 17-19). Sydney, Australia.

[5] Indraratna, B. H., Lackenby, J., & Christie, J. (2005). Effect of confining pressure on the

degradation of ballast under cyclic loading. Geotechnique, 55(4), 325-328.

[6] Indraratna, B., Khabbaz, H., Salim, W., & Christie, D. (2006). Geotechnical properties of ballast and the role of geosynthetics in rail track stabilisation. Ground Improvement, 10(3), 91-101.

[7] Indraratna, B., Shahin, & M. A., Salim, W. (2007). Stabillisation of granular media and formation soil using geosynthetics with special reference to railway engineering. Ground Improvement, 11(1), 27-43.

[8] Ionescu, D. (2004). Ballast Degradation and Measurment of Ballast Fouldin. 7th Railway Engineering Conference (pp. 169-180). London, UK: Commonwealth Institute.

[9] Kaewunruen, S. (2010). Sleepers and fastening. In: Track design fundamentals, Rail Corporation.

[10] Klein, C., & Carmichael, R. S. (2020). Encyclopedia Britannica. https://www.britannica.com/science/rock-geology.

[11] Kumawat, A., Raychowdhury P. & Chandra, S. (2019). Frequency-Dependent Analytical Model for Ballasted Rail-Track Systems Subjected to Moving Load. International Journal of Geomechanics, 19(4).

[12] Meeker, L. E. (1990). Engineering and economic factors affecting the cost of railroad ballast-MSc. Thesis. Reno, U.S.

[13] Meesit, R. (2017). Vibration Characteristics of Micro-Engineered Crumb Rubber Concrete for Railway Sleeper Applications. Journal of Advanced Concrete Technology, 15(2), 55-66.

[14] Raymond, G. P. (1985) Research on Railroad Ballast Specification and Evaluation. Transportation Research Record, 1006, 1-8.

[15] Remennikov, A. M., & Kaewunruen, S. (2006). Experimental Investigation on Dynamic Railway Sleeper/Ballast Interation. Experimental Mechanics, 46, 57-66.

[16] Remennikov, A. M., & Kaewunruen, S. (2007). A review of loading conditions for railway track structures due to train and track vertical interaction. Structural Control and Health Monitoring, 15(2), 207-234.

[17] Sadeghi, J. M., Ali Zakeri, J., & Emad Motieyan Najar, M. (2016). Developing Track Ballast Characteristics Guideline in Order to Evaluate Its Performance. International Journal of Railway, 9(2), 27-35.

[18] Sang Yeob, K., Junghee, P., Wonjun, C., Jong-Sub, L., & Santamarina, C. (2021). Soil Response during Globally Drained and Undrained Freeze-Thaw Cycles under Deviatoric Loading. Journal of Geotechnical and Geoenvironmental Engineering, 147(2), 14-31.

[19] Selig, E. T., & Boucher, D. L. (1990). Abrasion Tests for Railroad Ballast. American Society for Testing and Materials. Geotechnical Testing Journal, 13(4), 301-311.

[20] Selig, E. T., & Waters, J. M. (1994). Track Geotechnology and Substructure Management. London: Thomas Telford Publishing.

[21] Watters, B. R., Klassen, M. J., & Clifton, A. W. (1987). Evaluation of Ballast Materials Using Petrographic Criteria. Transportation Research Record, 1131, 45-63.

[22] Wright, S. E. (1983). Damage caused to ballast by mechanical maintenance techniques. British Rail Research Record, 16, 174-180.

VALUABLE INVESTMENTS IN ENERGY INFRASTRUCTURE TOWARDS SOCIOECONOMIC FOOTPRINT

Panagiotis Zeibekis, Researcher
Melina Karagkouni, Researcher
Dimitrios Dimitriou, Professor Associate
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece,
pzeimpekis@yahoo.gr, melkaragouni@gmail.com, ddimitri@econ.duth.gr

Maria Sartzetaki, Senior researcher/Adjusted lecturer
Department of Economics, Democritus University of Thrace
Panepistimioupoli, Komotini, Greece,
University Neapolis, Pafos, Cyprus
msartze@econ.duth.gr

ABSTRACT

The complexity of decision-making for investments in energy sector addressed to a variety of stakeholders with different expectations and to many business and financial risks and uncertainties. This especially applies to decisions regarding the implementation of capital-intensive energy projects where large amounts of capital are reserved. There are many cases where decision-makers do not clearly distinguish between the investment likelihood alternative scenarios and the socioeconomic footprint of the project. This paper deals with an assement framework to evaluate the socioeconomic footprint of energy projects in terms of economic, social, and environmental issues affected the decision process and the project development strategy. By a system of system approach key cost and benefits variables are highlighted and a series of key performance indicators are presented. The case study focus on natural gas facilities in Thrace Prefecture in North Greece.

Keywords:
Valuable investments, energy infrastructure, socioeconomic footprint

1. INTRODUCTION

As the need for a more sustainable and decarbonized future is continuously growing [1] gas has become the fastest growing primary energy source in the world, due to its environmental friendly nature and its multiple uses across a number of sectors. In Europe, the demand for natural gas is rapidly increasing, reaching 538 bcm in 2018. While the demand is growing so fast, the necessity for more expanded and well-developed natural gas distribution infrastructure is rising as well. Currently, Europe has a quite large natural gas distribution infrastructure, where countries such as Belgium, France, Germany, Norway, Italy the Netherlands and the UK have the best-developed distribution network. However, Russia and Norway remain the main natural gas suppliers to the remain European countries, since combined they provide almost 2/3 of the natural gas supplied to the EU [2] in contrast to Portugal, Greece and Northern Ireland, that began to develop their NG distribution network during the latter half of the 1900s.

Natural gas was brought in Greece during 1997, therefore its distribution networks are still not fully developed. Currently, its National Natural Gas Distribution System consists of the main natural gas distribution system, which spreads through transmission pipelines and branches across the biggest part of the country, the Trans Adriatic Pipeline (TAP), an 878 kg long pipeline, that will transfer natural gas to Europe from the Caspian region through Greece, Albania and Italy and also the under construction IGB and IGI distribution pipelines, which will be natural gas interconnector spreading from Bulgaria to Komotini Greece and from Turkish – Greek boarder to Italy, respectively.

This paper deals with the sustainability assessment and evaluation of the feasibility of Natural Gas distribution facilities project. The application of the proposed methodology framework is the

Alexandroupolis Independent Natural Gas System (INGS) project.

2. METHODOLOGY FRAMEWORK

Sustainability has become a key priority for every operational process and has mainly three pillars: economic, environmental and social as shown in figure 1. Investments in energy infractruture have an enormous impact in the three sustainability pillars mentioned above. More specifically, natural gas is responsible for various environmental implications.Even though natural gas emits approximately half as much CO_2 as coal when combusted, there have been various considerations lately regarding its upstream emissions due to variations in production basins [6]. However, natural gas has been gaining penetration due to the lower pollutants of conventional energy sources and improved efficiency.Ultimately, the environmental impacts of natural gas operations are depedant upon the life cycle methane leakages rates, in combination with other factors regarding subjective policy considerations.

Also, there are several social considerations regarding natural gas distribution processes. The construction and operation of an FSRU station has the potential to bring significant advantages on a social level to the country. First of all, it is a source of producing energy with low carbon emissions, which effects positively not only the environment but also the the health of the people located near the plant. Moreover, it can contribute to increase significantly the employment rates of the community, since an adequite workforce is essential in driving and running the sustainable FSRU station. The availability of enough workers on site and enough working time is essential for the plant to be operated properly.

The feasibility evaluation of the Natural Gas distribution project in Alexandroupolis is assessed based on these three pillars, giving significant attention to social and economic and includes a variety of different factors that are assessed in order to obtain a conclusion on the sustainability of this project.

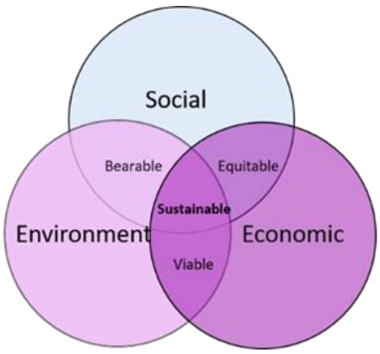


Figure 1: Sustainability Pillars

2.1. Economic Assessment Framework

A fundamental part of evaluating the feasibility of the project is conducting an economic assessment, which aims to examine the expenses and the costs occurring during the development of the project. After determining the different costs and expenses, Payback Time and ROI have to be calculated. Payback time is the time it takes a project to repay its initial investments and it is an effective criterion in evaluating the economic viability of a project. Therefore, the smaller the payback time is the quicker the project will generate a profit and will awake the interest of various investors. This payback time method assumes that all the investment is made in year zero and revenues begin immediately. Therefore, it may be assess whether a project can generate a profit or not, but it can be a trusted criterion of its profitability. Return on Investment (ROI) is a performance measure used to evaluate the efficiency of one or several investments and evaluate the financial consequences of investments and actions. When a ROI turns out to be negative that suggests that the total cost overcomes the return and so the project will generate a loss instead of a profit and an investment should not be made. On the other hand, when a ROI is positive, then the project is a success and the bigger the ROI is, the more probable is for a company to produce a profit [4,5,7].

2.2. Balance scorecard

Performance indicators support the management of infrastructures. As no single performance indicator can give a full picture regarding infrastructure performance, each indicator presents a partial view from a specific viewpoint and is therefore not enough to serve as a basis for management decisions. A popular performance measurement scheme suggested by [8] is the Balanced Scorecard (BSC) that developed to measure performance metrics from financial, customer, internal processes and growth perspectives. By combining these different perspectives, BSC helps decision makers to understand the inter-relationships and trade-offs between alternative performance dimensions, thus leading to improved decision making and problem solving. The base point of BSC would be the identification of the unit's strategic plan. This would involve the development of a goals, strategy, outputs, measures, targets and four different financial perspectives [8]

3. APPLICATION

3.1. Overview of the project

This project comprises an offshore floating unit for the reception, storage and re-gasification of LNG and a system of a subsea and an onshore gas transmission pipeline through which the natural gas is shipped into the Greek National Natural Gas

System (NNGS) and onwards to the final consumers. The Alexandroupolis INGS has also the capacity to connect with and transmit gas into other gas transmission systems such as TAP . This assessment framework will be based on the evaluation of the three economic, social and environmental Pillar [9,10].

3.2. Greece NG submission system

Natural Gas is, along with the petroleum products, the second biggest energy source used in the 21st century while the natural gas market of Greece, the total consumption of natural gas for 2017 was 4,9 billion cubic meters (bcm), while the prediction for 2028 is to be increased in 6,2 bcm. Greece is fully dependent in natural gas imports, since consumes in a year 5.0 times more than its total reserves. It substains this consumption by importing 175,727 MMcf of natural gas per year (in 2017) [11]. The entrance points are three: the entrance point of Sidirokastro in Greece to Bulgaria borders, the entrance point of Kipoi in Greece to Turkey borders and the Revithousa LNG land-based terminal. In 2017, the percentages of natural gas supply participation for the three entrace points mentioned above are 58,7%, 12,4% and 28,9%, respectively.

3.3. The Alexandroupolis FSRU Station

The Alexandroupolis FSRU Station has been decided to be the 4th entrance point of Greek's natural gas market. The capacity of its storages will be 170.000 cubic meters (cm), while the regasification capability will be 6, 1 bcm per year. The main component of the Project will be the offshore Floating Storage and Regasification Unit LNG vessel, with 300m length, 32,5m breadth and 26,5m height, which will be anchored at a fixed location at a distance of 17,6km southwest from the port of Alexandroupolis. The Project will also consist of the permanent offshore installations and the subsea and onshore sections of the gas transmission pipelines, which will be 24km and 4km long respectively. The pipeline of the FSRU station will be connected to the Greek natural gas transmission system in Amfitriti Station.

The FSRU Station is a cornerstone for the establishment of a gas hub in the area, since it will supply the Greek gas market, the Bulgarian and Balkan gas market via the IGB pipeline and the Italian market through the TAP pipeline.

3.4. Analysis results

The capital expenditures used for both calculations were taken from the study conducted by Brian Songhurst for Oxford University: “The Outlook for Floating Storage and Regasification Units (FSRUs)” [12]. The capital expenditures for FSRU stations construction are shown in Table 1.

Table 1: Capital expenditures for FSRU Station construction

	Cost (mil €)
New FSRU Vessel Cost	223,3
Infrastructure cost	98,25
CAPEX	321,55
Contigency (10% of CAPEX)	32,15
Owner's Cost	48,25
TOTAL CAPEX	401,95

The operational expenses were taken from the September 2016 press kit of the Toscana FSRU Station in Italy and Wartsila official website and were estimated at 38,58 mil € per year. The charge rates (€/cubic meter) were taken from the study of Brian Soghurst [12] and the official website of Toscana FSRU Station and are presented at Table 2.

Table 2: Charge rates of FSRU Stations around the world, (for all the above calculations we applied the dollar – euro exchange rate of 1st May 2019 (0,8929 $/€).

FSRU Station	Charge rate (€/cm)
Indonesia	0,034
Lithuania	0,026
Chile	0,022
Italy	0,039

The benefits of the operation of FSRU station were studied both at national and local level. Primarily, at national level the station will contribute to the Greek gas market security, becoming the fourth entrance point of Greek national gas transmission system. With Turkey stopping the supply of natural gas through the pipeline from Bulgaria at the end of 2019, 80% of natural gas will come to Greece from pipelines that cross Turkey. Therefore, the necessity of an extra entrance point is extremely high. Also, it will strengthen the geostrategic position of Greece in the Balkan area. The FSRU station will feed the Bulgarian gas market through the IGB pipeline and feed gas to other countries as well, such as Serbia, North Macedonia, Romania and Hungary.

On the other hand, at a local level the benefits would be 120 new job positions and supply of services from local companies, that are expected to cost at around 20 mil € per year [13]. Increase at the earnings of local companies from the professional tourism. The economic impact of the professional tourism is estimated to be from 187.500€ to 457.500€ per year. Last but not least, the FSRU station will have positive impact to the environment, due to the lower carbon and sulphur emissions of natural gas compared to diesel and coal. The project shall also have zero effect on the sensitive ecosystem of the Evros river delta and shall be located far from any Natura 2000 area. To ensure this, systematic monitoring for marine, physical, biological and eco-toxicological surveys shall be carried out.

4. CONCLUSIONS

Access to energy resources, energy supply security, insufficient investment in energy distribution infrastructure and sluggish progress in mitigating CO2 emissions are wide recognized in the market. The evaluation of the efficiency of energy distribution project should analyses the impacts in above issues and in literature there is a lack of publication regarding multidimensional assessment. This paper present the outline of methodology and assessment framework on evaluation the feasibility of an energy distribution system by adopting a system of system approach.

While most economics accept the importance of sustainable development and social justice within and between generations, they also point to historical experience suggesting that decisions about new infrastructure should meet the social targets toward welfare. On the investors side energy market prices are key driver for efficient energy facility projects as in the concept most of them are capital intensive projects meaning some times the payback period is too long. Risk assessment assumptions and energy market forecasting its crusinal in the feasibility study and this paper provides some key highlights in this area.

REFERENCES

[1] Dester, Mauricio & Andrade, Moacyr & Bajay, Sergio. (2012). New Renewable Energy Sources for Electric Power Generation in Brazil. Energy Sources, Part B: Economics, Planning, and Policy. 7. 390-397. 10.1080/15567249.2011.623502.

[2] Dediu, D., Czajkowski, M. and Janiszewska-Kiewra, E., 2019. How did the European natural gas market evolve in 2018?. [online] McKinsey & Company. Available at: <https://www.mckinsey.com/industries/oil-and-gas/our-insights/petroleum-blog/how-did-the-european-natural-gas-market-evolve-in-2018> [Accessed 25 April 2019].

[3] Dimitriou D., Sartzetaki M. et al. (2020). Chapter 1: Mobility as A Service: concept, objectives, actors, models and development, Book title: Mobility as a Service, Transport Trends and Economics 2018–2019, United Nations, UNECE, ISBN: 978-92-1-117238-6, pp. 9-13.

[4] Dimitriou D. and Sartzetaki M., Chapter title: Social dimension of aviation on sustainable development, Book title: Sustainable aviation - Greening the Flightpath, Palgrave Macmillan, pp. 173-191

[5] Dimitriou D., Sartzetaki M. (2020). Assessment framework to develop and manage regional intermodal transport network, Journal: Research in Transportation Business & Management, Elsevier, https://doi.org/10.1016/j.rtbm.2020.100455

[6] Draucker L., Bradbury J., Obeiter M., Wang W., Stevens A. (2013). Clearing the air: Reducing upstream greenhouse gas emissions from U.S. natural gas systems, World Resources Institute.

[7] Sartzetaki M. (2019). Appraisal of capital-intensive investments upon effectiveness to business ecosystem, Journal of Economics and Business, Vol.2, No.3, pp. 900-909.

[8] Kaplan, R. and Norton, D. (1992) The Balanced Scorecard—Measures That Drive Performance. Harvard Business Review, 79.

[9] Dimitriou D., Sartzetaki M., (2016). Sustainable Development Variables to Assess Transport Infrastructure in Remote Destinations, International Journal of Civil and Environmental Engineering, Vol.10(10), pp 1314-1321.

[10] Sartzetaki M. (2019). Attribute of the Effects Diversification to Eco-System from Large Investments in Supply Chain, Global Journal of Management and Business Research, Volume 19 Issue 5.

[11] Worldmeter, 2017. https://www.worldometers.info/ (accessed, February 15, 2021)

[12] Songhurst, B. (2018) LNG Plant Cost Reduction 2014 –18. Available from https://www.oxfordenergy.org/wpcms/wp-content/uploads/2018/10/LNG-Plant-Cost-Reduction-2014–18-NG137.pdf. Accessed on 5 August 2020.

[13] Toscana, 2016. https://www.oltoffshore.it/en/terminal/location/