Advances in Production Engineering & Management
Volume 20 | Number 1 | March 2025 | pp 29–42
https://doi.org/10.14743/apem2025.1.525
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Closed-loop simulation of workload control: Integrating input-output regulation with feedback

Oliveira, A.F.[a], **Sagawa, J.K.**[a], **Mušič, G.**[b,*]

[a]Federal University of São Carlos, Department of Production Engineering, Rodovia Washington Luís, São Carlos, Brazil
[b]University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

ABSTRACT

When responsiveness to demand fluctuations is a key performance factor, continuous dynamic models can advantageously replace discrete event simulation, often employed in Workload Control (WLC) studies. Nonetheless, dynamic modelling tools and feedback control are scarcely applied to WLC or other shop floor control systems. To fill this gap, this paper presents a closed-loop model of WLC incorporating feedback control and shop floor input-output control. The bond graphs' dynamic modelling technique is employed. The model is implemented in Simulink® and its behaviour in face of disturbances is analysed. The load of the machines and of the job pool of WLC is considered to adjust order release (input control). The capacity of each machine (output control) is altered in function of the level of its preceding buffer. The machines' processing rates stabilize and the reference levels for the buffers are reached when a step disturbance in order entry is simulated. Also, the system responded with a maximum capacity increment of 15 % when cyclic demand is simulated. This novel approach of a smart production control system can help managers to better control shop floor load in response to disturbances.

ARTICLE INFO

Keywords:
Workload control;
Closed-loop systems;
Simulation;
Smart production systems;
Control theory;
Bond graphs;
Input-output control;
Shop floor

**Corresponding author:*
gasper.music@fe.uni-lj.si
(Mušič, G.)

Article history:
Received 14 January 2024
Revised 29 March 2025
Accepted 31 March 2025

1. Introduction

The Production Planning and Control (PPC) can be seen as a determinant factor to the performance of an organization, thus, the choice of adequate PPC Systems remains as a key point. Some classic PPC approaches cannot be well applied to production systems that are not based on demand forecasts, like make-to-order (MTO) systems, due to the fact that these approaches focus on large systems within a repetitive production environment [1]. The Workload Control (WLC) is an alternative to MTO systems with a job shop production flow. It balances the load of the workstations, allowing only controlled releases to the shop floor, reducing the variability in the order entry level [2]. The term load refers to the capacity required in a production system.

The production in job shops lead to queues of orders/work-in-process (WIP) competing for the workstation's capacity. The Workload Control concept seeks to achieve small and stable queues, i.e., to establish low and stable direct load levels [3].

The Workload Control rules apply to the discrete domain, i.e. to systems with discrete production orders. Most studies in the literature model and test this approach by means of discrete event simulation, with order arrival, due dates and processing times defined by probability distributions [4]. Other approaches from the dynamic modelling field are a relevant alternative to discrete event simulation (DES), since they allow a more analytical understanding of the fundamental dynamics of the system allow to automatically adjust WLC parameters in response to dynamic changes and other disturbances. While the discrete event simulation is a trial-and-error method [5], if used alone, the dynamic modelling allows the use of feedback loops, i.e., automatic control, generating smart closed loop models, i.e. models that yield prescriptive guidelines for WLC operation.

Different approaches for dynamic modelling and control can be used, such as transfer functions, difference equations, optimal control, etc. In [6], the author uses bond graphs (BG) and proposes the representation of production stations based upon analogies with electrical components like resistor, capacitor and source of effort. This representation can also be found in [7], and can be applied to Workload Control simulation.

This paper aims to model a Workload Control system as a continuous dynamic system with feedback loop (i.e., a closed-loop system), and analyse the behaviour of this system in face of demand disturbances. The proposed model is smart in the sense that the WLC parameters are automatically changed in function of the current state of the shop floor; this parameterisation is responsive to disturbances. In the extant WLC models, the parameterisation is done by means of a set of experiments with open-loop systems (the DES systems proposed in the extant literature are open-loop systems, i.e., systems without feedback loop).

The proposition of a closed-loop model (automatically controlled by a controller) is an unusual approach for WLC simulations. It provides intelligence to the production planning system. Currently, smart production planning and scheduling [8, 9] are relevant research trends. An additional gap in the Workload Control literature refers to research that deals, simultaneously, with input and output control. Input control in WLC is about restraining order release, while output control corresponds to adjustments in the capacity of the machines. The vast majority of WLC studies focus on input control; publications related to output control are scarce, and even scarcer is the research related to the simultaneous application of both (e.g. [10]). The proposed model also addresses this gap by proposing control rules for the application of both input and output controls.

Thus, the novelty and contributions of this research are summarized as follows:

- A novel closed-loop continuous model for WLC simulation is proposed;
- Input and output control are simultaneously operating in the same model, expanding the extant literature;
- The input-output control adjustments are given by a single continuous simulation run (for a given scenario) in contrast to prior research where the usual approach is to determine the control adjustments by multiple experiments;
- The resulting capacity increments are gradual, based on continuous curves, in contrast to fixed-step increments proposed in the literature (see Section 2).

2. Literature review

A search for papers presenting any type of relationship between Control Theory and Workload Control was conducted as theoretical base for this research. No closed-loop models (i.e. models with a controller and a feedback loop) of WLC systems were found. In [11], capacity adjustments of the stations are performed in function of the shop floor current state; however, the model does not present a controller. Instead, pre-defined rules are used to perform the adjustments, with constant values as increments/decrements. A few WLC papers with multi-agent systems were found. Moreover, the review included papers with the simultaneous application of input and output control in WLC. These two groups of papers will be discussed as follows.

In [12], a system with 3 agents is proposed for WLC implementation: order entry agent (OEA), job agent (JA), and machine agent (MA). The due date settings and customers' enquiries are performed by the OEA. The JA defines the job routings and the MA allocates the jobs in each machine queue.

A similar approach is presented by [13] using 4 agents. The first agent (OEA) demands information of current shop status and communicates with the JRA, which operates the job release using a continuous aggregate loading (CAGG) mechanism: jobs are continuously released to the shop if the workload is below the norm. The JRA communicates with the RSA to calculate the current workload [13]. The RSA encompasses job agents (JA) and machine agents (MA), with the same functions as in [12]. The fourth agent, IFA, estimates the average lateness, used by the OEA for decision-making [13].

The authors in [14] propose a WLC dynamic model based upon product agents and planning agents. The job pool agent (JPA) keeps data regarding products waiting in the job pool. The shop floor assessment agent (SAA) collects shop floor data about orders already released, and computes the shop floor simulation. The results are sent to the JRA, together with the information from the pool. The JRA assesses the impact of each potential order release and makes continuous release decisions.

The discussed multi-agent systems present similarities. Both [13] and [14] use a job release agent (JRA) for the release decision-making, and the job routing and sequencing agent (RSA) has mostly the same responsibilities as the shop floor assessment agent (SAA). However, in [13], CAGG is the job release mechanism, while [14] use priority indexes, defined to each product agent, for order sequencing and release.

The second branch of this review contains the papers that discuss WLC approaches with simultaneous control of input (order release) and output (capacity adjustment).

The authors in [15] present a conceptual view of different approaches in WLC. Fig. 1a portrays the input control, where the released work quantity is based on the shop floor queues. This is the most common approach in workload control research. Fig. 1b shows output control; the released work quantity is uncontrolled and the work centre capacities are controlled with the aim of keeping the queues at a constant level. Figs. 1c and 1d show possibilities for simultaneous input/output control. In the approach shown in Fig. 1c, both order release and work centre capacities are controlled on the basis of shop floor load. In Fig. 1d, only the order release is controlled based on the shop floor load while the work centre capacities are controlled according to the pool level.

In [11], the authors analyse the effects of input-output control in a pure job shop. Urgent and non-urgent jobs are organised in the pool according to different rules. The release is based on the LUMS COR method, which is periodic but has a trigger against station starvation. This represents the input control of the system. The capacity adjustments (output control) are based on three parameters: the size of the increment (α), the load threshold that triggers the capacity adjustment (upper limit of utilisation, β) and the load reduction threshold that triggers the adjustment setting (γ). The authors perform simulations with 5 levels for α (0, 10, 20, 30 or 40 %), 3 levels for β (85, 90 and 95th percentile) and 3 levels for γ (0, 5 and 10 %).

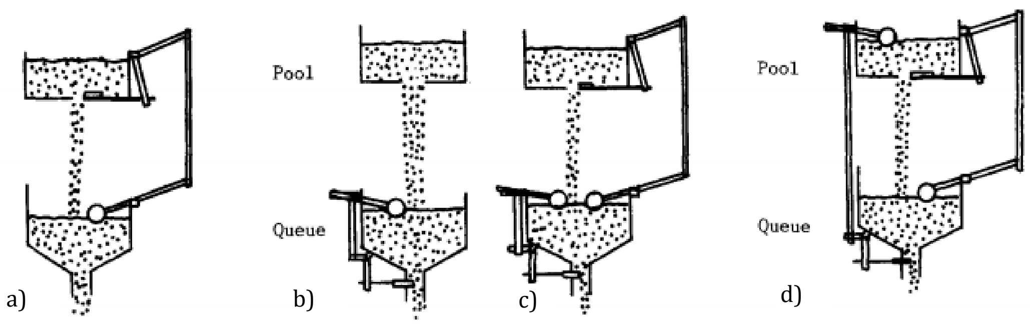

Fig. 1 Representation of the concepts of input-output control: a) input control; b) output control; c) and d): approaches for simultaneous input and output control [15]

The aforementioned study is based on discrete event simulation (DES), and constant capacity increments, chosen from a finite set of values, are applied. When utilization reaches a defined threshold (β), e.g. 85 %, a step-shaped increment (α) is applied; the extra capacity is deactivated after the utilization reduces a certain amount (γ). The representation of a WLC system as a continuous dynamic system (using bond graphs) allows the application of automatic controllers, yielding a closed-loop model. In this way, the WLC parameters, i.e. order release level and size of capacity increments, can be defined as continuous curves, based on the instantaneous shop floor load, instead of being defined as fixed/constant values based on experiments.

There is a paucity of research investigating the simultaneous application of WLC input and output control, as mentioned. The proposed closed-loop model and simulations serve as proof of concept for the alternatives presented in [15] and shown in Fig. 1, also complementing the (scant) existing input-output control research.

Considering a wider view, other advanced methods of production control were analysed in [16]. Model-based production control was proposed in combination with computational intelligence and data-based modelling methods. Among those, evolving model identification has been successfully applied to production systems with non-linear process behaviour [17] and combined with design of experiments [18]. A data-driven modelling approach has been successfully applied to energy consumption optimization in steel production [19].

The application of reinforcement learning (RL) to production control is also a recent research trend. Most models found in the literature that consider work-in-process (WIP) or workload control employ the CONWIP (Constant Work in Process) policy, such as [20-22], and not WLC. CONWIP is simpler than WLC, because only the total workload of the system matters, instead of the workload of each machine. In CONWIP, a new order or amount of work is released only when an order or amount of work outputs the system, keeping the total work-in-process constant. A fixed number of cards or containers is used to limit the WIP. In the aforementioned works, an agent that learns by reinforcement is coupled to a DES system representing the shop floor. The agent observes the current state of the system and modifies the amount of cards in it, i.e., it adjusts the total WIP allowed in the system dynamically, influencing the system's throughput. This production control policy is simpler than WLC or than other production control systems because there is just one parameter to be adjusted. No works specifically applying reinforcement learning to WLC were found.

A similar framework (i.e., combining DES and decision-making algorithms) is employed by [23] to a system ruled by the Make-To-Availability (MTA) policy. With MTA, for a given product, the orders are released and sequenced according to the difference between the established target level for the buffer of that product and the actual inventory on hand. The bigger this difference is, the higher is the priority of the order. In [23], the target level of the buffers is adjusted considering the current state of the plant, which is influenced by demand fluctuations and variability of the processing times, among others.

A literature search in Scopus database with the string *("artificial intelligence" OR "machine learning" OR "reinforcement learning") AND "workload control"* was carried out, returning 10 papers. The string was applied to the field "Title, Abstract and Keywords". Among these 10 papers: two were from other areas (i.e., used the term "workload" within other contexts); one [20] applied reinforcement learning to CONWIP, and not to WLC; six investigated specific parameters/aspects of WLC but not related to machine learning (ML) or artificial intelligence (AI), and one treated about WLC software development. This reinforces the existence of a relevant gap to be tackled, concerning the dynamic parameterisation of WLC and the incorporation of AI and ML techniques into WLC research.

3. Modelling and simulation

The system simulated in this paper was presented by [4]. Jobs and material are provided by the source $S_{01\text{-}pool}$, which is connected to the job pool, i.e. the queue of jobs/orders awaiting release. The shop floor has four workstations (Fig. 2), representing a general flow production [24]. This means that there are different product families with different production routings, without reentrant flows (the flow is directional). Earlier studies on output control, such as [11], considered pure job shops and not general flow shops. Using the data on product families, routings and demand (Table 1, taken from [4]), the percentage of material that must flow into each branch of the system can be calculated (Fig. 2).

Table 1 Production routings and demand rates for the unidirectional job shop

	Station 1	Station 2	Station 3	Station 4	Demand rate (units/day)
Family 1	X	X			1.5
Family 2	X	X		X	1.25
Family 3			X		1.3
Family 4			X	X	1.1

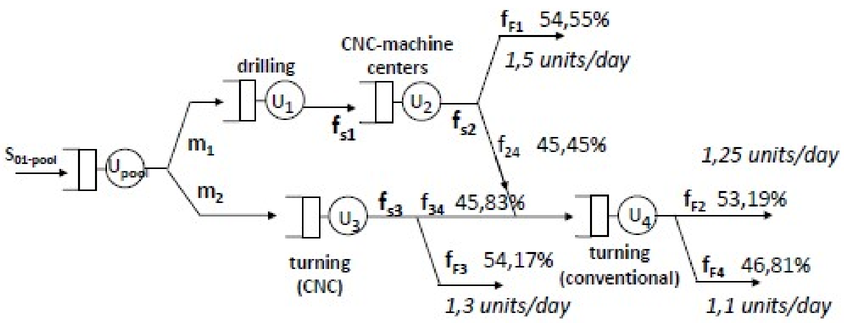

Fig. 2 Schematics of the modelled WLC system

A modelling approach based on bond graphs is chosen for the development of a mathematical model of the production system. Bond graphs capture the power and energy exchange between the model elements. The graphical representation of these elements and their interconnections can be translated into equations that form a mathematical model.

To represent a shop floor with bond graphs requires some adaptations of the standard bond graph formalism. A corresponding bond graph model for a production station, based on analogies to capacitor, resistor and source of effort, was proposed in [6]. The representation focuses on fluidised, continuous material flow (f), which can also be interpreted as the amount of work per unit of time. The approach in [6] assumes an unlimited buffer capacity, which allows the output flow of the i-th production station to be expressed as:

$$f_i = U_i \min(1, q_i), \tag{1}$$

where U_i is the processing frequency of the station (i.e. the reciprocal of a processing time) and q_i is the amount of material stored in the i-th buffer. These variables are dependent on time, e.g. $f(t)$, but the time dependence is not explicitly shown in the equations for simplicity. An analogy is assumed between the material flow rate of a machine and the electric current through a resistor, where U_i is the reciprocal of the resistance ($U_i = 1/R_i$). The expression $\min(1, q_i)$ stands for the effort (analogue to voltage), which gradually decreases with a small amount of material storage. Further details on this modelling approach can be found in [6].

The material quantity rate in the station buffer ($\dot{q}_i$) depends on the balance between the material input rate and material consumption rate, which is expressed by the difference between material input and material output flows of the station. This leads to equation:

$$\dot{q}_i = f_{ei} - U_i \min(1, q_i), \tag{2}$$

where f_{ei} is the material input flow of station i while U_i and q_i have the same meaning as in (1). The use of these equations for modelling multiproduct manufacturing systems is presented in [7, 25, 26]. In the simulated model, the controlled variables are the levels of the buffers ($q_i, i = 1,2,3,4$) and of the job pool (q_{pool}), and the manipulated variables are the processing frequencies of the machines (U_i, with $i = 1,2,3,4$), and the release frequency of the job pool (U_{pool}). The most important variables are summarised here:

- $\dot{q}_{pool}$: change rate of waiting orders/materials in the pool;
- $\dot{q}_i$: rate of change of the material quantity in buffer *i*;
- q_{pool}: quantity of orders waiting to be released;
- q_i : quantity of material in machine buffer *i*, where *i* = 1,2,3,4;
- U_{01pool}: rate/frequency of orders/materials entering the job pool;
- U_{pool}: release frequency. In the model, the release is considered an operation and the pool operates similarly to a work centre;
- U_i : processing frequency of machine *i*, with *i* = 1,2,3,4.

The products' demand rates (Table 1), which can be modified or modelled as curves to represent volatile demand, are the main parameters of the system. The model is also parameterized by m_1 and m_2, which will be described later.

Applying (2) to the pool and buffer of station 2 gives:

$$\dot{q}_{pool} = U_{01pool} - U_{pool} \min(1, q_{pool}) \tag{3}$$

$$\dot{q}_2 = U_1 \min(1, q_1) - U_2 \min(1, q_2). \tag{4}$$

Stations 1 and 2 are connected in series, so that the input of station 2 in (4) is replaced by the output of station 1.

Convergent and divergent junctions conduct parts of the flow to different stations, according to the production routings shown in Fig. 2. Convergent junctions establish the flow conservation and are represented as a 1-junction in bond graphs, leading to:

$$f_{si} = \sum_{p=1}^{P} f_{ep}, \tag{5}$$

in which f_{si} represents the output flow of a *i*-th convergent junction, and f_{ep}, *p* = 1,2, ..., *P*, represent the branches of incoming flow. The divergent junctions can be represented as transformers in bond graphs. The q-th transformer with module m_q receives an input flow f_{ei} and converts it into an output flow f_{sq}:

$$f_{sq} = m_q f_{ei}, \tag{6}$$

with $\sum_q^Q m_q = 1$ and *q* = 1,2, ..., *Q*, where Q is the total number of diverging branches [4].

The proposed model has modules m_1 and m_2, which describe how the material flow is distributed between the two processing branches coming from the pool (see the first junction of Fig. 2). The application of (6) to this junction and of (2) to station 1 and station 3 yields:

$$\dot{q}_1 = m_1 U_{pool} \min(1, q_{pool}) - U_1 \min(1, q_1) \tag{7}$$

$$\dot{q}_3 = m_2 U_{pool} \min\left(1, q_{pool}\right) - U_3 \min(1, q_3). \tag{8}$$

In the case considered, the parameters were set to m_1 = 0.53398 and m_2 = 0.46602 to obtain the desired production mix. The application of (5) and of (6) considering the percentages of split shown in Fig. 2 yields:

$$f_{e4} = f_{24} + f_{34} \tag{9}$$

$$f_{24} = 0.4545 U_2 \min(1, q_2) \tag{10}$$

$$f_{34} = 0.4583 U_3 \min(1, q_3) \tag{11}$$

$$\dot{q}_4 = 0.4545 U_2 \min(1, q_2) + 0.4545 U_3 \min(1, q_3) - U_4 \min(1, q_4). \tag{12}$$

The instantaneous WIP in the shop floor (q_i) is calculated by the integral of the rates of material storage or consumption in the buffers, and the amount of orders waiting in the pool (q_{pool}) is obtained by the integral of the release rate of orders/material in the pool. These are the state variables of the system (q_i, q_{pool}). The algebraic manipulation of (3), (4), (7), (8), and (12) yields the state model:

$$\begin{bmatrix} \dot{q}_{pool} \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_4 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 1 \\ m_1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ m_2 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0.4545 & 0.4583 & -1 & 0 \end{bmatrix} \begin{bmatrix} U_{pool} min\,(1, q_{pool}) \\ U_1 min\,(1, q_1) \\ U_2 min\,(1, q_2) \\ U_3 min\,(1, q_3) \\ U_4 min\,(1, q_4) \\ U_{01pool} \end{bmatrix} \tag{13}$$

Details about this modelling, the development of the equations, and the variables and parameters can be obtained in [4, 25, 26].

In WLC, input control refers to the release of orders, which corresponds to the control of the processing frequency of the job pool, U_{pool}. Output control refers to the control of the processing frequencies of each machine, U_1 to U_4, which correspond to capacity adjustments. These controls are simultaneously executed in the proposed model. Release decisions are made in function of the shop floor load and the pool's load. Since all the production routings start either in station 1 or station 3 (Fig. 2), the buffer level of these stations is considered as a measure of the shop floor load. The load of the pool is measured based on the level of its own buffer of orders, q_{pool}. The level of the buffer upstream the i-th machine is assessed to define the capacity adjustment of the respective machine. As an analogy to Fig. 1, Fig. 3 represents the input and output control approaches used in all the simulations.

The approach adopted is similar to the one of Fig. 1c, with the releases based on the shop floor load and the capacity adjustments based on the machines' preceding buffers. However, in the proposed model, the releases also rely on the job pool load, differing from [15] and from other WLC studies. Moreover, in none of the works reviewed in Section 2, input and output control of Fig. 1c is implemented as a dynamic closed-loop system. To implement this system, a controller is attributed to each machine and to the pool. Fig. 4 shows the information that is used by each controller to adjust the processing frequencies U_{pool} and U_i.

Three different simulations were carried out. In all of them, the reference values for the buffer levels are: $q_{pool\_c} = q_{1c} = q_{2c} = q_{3c} = 7$ and $q_{4c} = 6$. In simulation A, the system starts with empty buffers, to verify if stable operation would be achieved. Simulations B and C aimed to verify the system's behaviour in face of disturbances in the order entry level (which represents external demand). In both simulations, the buffer levels (WIP) start at the reference values, to check how the disturbances impact a system initially in balance. For Simulation B a step signal was chosen to represent the disturbances, and for Simulation C, a sinusoidal signal. Reference processing frequencies (U_{ip}) for the machines and for the sources were calculated based on (13) and on the demand rates shown in Fig. 2. These processing frequencies (U_{01pool_p} = 5.15, U_{pool_p} = 5.15, U_{1p} = 2.75, U_{2p} = 2.75, U_{3p} = 2.4, U_{4p} = 2.35) lead to the attendance of the demand in the medium term (in steady state) and are used as parameters for the control of the machines and the pool.

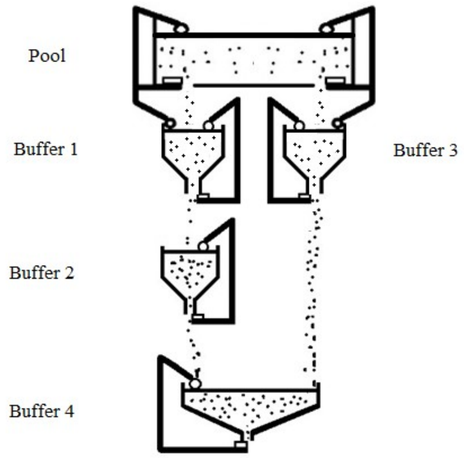

Fig. 3 Representation of simultaneous input and output control applied (analogy with Fig. 1)

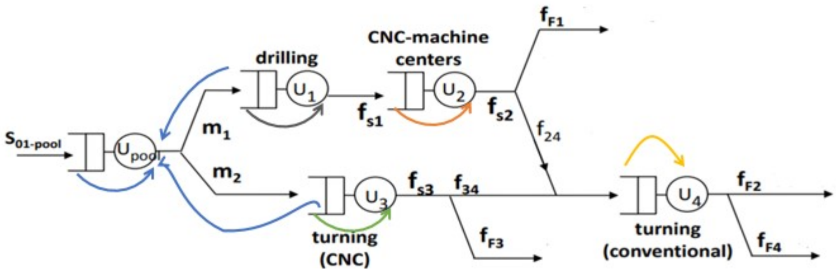

Fig. 4 signals sent to the controllers to implement the control rules (right side)

4. Results and discussions

4.1 Implementation

The model was implemented in Matlab and Simulink® (Fig. 5). The blue box represents the proportional integral (PI) controllers added to the system, which will define the processing frequencies of the pool and of each machine (U_{pool} and U_i). These instantaneous processing frequencies fluctuate around the steady state processing frequencies presented at the end of Section 3 (i.e. the output of the controllers are used as relative values to modify the steady state frequencies). The components highlighted in green represent the arrival of data for the controllers. The pool's controller receives 3 signals: the error from its own buffer, and the error from buffer levels 1 and 3. The controller of each machine receives only error signal of its preceding buffer. This implementation is aligned to the proposed control rules (Fig. 4).

The control signals of the system will be inputs for the production function, represented by the yellow box. This function calculates the rates of production or consumption of material ($\dot{q}_{pool}$ and $\dot{q}_i$) according to the state equation (13). In this block, there is also an integrator to convert the rates into work in process (WIP) stored in the pool and in the buffers. These values are, then, fed back into the system – as represented by the purple lines – and compared with the reference values in (red box). This comparison defines the error, which will serve as input for the controllers, closing the loop.

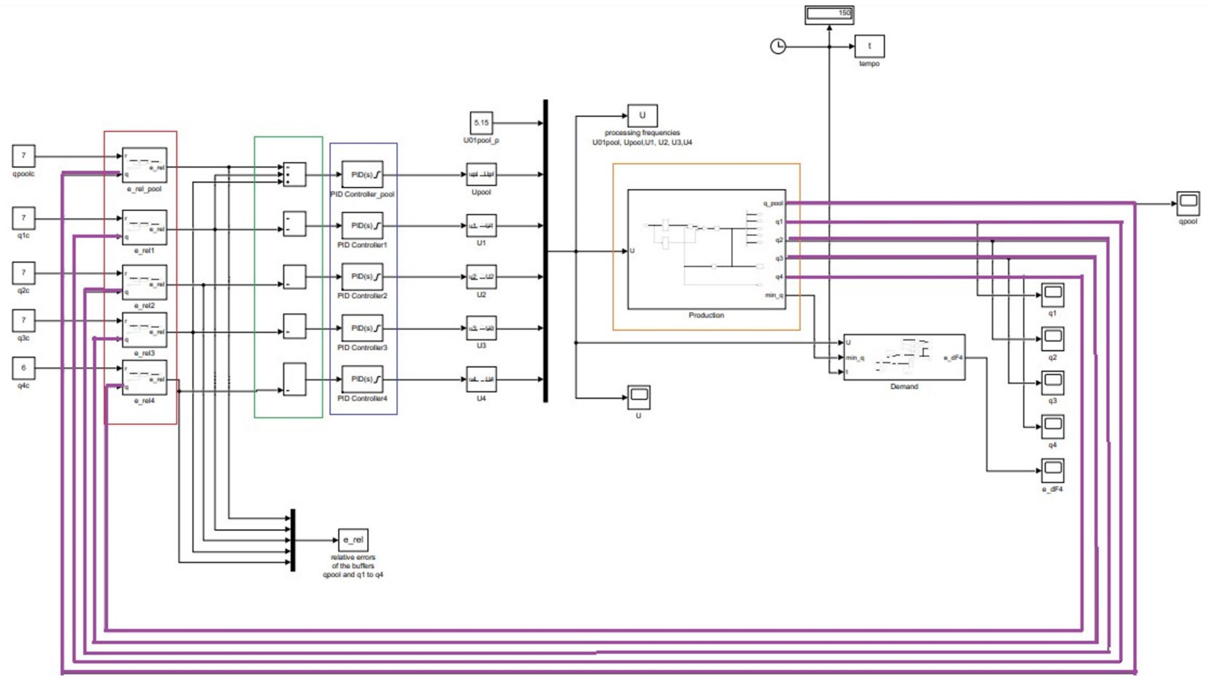

Fig. 5 Simulink® representation of the proposed model.

4.2 Results

The results of simulation A are shown in Figs. 6a and 7a. The controller of the pool receives signals from the buffers 1 and 3 that lead to the increase of the order release frequency, since the levels of both buffers are below the reference. A signal in the opposite direction comes from the pool itself. Since the level of orders in the pool is also below the reference, it induces a reduction in the release frequency, to clog the orders. The signals coming from the buffers 1 and 3 have a stronger influence on the control at the initial moments, and the order release frequency (U_pool) increases: the pool release frequency starts at 200 % of the release frequency in the steady state. This value decreases gradually as the levels of the buffers 1 and 3 increase and get closer to the reference (Fig. 7a). The inflection point of the release frequency curve occurs when machines 2 and 4 start processing, stimulating an increase in the flow through the whole system (i.e. the release frequency starts to increase again). After the transient period, the WIP stabilize at the reference values, showing that the control rules are effective.

In simulations B and C, the initial buffer levels match the reference levels. In Simulation B, there is a 20 % step-shaped increase in the order's arrival until the sixtieth day. In response to that, the release frequency of the pool increases due to the high levels of orders on wait. Next, there is an increase in the processing frequencies of the machines immediately downstream (machines 1 and 3). Finally, the reaction reaches the machines 2 and 4 (Fig. 6b).

The variations of the work in process in the buffers is not accentuated, showing that even with the disturbances – almost 20 % increase on the entry signal – the system did not suffer great penalties related to the buffers' levels (Fig. 7b). This behaviour results from the coordinated increase in the release frequency and in the processing frequencies of the machines (capacity adjustments).

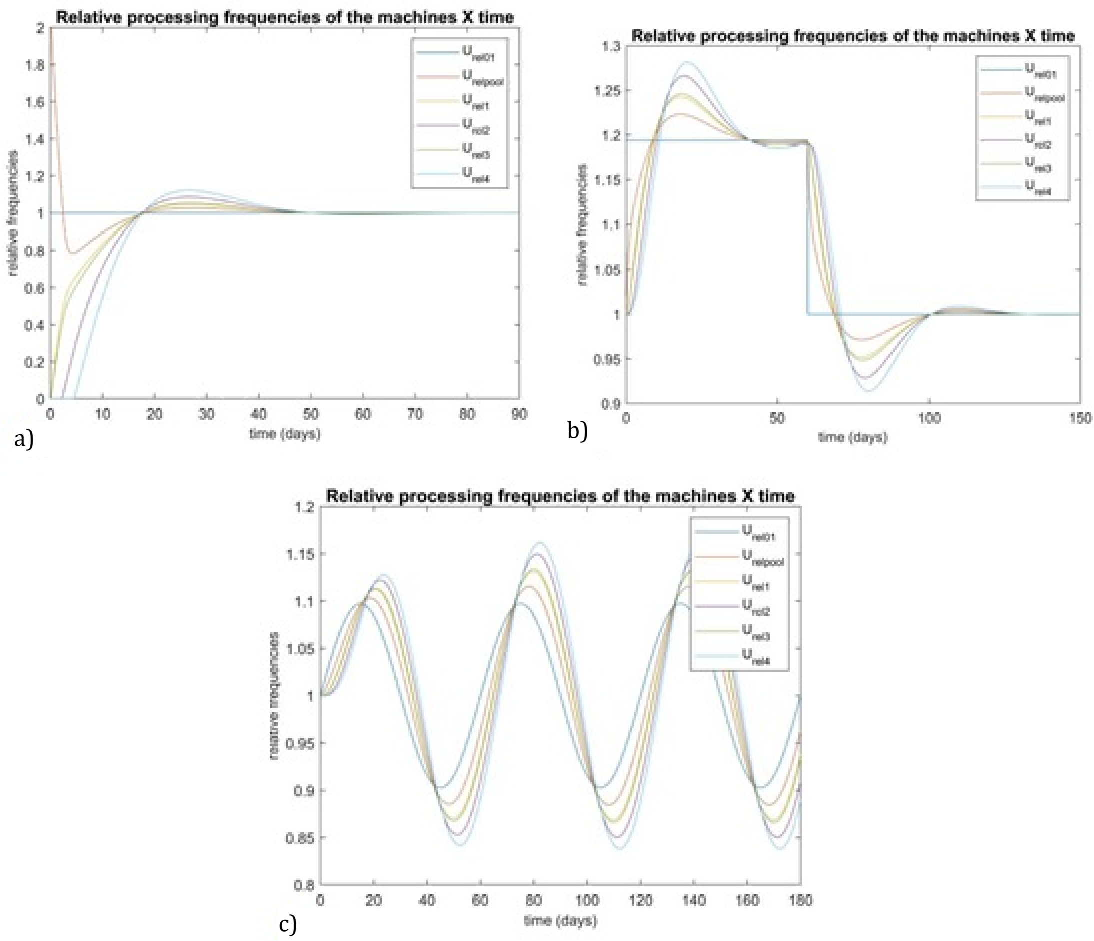

Fig. 6 Relative processing frequencies of the release and of the machines for:
a) Simulation A; b)Simulation B; c) Simulation C

With the return to the standard demand/input values (sixtieth day), the processing frequencies decrease and there is an undershoot before the system returns to operate at its regular capacity. This also reflects in an undershoot in the buffer levels. After that, they are led to the reference values again.

In Simulation C, the entry signal presents a sinusoidal variation of 10 % of the nominal value during the simulation (see curve of U_{rel01} in Fig. 6c).

The system regulates itself. The reaction begins with a higher release in the pool, and more material flows downstream with the increase of the processing capacity of the machines. A small delay in the reactions can be seen. The amplitude of the oscillations also gradually increases downstream. The last machine of the routings (Machine 4) has the higher peak and valley, with amplitude of almost 15 % of the nominal value. Meanwhile, the pool, the element most upstream, has the smallest variation, of 10 % (Fig. 6c).

The general behaviour of the system follows the behaviour of the entry signal, with a certain delay and with some amplification of the disturbance. The downstream stations are more affected. The system responds to the cyclic demand variations with a maximum 15 % of capacity variation (Fig. 6c).

The system reaches the steady state and the reference values for the buffers are achieved (simulations A and B). In simulation C, even with an increase of almost 20 % in the entry signal (demand signal), the WIP does not variate excessively (Fig. 7c). These are positive results, considering that the goal of WLC is to obtain a more stable and predictable shop floor. From the managerial point of view, it means that the system reacted to the demand increase without a strong increase the throughput time. This was only possible due to the increase in the processing frequencies (capacity adjustments). Managers, however, must analyse the viability of implementing the suggested increases in the machine's capacity.

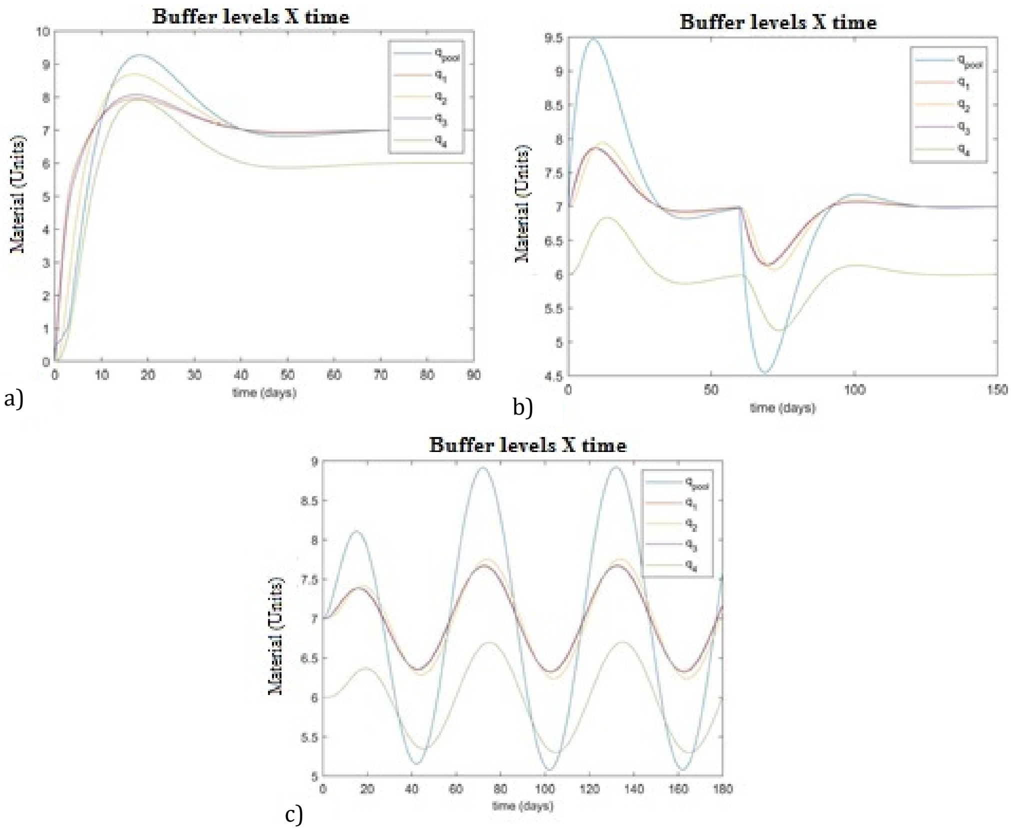

Fig. 7 Buffer levels for: a) Simulation A; b) Simulation B; c) Simulation C

The model developed presents a prescriptive characteristic. The results of simulation of the closed-loop model (i.e., with an automatic controller) point out when and with which amount adjustment actions must be taken to attend the demand fluctuations. Based on that, managers can define actions to implement the capacity increase or decrease, such as the reallocation of personnel, application of extra shifts, or the reduction of working hours of determined resources (i.e. defining programmed pauses in the schedule of the machines), when demand falls.

4.3 Discussions and model extensions

The applicability of the model in practice as well as extensions for its application to real and more complex shop floors and to supply chains are discussed as follows.

In the real manufacturing environment, the model could be implemented relatively easily from a technical point of view if the manufacturing environment is already supported by information systems, e.g. a Manufacturing Execution System (MES). Such a system collects data that can be used to align model states with the real environment and also offers options for adjusting capacities. More challenging is the underlying organisational change that would redefine the established concepts of production control. Transition to Industry 4.0 brings increasing demands of digitalisation of production processes in smart factories, e.g. through the use of Digital Twins and digital agents [27] in production control-related decision-making processes. This increases also the opportunities for incorporating advanced feedback control concepts such as the one proposed here.

Capacity adjustments can be implemented, in practice, in different ways. Small increments are usually implemented by means of overtime; substantial increments require long term managerial actions, as extra shifts, subcontracting or plant expansion. In the simulated scenarios, the capacity adjustments do not exceed 30 % (i.e. relative processing frequencies of 1.3 times the usual frequency, as shown in Fig. 6). Hence, they can be implemented by means of overtime and are associated only to variable costs, that is, it is supposed that extra fixed costs are negligible. Then, if a constant line $y = 1$ is drawn in the graphs of the relative processing frequencies, the cost of capacity adjustments can be related to the areas limited by the curves of processing frequencies and this reference line. For capacity increments, the areas above the line can be multiplied by a unitary cost of 1.5, if the overtime costs 50 % more than the normal work hour. Capacity decrements (or machines' downtime) also represent costs, because there are regular fixed costs related to the overall structure even when the machine is idle. So, the areas below the line $y = 1$ can be multiplied by a unitary cost (e.g., of 0.7) to estimate the cost of capacity decrements. These costs of gradual/continuous capacity adjustments can be compared to the costs of other policies, based on fixed adjustments, as the policy proposed by [11].

The advantage of dynamic models compared to discrete event simulation (DES) as a decision-making aid lies in the significantly lower model complexity. The probabilistic nature of DES models requires lengthy simulations before appropriate conclusions can be drawn. The dynamic model represents an aggregation of the complex DES dynamics. This comes at the cost of lower model accuracy, but the reduced complexity enables very fast simulation and even optimisation in real time, e.g. in the context of model predictive control. The lower model accuracy is compensated by the robustness of the feedback loop with the correctly tuned controller. Furthermore, aggregated dynamic models can also be used much more efficiently for sensitivity analyses and parameter calibration compared to DES. In contrast, DES models offer more flexibility and accuracy in simulating real emergency scenarios, such as machine failures and emergencies on the supply and/or demand side.

The model can be scaled to more complex shop floor configurations due to the modularity of the bond graphs' modelling technique. There is a bond graph model to a production station, with corresponding constitutive equations, derived from bond graphs' ideal elements, and there is a unique/fixed correspondence between an element of BG and its mathematical equation. The stations are linked by means of 1-junctions or transformer elements, as presented in Section 3. These elements allow representing different production routings, using different machine combinations. Thus, to represent shop floor configurations with much more machines and intricate flow, the modeller should add the machines and junctions into the bond graph pictorial representation, and then add the equations, corresponding to each element, into the mathematical model. After that, adequate algebraic manipulation leads to the matrix representation of the system. There is software, as 20-sim [28], that can automatically generate the mathematical/simulation model when the user elaborates the BG model by selecting the pictorial elements from a specific library. A real system in the textile industry with 11 machines and 9 different product routings was modelled by [7], showing that it is possible to scale the model. In the cited work, however, the shop floor control does not follow the principles of Workload Control. Since the goal is not to find an optimal solution to the models, but to simulate them with feedback control, the curse of dimensionality does not apply, i.e. the models are tractable and it is possible to simulate models with big dimensions using the computational capacity currently available.

External factors such as supply chain disruptions can affect the supply of raw materials to the system. The effect of these factors could be observed by simulating the temporary interruption of the flow coming from the source $S_{01\text{-}pool}$; i.e., the lack of raw material would prevent the entry of orders on the shop floor. If the suppliers take a significant time to recover from the disruption, the situation is similar to the one of simulation A, where the system has to start from the beginning, with empty buffers. The reference levels for the buffers is reached but with an overshoot, as seen in Fig. 7a. Another way to simulate supply chain disruptions would be to explicitly represent suppliers and customers as production stations linked to the plant already modelled. A delay in the output flow of the supplier's station can be added to represent the transportation and logistics times between the links; a delay in the output flow of the modelled plant can be also added to represent the logistics times needed to deliver the goods to the next or final customer. Other dynamic modelling tools such as block diagrams and transfer functions have been widely used to study the supply chain dynamics. For a review of these models, [29] can be consulted.

5. Conclusion

Simulations of WLC in the literature are based on open-loop discrete models. An alternative approach is presented in this paper: a continuous and closed-loop dynamic model for simultaneous application of input and output control in WLC. A closed-loop model includes automatic control, so that the job release and the capacity adjustments of the machines are defined in a dynamic way (as function of the shop floor real-time state) and vary in a continuous way.

Three simulations were carried with a control rule that applies input and output control simultaneously (simultaneous release and capacity adjustments). Simulation A is conducted to study the transient response of the system. Simulations B and C include fluctuations in order entry/demand according to a step and a sinusoidal signal.

The implemented feedback control leads the buffer levels to achieve the defined targets and the processing frequencies to stabilize in simulations A and B. In Simulation C, the system responded to the cyclic variations of demand with a maximum 15 % variation of capacity. Even with an increase of almost 20 % in the order entry signal, the WIP did not increase significantly. This means shorter throughput times and a balanced and predictable shop floor, reactive to disturbances. In the proposed closed-loop model, the adjustments are a function of the real-time state of the system.

The first contribution of this paper is the proposition of a prescriptive model that indicates when to release orders and to which amount to adjust the system's capacity. A managerial analysis should define how these adjustments shall be implemented, and its viability. Another relevant contribution is the simultaneous application of input and output control in WLC, considering that this kind of study is scarce in the literature.

In the literature review, no model of WLC that applies the concept of feedback control – as seen in Control Theory – was found. Thus, there is still space for the implementation of continuous and closed-loop simulations of Workload Control, bringing relevant information about the dynamic behaviour of the system. Interdisciplinary models based on the integration of Control Theory and Operations Management can lead to smart production control systems, and expands the range of tools to be used in Production Engineering and Manufacturing research.

For future research, the proposition of different scenarios (changes in the initial/reference values, use of different signal shapes for the order entry) and the development and simulation of different control rules for both input and output control are suggested. This will allow studying the system's behaviour when the controllers have a more global view of the system. This paper is a proof of concept of the closed-loop simulation of WLC. Thus, future studies could apply this model to more complex real systems. Another suggestion is the implementation of feedback loops to models based on discrete event simulation (DES). This is not trivial to be developed.

Acknowledgement

The second author acknowledges the financial support from the São Paulo Research Foundation under process #2019/12023-1 and from the National Council for Scientific and Technological Development under process #310812/2021-6. The third author acknowledges the financial support from the Slovenian Research and Innovation Agency (ARIS) (research core funding No. P2-0219).

References

[1] Shaladdin Muda, M., Hendry, L. (2003). The SHEN model for MTO SMEs: A performance improvement tool, *International Journal of Operations & Production Management*, Vol. 23, No. 5, 470-486, doi: 10.1108/01443570310471820.

[2] Lödding, H. (2013). *Handbook of manufacturing control: Fundamentals, description, configuration*, Springer, Heidelberg, Germany, doi: 10.1007/978-3-642-24458-2.

[3] Breithaupt, J.-W., Land, M., Nyhuis, P. (2002), The workload control concept: Theory and practical extensions of load oriented order release, *Production Planning & Control,* Vol. 13 No. 7, 625-638, doi: 10.1080/0953728021000026230.

[4] Sagawa, J.K., Land, M.J. (2018). Representing workload control of manufacturing systems as a dynamic model, *IFAC-PapersOnLine*, Vol. 51 No. 2, 825-830, doi: 10.1016/j.ifacol.2018.04.016.

[5] Duffie, N.A., Roy, D., Shi, L. (2008). Dynamic modeling of production networks of autonomous work systems with local capacity control, *CIRP Annals,* Vol. 57, No. 1, 463-466, doi: 10.1016/j.cirp.2008.03.018.

[6] Ferney, M. (2000). Modelling and controlling product manufacturing systems using bond-graphs and state equations: Continuous systems and discrete systems which can be represented by continuous models, *Production Planning & Control*, Vol. 11, No. 1, 7-19, doi: 10.1080/095372800232441.

[7] Sagawa, J.K., Nagano, M.S., Speranza Neto, M. (2017). A closed-loop model of a multi-station and multi-product manufacturing system using bond graphs and hybrid controllers, *European Journal of Operational Research*, Vol. 258, No. 2, 677-691, doi: 10.1016/j.ejor.2016.08.056.

[8] Rossit D.A., Tohmé, F., Frutos, M. (2019). Industry 4.0: Smart scheduling, *International Journal of Production Research*, Vol. 57, No. 12, 3802-3813, doi: 10.1080/00207543.2018.1504248.

[9] Cañas, H., Mula, J., Campuzano-Bolarín, F., Poler, R. (2022). A conceptual framework for smart production planning and control in Industry 4.0, *Computers and Industrial Engineering*, Vol. 173, Article No. 108659, doi: 10.1016/j.cie.2022.108659.

[10] Thürer, M., Stevenson, M., Land, M.J., Fredendall, L.D. (2019). On the combined effect of due date setting, order release, and output control: An assessment by simulation, *International Journal of Production Research*, Vol. 57, No. 6, 1741-1755, doi: 10.1080/00207543.2018.1504250.

[11] Thürer, M., Stevenson, M., Land, M.J. (2016). On the integration of input and output control: Workload control order release, *International Journal of Production Economics*, Vol. 174, 43-53, doi: 10.1016/j.ijpe.2016.01.005.

[12] Wu, Z., Weng, M.X. (2005). Dynamic due date setting and shop scheduling for make-to-order companies, In: *Proceedings of the IIE Annual Conference*, Institute of Industrial and Systems Engineers (IISE), USA, 1-6.

[13] Weng, M.X., Wu, Z., Qi, G., Zheng, L. (2008). Multi-agent-based workload control for make-to-order manufacturing, *International Journal of Production Research*, Vol. 46, No. 8, 2197-2213, doi: 10.1080/00207540600969758.

[14] Akillioglu, H., Ferreira, J., Onori, M. (2013). Demand responsive planning: Workload control implementation, *Assembly Automation*, Vol. 33, No. 3, 247-259, doi: 10.1108/AA-12-2013-040.

[15] Kingsman, B.G., Tatsiopoulos, I.P., Hendry, L.C. (1989). A structural methodology for managing manufacturing lead times in make-to-order companies, *European Journal of Operational Research*, Vol. 40, No. 2, 196-209, doi: 10.1016/0377-2217(89)90330-5.

[16] Glavan, M., Gradišar, D., Strmčnik, S., Mušič, G. (2013). Production modelling for holistic production control, *Simulation Modelling Practice and Theory*, Vol. 30, 1-20, doi: 10.1016/j.simpat.2012.07.010.

[17] Andonovski, G., Mušič, G., Blažič, S., Škrjanc, I. (2018). Evolving model identification for process monitoring and prediction of non-linear systems, *Engineering Applications of Artificial Intelligence*, Vol. 68, 214-221, doi: 10.1016/j.engappai.2017.10.020.

[18] Ožbot, M., Lughofer, E., Škrjanc, I. (2023). Evolving neuro-fuzzy systems based design of experiments in process identification, *IEEE transactions on fuzzy systems*, Vol. 31, No. 6, 1995-2005, doi: 10.1109/tfuzz.2022.3216992.

[19] Tomažič, S., Andonovski, G., Škrjanc, I., Logar, V. (2022). Data-driven modelling and optimization of energy consumption in EAF, *Metals*, Vol. 12, No. 5, Article No. 816, doi: 10.3390/met12050816.

[20] Xanthopoulos, A.S., Chnitidis, G., Koulouriotis, D.E. (2019). Reinforcement learning-based adaptive production control of pull manufacturing systems, *Journal of Industrial and Production Engineering*, Vol. 36, No. 5, 313-323, doi: 10.1080/21681015.2019.1647301.

[21] Ragazzini, L., Negri, E., Macchi, M. (2021). A digital twin-based predictive strategy for workload control, IFAC-PapersOnLine, Vol. 54, No. 1, 743-748, doi: 10.1016/j.ifacol.2021.08.183.
[22] Vespoli, S., Mattera, G., Marchesano, M.G., Nele, L., Guizzi, G. (2025). Adaptive manufacturing control with deep reinforcement learning for dynamic WIP management in Industry 4.0, *Computers & Industrial Engineering*, Vol. 202, Article No. 110966, doi: 10.1016/j.cie.2025.110966.
[23] Ikeziri L.M., de Souza, F.B., da Silva Meyer, A., Gupta, M.C. (2021). Pulling the distribution in supply chains: Simulation and analysis of Dynamic Buffer Management approach, *International Journal of Systems Science: Operations & Logistics*; Vol. 10, No. 1, Article No. 1981480, doi: 10.1080/23302674.2021.1981480.
[24] Oosterman, B., Land, M., Gaalman, G. (2000). Influence of shop characteristics on workload control, *International Journal of Production Economics*, Vol. 68, No. 1, 107-119, doi: 10.1016/S0925-5273(99)00141-3.
[25] Sagawa, J.K., Mušič, G. (2019). Towards the use of bond graphs for manufacturing control: Design of controllers, *International Journal of Production Economics*, Vol. 214, 53-72, doi: 10.1016/j.ijpe.2019.03.017.
[26] Sagawa, J.K., Nagano, M.S. (2015). Applying bond graphs for modelling the manufacturing dynamics, *IFAC-PapersOnLine*, Vol. 48, No. 3, 2047-2052, doi: 10.1016/j.ifacol.2015.06.390.
[27] Resman, M., Pipan, M., Šimic, M., Herakovič, N. (2019). A new architecture model for smart manufacturing: A performance analysis and comparison with the RAMI 4.0 reference model, *Advances in Production Engineering & Management*, Vol. 14, No. 2, 153-165, doi: 10.14743/apem2019.2.318.
[28] Broenink, J.F. (1999). 20-SIM software for hierarchical bond-graph/block-diagram models, *Simulation Practice and Theory*, Vol. 7, No. 5-6, 481-492, doi: 10.1016/S0928-4869(99)00018-X.
[29] Lin, J., Naim, M.M., Purvis, L., Gosling, J. (2017). The extension and exploitation of the inventory and order based production control system archetype from 1982 to 2015, *International Journal of Production Economics*, Vol. 194, 135-152, doi: 10.1016/j.ijpe.2016.12.003.