EVALUATING THE REPEATABILITY OF RTK GPS MEASUREMENTS USING ANALYSIS OF VARIANCE

VREDNOTENJE PONOVLJIVOSTI OPAZOVANJ RTK GPS Z ANALIZO VARIANCE

Atınç Pirti

UDC:528.28:528.4

Klasifikacija prispevka po COBISS-u: 1.02

ABSTRACT

The purpose of this study is to evaluate the repeatability of RTK GPS method under varying satellite configurations by using different reference points. Furthermore, we conducted one way Analysis of Variance (ANOVA), which is a powerful tool for comparing the variability of related types of measurements. Multi-reference station is a well-known approach in RTK GPS. Using multi-reference points is an effective way to achieve consistent accuracy in the whole net by making errors less distance dependent on the reference stations. It is possible to achieve high reliability and availability by using multi-reference stations. If one station goes down or starts to provide suspicious values, it is possible to compensate the situation with other stations, while this is not possible when a single reference station fails. Here we present ANOVA which is an effective method to check the quality of corrections generated from each reference station.

IZVLEČEK

Namen raziskave je oceniti ponovljivost RTK-metode GPS-opazovanj v odvisnosti od različne geometrije razporeditve satelitov in z uporabo različnih referenčnih postaj. Uporabili smo metodo enosmerne analize variance (ANOVA), ki omogoča učinkovito primerjavo spremenljivosti podobnih meritev. RTK-metodo GPS-opazovanj smo opravili z več referenčnimi postajami, saj je tako mogoče doseči homogeno natančnost v celotni mreži. Prav tako je natančnost manj odvisna od oddaljenosti izmeritvenih točk od referenčne postaje. Tako sta omogočeni visoka zanesljivost in dostopnost opazovanj, saj lahko eno referenčno postajo ob morebitnem izpadu nadomestimo z drugo, kar je velika prednost pred uporabo samo ene referenčne postaje. Z metodo ANOVA lahko učinkovito preverimo kakovost popravkov, ki jih ustvarijo posamezne referenčne postaje.

KEY WORDS

RTK GPS, Accuracy, Repeatability, ANOVA

KLJUČNE BESEDE

RTK GPS, natančnost, ponovljivost, ANOVA

1 INTRODUCTION

Repeated measurements of the same quantity will often yield different values due to errors; discrepancy is defined as the algebraic difference between two or more measurements of the same quantity. When a small discrepancy occurs between repeated measurements, it is generally believed that only random, systematic and gross errors will occur. Therefore, there is a tendency to give higher credibility to such data and to define the measurements precisely. However, precise values are not necessarily accurate values. To clarify the difference between precision and accuracy, the following definitions are presented:

Precision: The degree of consistency between measurements based on the sizes of the discrepancies in a data set. The degree of precision attainable depends on the stability of the environment during the measurement, the quality of the equipment used to make the measurements and the observer's skill with the equipment and measurement procedures.

Accuracy: The measure of the absolute nearness of a measured quantity to its true value. Since the true value of a quantity can never be determined, accuracy is always unknown.

Precision is calculated with the most standard GPS receivers. Achieving location repeatability we could determine an accurate, repeatable position or location in the first instance, which ensures that positioning can be replicated repetitively over any period. The periods for accuracy and time will be predicated by each individual situation. However, these scores may be displaced from the true value by a large but unknown amount and therefore result in being inaccurate. Common sources of errors for GNSS are signal multipath, atmospheric refraction or the result of using incorrect datum or erroneous datum transformation parameters. Reliability refers to the quality of the position result with respect to biases. In a highly reliable position result, even small outliers in the data will be noticed. Conversely, in an unreliable position result, large outliers will go unnoticed. Position reliability is driven by redundancy and is generally represented by parameters that describe the ability to detect outliers and to estimate the effects of undetectable outliers on the estimated parameters. Reliable positioning means the ability to consistently repose oneself to the same location as well as remain within the prescribed limits to suit the requirements and the given situation. It would be a combination of precision and accuracy.

Repeatability: The variation arising when all efforts are made to keep conditions constant by using the same instrument and operator, and repeating them in a short time period. Repeatability value r is the value below which the absolute difference between two single results obtained under repeatability conditions may be expected to lie with a probability of 95%. Assuming that the distribution of the random errors occurring in every single test result is approximately normal, the repeatability value (r) can be calculated as follows:

$$r = 2\sqrt{2}\, S_r \tag{1}$$

where S_r is the standard deviation of repeatability (McClave 2000), (Robouch 2003), (Walpole 1993), (Wolf 1997).

2 ANALYSIS OF VARIANCE (ANOVA)

Variance is an important statistic used in some statistical calculations, such as analysis of variance. Statisticians use the variance and standard deviation of a continuous random variable *X* as a way of measuring its dispersion. Let *X* be a continuous random variable with density function *f* defined on the interval (*a, b*), and let $\mu = E(X)$ be the mean of *X*. Then the variance of *X* is given by

$$S_X^{\,2} = \mathrm{E}((\boldsymbol{X}-\mu)^2) = \int_a^b (\boldsymbol{X}-\mu)^2 \mathrm{f}(\boldsymbol{X})\, \mathrm{d}\boldsymbol{X}. \tag{2}$$

The standard deviation of *X* is the square root of the variance,

$$S_X = \sqrt{S_X^2} \quad (3)$$

In order to calculate the variance and standard deviation, the mean should be calculated first. S_x^2 is the expected value of the function $(X - \mu)^2$, which measures the square of the distance of X from its mean. Therefore, S_x^2 is sometimes called the mean square deviation, while S_x is named the root mean square deviation. S_x^2 will be larger if X tends to wander far away from its mean, and smaller if the values of X tend to cluster near its mean. The reason to take the square root in the definition of S_x is that S_x^2 is the expected value of the square of the deviation from the mean, and thus it is measured in square units. Its square root S_x, therefore, provides us a measure in ordinary units. The analysis of variance, or ANOVA as it is often referred to, is a particularly powerful way to analyse experimental data. Analysis of variance or ANOVA will allow us to test the difference between two or more means. This method is widely used in industry, science, statistics and engineering to help identify the source of potential problems in the production process and to identify whether variation in measured output values is due to variability between various manufacturing processes, or within them. By varying these factors in a predetermined pattern and analysing the output, it is possible to use statistical techniques to make an accurate assessment as to the cause of variation in a manufacturing process (McClave 2000), (Robouch 2003), (Walpole 1993), (Wolf 1997).

2.1. ONE- WAY ANOVA

One-way ANOVA is a method of statistical analysis that requires multiple experiments or readings to be taken from a source that can take on two or more different inputs or settings. One-way ANOVA performs a comparison of the means of a number of repetitions of experiments performed where a single input factor is varied at different settings or levels. The model deals with specific treatment levels and is used for testing the null hypothesis $H_0: \mu_1 = \mu_2 = \ldots = \mu_\alpha$, where μ_i represents the level mean. Basically, rejection of the null hypothesis indicates that variation in the output is due to variation between the treatment levels, but not due to random error. Rejection of the null hypothesis means that there is a difference in the output of the different levels at significance (α) and it remains to be determined (McClave 2000), (Robouch 2003), (Walpole 1993). An ANOVA table looks like as follows:

	SS	***df***	***MS***	***F***
Between	*SS(B)*	*k-1*	$\frac{SS(B)}{k-1}$	$\frac{MS(B)}{MS(W)}$
Within	*SS(W)*	*N-k*	$\frac{SS(W)}{N-k}$	
Total	*SS(W)+SS(B)*	*N-1*		
Repeatability Std Dev $\sigma_r = \sqrt{(MSW)}$				

Table 1. One-way ANOVA table

The mean of a set of samples is the total sum of all the data values Σx divided by the total sample size N, which requires all of the sample data available to be obtained, as it is usually the case, but not always (McClave 2000), (Robouch 2003), (Walpole 1993), (Wolf 1997).

$$\bar{X}_{GM} = \frac{\sum x}{N} \tag{4}$$

Another way to find the grand mean is to find the weighted average of the sample means. The weight applied is the sample size,

$$\bar{X}_{GM} = \frac{\sum n\bar{x}}{\sum n} \tag{5}$$

The total variation (not variance) includes the sum of squares of the differences of each mean with the grand mean. The whole idea behind the analysis of variance caused by the interaction between the samples is much larger when compared to the variance that occurs within each group, which results from the fact that the means are not the same. The total variation is calculated as:

$$SS\,(T) = \sum (x - \bar{X}_{GM})^2 \tag{6}$$

In between group variations, the variation due to the interaction between the samples is denoted SS (B) for the sum of squares between groups. If the sample means are close to each other (and therefore to the Grand Mean) the mean (Grand Mean) will be small. There are *k* samples involved with one data value for each sample (the sample mean), so there are k-1 degrees of freedom. The variance due to the interaction between the samples is denoted MS (B) for the mean square between groups. It is denoted by s_b^2. The between group variation is calculated as:

$$SS(B) = \sum n(\bar{x} - \bar{X}_{GM})^2 \tag{7}$$

In within group variation, the variation due to differences within individual samples is denoted SS (W) for the sum of squares within groups. Each sample is considered independently and no interaction between samples is involved. The degree of freedom is equal to the sum of the individual degrees of freedom for each sample. Since each sample has degrees of freedom equal to one point less than their sample sizes, and there are *k* samples. The total degree of freedom is k less than the total sample size $df = N\text{-}k$, and within group variation it is calculated as:

$$SS(W) = \sum df\, s^2 \tag{8}$$

The variance due to the differences within individual samples is MS (W) for the mean square within groups, which is the within group variation divided by its degrees of freedom. It is denoted by s_w^2 and it is the weighted average of the variances (weighted with degrees of freedom). Please note that an F variable is the ratio of two independent chi-square variables divided by their degrees of freedom and also remember that the F test statistic is the ratio of two sample variances. The F test statistic is calculated by dividing the between group variance by the within group variance. The degrees of freedom for the numerator are the degrees of freedom for the between group (k-1), whereas the degrees of freedom for the denominator are degrees of freedom for the within group (N-k) (McClave 2000), (Robouch 2003), (Walpole 1993), (Wolf 1997). The F statistic is calculated as:

$$F = \frac{s_b^2}{s_w^2} \tag{9}$$

Note that each mean square is just the sum of squares divided by its degrees of freedom, and the F value is the ratio of the mean squares. If the between variance is smaller than the within variance, the means are really close to each other, which means that one will fail to reject the claim that they are all equal. The degrees of freedom of the F test are in the same order as they appear in Table 1. The null hypothesis cannot be rejected if the statistic in the table is greater than the F critical value with k-1 numerator and N-k denominator degrees of freedom. The numerator of the F test statistic measures the variation between the sample means. The estimate of variance in the dominator depends only on the sample variances and is not affected by the differences among the sample means. Consequently, the sample means that are close in value result in an F test statistic that is close to 1, which leads to the conclusion that there is no significant difference among the sample means. But if the value of F is excessively large, then the claim of equal means could be rejected (The vague terms “close to 1” and “excessively large” are made objective by the corresponding P-value, which tell us whether the F test statistic is in the critical region or not). Since excessively large values of F reflect unequal means, the test is right-tailed.

More data should be collected before it is possible to assess whether the distance comes from normal distributions. ANOVA has been shown to be a very robust method when the assumption of normality is not satisfied exactly; in other words, moderate deviations from normality do not have much effect on the significance level of ANOVA F-test or confidence coefficients. Rather than spending time, energy, or money to collect additional data for this experiment in order to verify the normality assumption, we will rely on the robustness of the ANOVA method. Table 1 shows the counts, means, and variances for the data. The ANOVA table shows the results of the completely randomised analysis of variance. In this set of data, the calculated F is greater than the tabled F. Thus, for Fcrit the null hypothesis cannot be accepted (McClave 2000), (Robouch 2003), (Walpole 1993), (Wolf 1997).

3 REAL TIME KINEMATIC GPS (RTK GPS)

Real-time precise positioning is possible even when the GPS receiver is in motion, through the use of "on-the-fly" (OTF) ambiguity resolution algorithms. These systems are commonly referred to as RTK (real-time-kinematic) systems, and make feasible the use of GPS-RTK for many time-critical applications such as machine control, GPS-guided earthworks/excavations, automated haul truck operations, and other autonomous robotic navigation applications. The limitation of the single base RTK is the distance between the base receiver and the rover receiver due to distance-dependent biases, i.e. orbit bias, ionosphere bias and troposphere bias. Techniques have been developed to overcome this distance dependence, whereby the resolving of the wide lane integer ambiguities is attempted first, then using the ionosphere-free combination to resolve the integer ambiguity with the wavelength of 10.7 cm. The performance cannot be as good as with short-range-RTK, as it cannot actually be used in practice in real-time. On the other hand, the Wide Area Differential GPS (WADGPS) and the Wide Area Augmentation System (WAAS) have been investigated extensively, but they are pseudo-range based systems intended to deliver accuracies at the one metre level. Both WADGPS and WAAS require a network of master and monitor stations spread over a wide geographic area. Because the measurement biases will be modelled and corrected, the positioning accuracy will be almost independent of the inter-receiver distance (or baseline length). Carrier phase observations in these systems will generally be used to smooth the pseudo-range data.

The limitations described above have initiated the development of "network" RTK, in which the data from an entire network of RTK Base Stations are considered in real-time to allow fast and accurate rover initialisations over a large area. These techniques have been developed since the early 2000s, and the results seem very promising. Future RTK implementations/refinements will take into consideration these developments as well as the GPS modernization impacts of L2C and L5 and future GALILEO signals (Mowafy 2000), (Mowafy 1997), (Langley 1998), (Satalich 1998), (Thales Navigation 2002).

The geometrical strengths of the GPS satellite constellation at a certain location and time are numerically represented using DOP (Dilution of Precision) numbers. Note that even though the PDOP (Position Dilution of Precision), HDOP (Horizontal Dilution of Precision), and VDOP (Vertical Dilution of Precision) values actually indicate the geometric strengths for pseudo ranging, they can be considered as an indication of the strengths for the respective baseline components as well. For simplicity reasons, most RTK surveys take into account just the 3D PDOP value. Good PDOP values are <3, acceptable PDOP values range between 3 and 6, and a poor PDOP is >6 (Mowafy 2000). With efficient PDOP values, the ambiguity resolution is normally fast and reliable, and the instantaneous RTK positions are consistent. With higher PDOP values, the ambiguity resolution may take longer and become less reliable, and the RTK instantaneous positions may fluctuate more (although the mean of several epochs may still generate acceptable results). With poor PDOP values, a successful ambiguity resolution may not be achieved, and the poor geometry does not allow a reliable initialisation check. The satellite geometry can be predicted in advance using planning software. The instantaneous satellite geometry reflecting the actual satellites being used is shown on the RTK controller, and a mask value should be set to prevent working with poor geometry. As satellites move close to the horizon, observation errors become unstable and more difficult to model. The errors are mostly due to the grazing path of the signal through the troposphere and greater susceptibility to multipath.

RTK rover positioning can be done continuously during movement, and it can be positioned with a short-time occupation (typically less than 15 seconds). All GPS observations are subject to random errors (e.g. receiver measurement noise), and short-term systematic errors can affect the observations (e.g. signal multipath, ionospheric and tropospheric disturbances, signal refraction through foliage, etc.). The quick nature of RTK surveys explains that these errors can influence the final positions more than during static carrier-phase surveys with much longer observation time-spans. Longer periods of measurements reduce the impact of both random and short-term systematic errors and this is why static carrier-phase surveys are used when the best results are required. By viewing the time-series plots of static carrier phase residuals, the findings can be supported. RTK field surveyor has a limited scope of actions available to reduce the impact of random errors (observations may last for many minutes, which defeats the purpose of a normal RTK survey). Some of the short-term systematic errors can be minimized with careful attention to the antenna placement (avoiding nearby multipath reflector surfaces) and possibly to receiver configuration settings (using a higher SNR (signal-to-noise ratio) mask to prevent foliage-refracted signals from being accepted, and raising the elevation mask to minimize tropospheric & multipath errors). If it is possible and practical to re-occupy a station at a later time (e.g. using a different satellite constellation with >2 hour time difference), then a second short occupation should improve the point accuracies, which will in turn improve the reliability through an independent occupation. If the two occupations show good coordinate agreement, both results can be combined to determine the final coordinate (Mowafy 1997), Hoffmann 2000), (Lemmon 1999), (Thales navigation 2002), (Van Diggelen 1997). The limitations of the RTK system include the following:

- **Initialisation** - The receiver must be initialised in good GPS conditions up to 15 minutes before achieving sub-meter accuracy. If the receiver tracks less than 4 satellites at any given time after being initialised, it must re-initialise before achieving again the sub-meter accuracy.
- **Baseline Length** - As the distance between the base and the remote receivers grows larger, the errors observed between the GPS receivers become less and less common, degrading accuracy at the remote. Good accuracies can normally be achieved through baselines (line between base and remote) in the order of 10 - 15 km. Before GPS signals reach the antenna on the Earth, they pass through a zone of charged particles called ionosphere, which changes the speed of the signal. However, if the rover works too far from the reference station, problems may appear, particularly with initialising the RTK fixed solution. Troposphere is essentially the weather zone of our atmosphere, and droplets of water vapour may affect the speed of the signals.
- **Radio Transmission** - The base and the remote unit must maintain communications at all times in order to obtain good accuracy. Any obstructions along the propagation path will affect the signal's range. Signals might be blocked or reflected by buildings or other objects, diffracted over and around mountain peaks and ridges and the corners of structures, or even travel much longer distances than normal. Successful application of RTK depends on the radio link viability. If there is an interruption in the radio link between a reference receiver and a rover for any reason, the rover is left with an autonomous position. It is very important to set up a network of radios and repeaters, which can provide uninterrupted radio link required for the best GPS results.
- **Visibility and Multipath** - In general, at least 5 satellites must be available in order to achieve good results. Despite being less susceptible to multipath after initialisation compared to other techniques, RTK results can be seriously degraded by obstructions such as trees, fences and buildings.

- **Accuracy of Reference Point -** The absolute accuracy of the position reported by the remote receiver is only as accurate in an absolute sense as is the position of the base station coordinates (Mowafy 2000), (Langley 1998), (Thales Navigation 2002), (Van Diggelen 1997).

4 DESCRIPTION OF THE EXPERIMENTS

In the evaluation of the performance of RTK GPS method, four tests were carried out in Samandira Region, Istanbul, Turkey. The objective of the study was to assess the achievable accuracy of RTK and check the repeatability of results under different satellite configurations by using four different reference points. For this purpose, four reference points (R1, R2, 34802 and 34140) were located in the project area (see Figs. 1 and 2). Static GPS surveys were conducted to determine the coordinates of the four reference points. The measurements on the primary network were taken with at least 10 hours of observation time. The minimum elevation cut-off angle and sample rate were 15° and 10 seconds, respectively. All the measurements were carried out by using Ashtech Z Surveyor receivers. Data processing and network adjustments were conducted by making use of Ashtech Solution Software (Version 2.60). In the adjustment procedure, the ITRF 2000 coordinates of IGS point ISTA were considered as fixed (Figure 1). Table 2 lists the reference points, measurement intervals and the dates the observations were acquired (Pirti 2011).

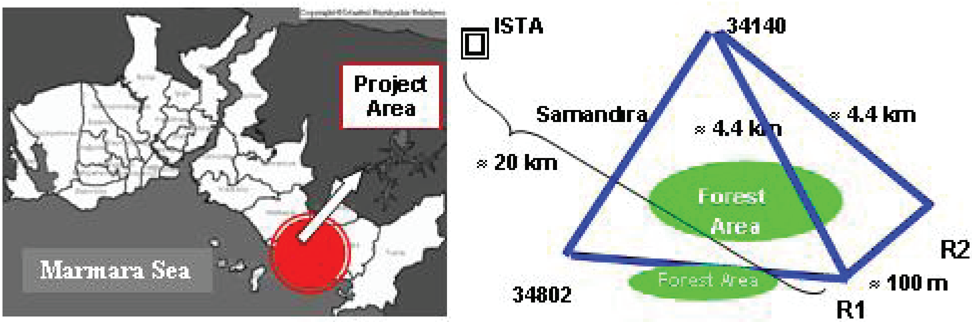

Figure 1. Project area and GPS network (Pirti 2011)

In the four tests, the GPS equipment used with the RTK surveying consisted of a pair of Ashtech Z Surveyor receivers with UHF radio modems with a power of 2 Watts. In addition, Ashtech GPS Fieldmate Software and Ashtech SSRT modem were used. The receivers were dual frequency systems with 12 channels. The data acquiring and processing rate were set to one second, with a cut-off elevation mask angle of 10 degrees. The four tests were conducted in different times during four days, with substantial changes in the satellite configuration to ensure the independence of the results (see Table 2). Throughout the testing, the number of tracked GPS satellites and their distribution were generally "normal", with 5 to 8 satellites being observed, and with PDOP ranging between 2.0 and 4.8 (Pirti 2011).

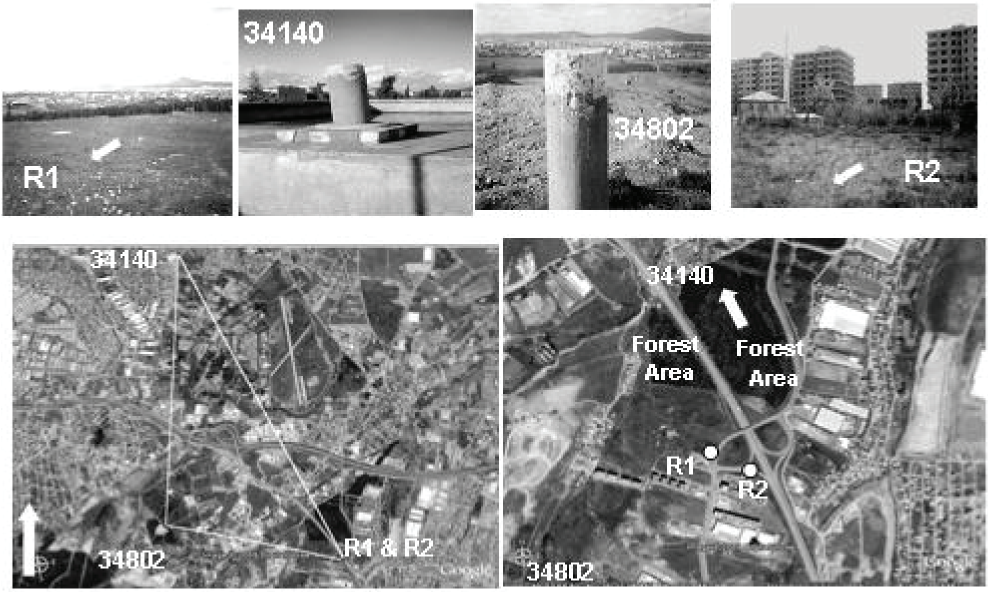

Figure 2. The four reference points in the project area (above) and the view of these reference points (Pirti 2011)

Reference Point	Date	Time interval (h)
R1	15th July 2002	09:00-11:00
R2	16th July 2002	11:00-13:00
34140	17th July 2002	15:00-17:00
34802	18th July 2002	13:00-15:00

Table 2.Time schedule of the measurements by using four reference points (Pirti 2011)

4.1. RESULTS OF THE EXPERIMENTS

Figure 3 shows the coordinate differences between RTK GPS surveys from reference points R1 and R2 (by using ISTA (IGS station)). Carrying out surveys under different satellite constellations provided better repeatability for RTK GPS surveys (see Table 2). The first test was performed taking R1 as the reference point, whereas the second test was performed using R2 as the reference point. Figure 3a also shows means and standard deviations of the estimated coordinates (i.e. coordinate differences) of the first and second tests for about 90 points. Comparing the results of these two tests, all components (northing, easting and height) of the points as separately determined by these tests seem very consistent, with changes between a few millimetres up to 5 cm. The reason for the jump marked in the bottom chart could be attributed to the effect of signal attenuation on the height component due to a forest nearby the survey point, see Fig. 3b (Pirti 2011).

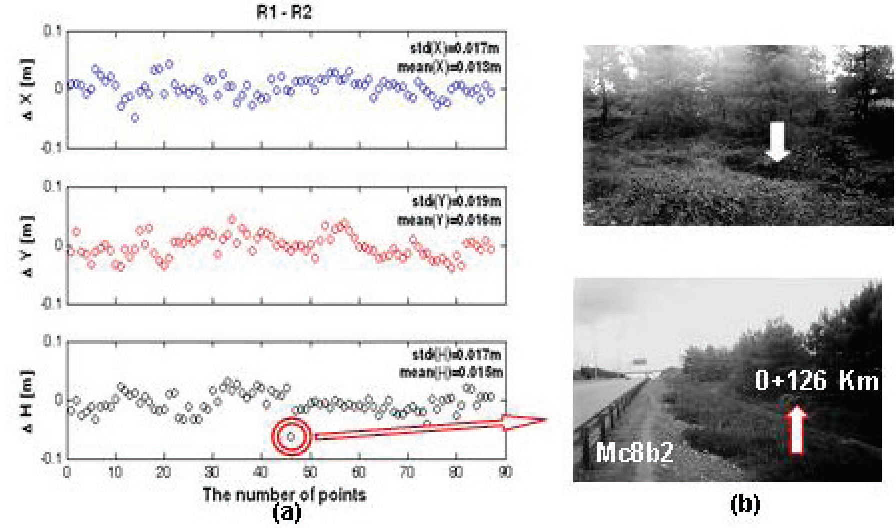

Figure 3. Comparison of the estimated coordinates from reference point R1 on Day I (15 July 2002) with the estimated coordinates from reference point R2 on Day II. (16 July 2002), (Pirti 2011)

Figures 4a and 4b show the coordinate differences between RTK GPS surveys using reference points 34140, R1, and R2. The means and the standard deviations for about 90 points (the majority of points near the forest) are also presented in Figure 4. Figure 4a shows the comparison of the estimated coordinates from reference points R1 and 34140. Figure 4b compares the estimated coordinates obtained from reference points R2 and 34140. The results of the second test presented in Figures 4a and 4b indicate that the northing and the height components show large offsets with respect to the zero mean in comparison with the east component, with changes varying between a few millimeters and 20 cm. Only a small part indicated with grey colour in Figure 4 showed minor scatter, but these points are dislocated from the others. They are located in unobstructed areas. Results indicate that ambiguities were successfully fixed throughout the survey. Five satellites were tracked, as usual. During the time span of the grey area the number of the tracked satellites was six. It is suspected that signal attenuation or blockage occurred because of the forest area and the grey part was less affected (Pirti 2011).

Figures 5a and 5b indicate the coordinate differences for the survey points measured using various reference points (i.e. R1, R2, and 34802). Please note that the surveys were taken under different satellite constellations (c.f. Table 2) so that better repeatability values for RTK GPS were provided. Figures 5a and 5b also show means and standard deviations for about 90 points for the third test. Figure 5a compares the estimated coordinates from reference point R1 with the estimated coordinates from reference point 34802. Figure 5b compares the estimated coordinates between reference points R2 and 34802. Comparing the results of the test, the horizontal coordinates of the points as separately determined by these tests seem to be consistent with changes between a few millimetres up to 5 cm. The height component was, however, less consistent and sometimes had differences by up to 10 cm at the same point between two RTK sessions (Pirti 2011).

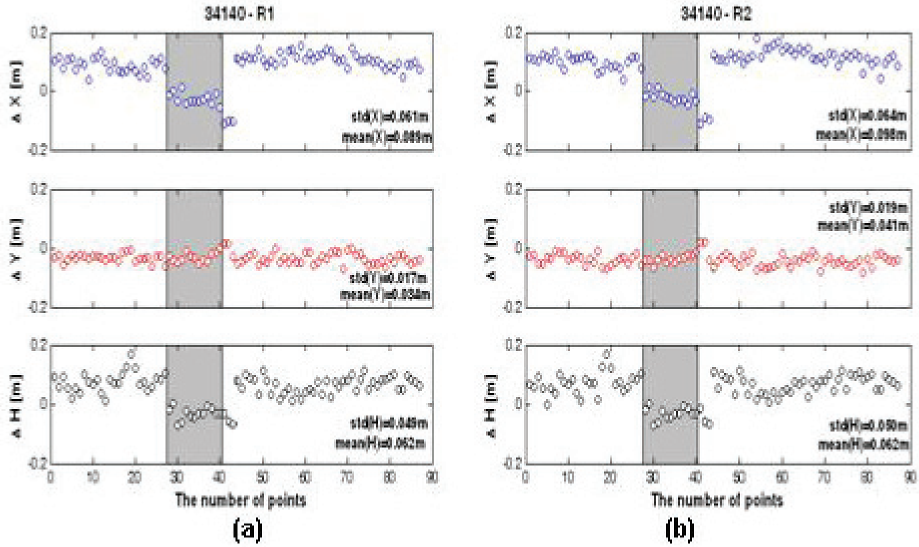

Figure 4. Comparison of the estimated coordinates from reference points R1 and R2 surveyed on Day I. (15 July 2002) and Day II. (16 July 2002), respectively, with the estimated coordinates of 34140 reference points surveyed on Day III (17 July 2002) (Pirti 2011)

In order to find out whether there were any significant differences between individual solutions (i.e. using different reference points under different satellite constellations), ANOVA was performed on the set of data presented in Figs. 3, 4 and 5. In RTK practice, repeatability is usually calculated by taking the averages of the solutions produced from different sessions (i.e. variation of satellite receiver geometry is taken into account).

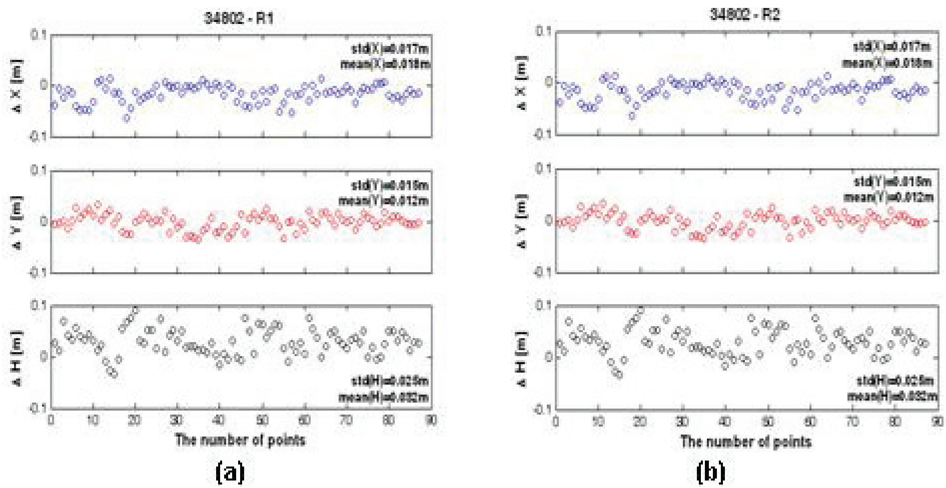

Figure 5. Comparison of 34802 coordinates surveyed on Day IV (18th July 2002) with the coordinates derived from reference points R1 and R2 surveyed on Day I and Day II (15th and 16th July 2002), respectively (Pirti 2011)

ANOVA is a relatively robust method, as it is based on statistical criteria, and can tolerate only a fair amount of gross errors. In other words, large outliers or extreme differences in variability among groups are not easily tolerated in the analysis. As a result, we propose ANOVA as an accurate method for quantifying repeatability. We assume the null hypothesis H_0: $\mu_1 = \mu_2 = ... = \mu_\alpha$,, i.e. the means from the four different tests are equal (McClave 2000), (Robouch 2003), (Walpole 1993). The test results for the hypothesis are given for each baseline component in Tables 3, 4 and 5, called ANOVA tables (see Table 1). The information here is used to calculate repeatability. The statistics is the F test, with degrees of freedom (df) and the P-Value. The P-value is the likelihood of finding as big a difference between groups as if there was no true difference. If the P-value is less than alpha (α = 0.05), the null hypothesis is rejected (i.e., the means are different) (McClave 2000), (Robouch 2003), (Walpole 1993).

ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**$F_{crit.}$**
Between Groups	0.0264	86	0.0003075	0.9945	0.4990	1.3067
Within Groups	0.1076	348	0.0003092			
Total	0.1340	434				
Repeatability Std	S_r	0.018	r = 0.05			

Table 3. ANOVA table for the Y values from the four different tests

ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**$F_{crit.}$**
Between Groups	0.1058	86	0.001230	0.5329	0.9997	1.3067
Within Groups	0.8035	348	0.002309			
Total	0.9093	434				
Repeatability Std	S_r	0.048	r = 0.14			

Table 4. ANOVA table for X values from the four different tests

ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**$F_{crit.}$**
Between Groups	0.1162	86	0.00135	1.5694	0.0026	1.3067
Within Groups	0.2996	348	0.00086			
Total	0.4158	434				
Repeatability Std	S_r	0.029	r = 0.08			

Table 5. ANOVA table for H values from the four different tests

When testing the hypothesis, the values of F_{ratio} should be checked first. If the ($F_{ratio} \geq F_{crit}$) result is significant, there is a high degree of confidence that the four solutions from the four different RTK campaigns are indeed different (Tables 3, 4 and 5). How much the F_{ratio} ($F = MS(B)/MS(W)$) is larger than 1 depends on how far apart the means actually are. The larger the value of the F statistic, the more evidence there is that at least one of the means is different from the others. Another criterion to check the hypothesis is also the P value. If the P value $\leq\alpha$, the null hypothesis of equal means is rejected. In this study, the tests were performed at the level of significance α = 0.05. Since in Table 5 $F_{ratio} \geq F_{crit}$ (1.5694>1.3067), the null hypothesis cannot be accepted. In other words, calculating one average value from the four different solutions to determine the value of H (hypothesis) is not a correct approach since the hypothesis test fails. In addition, we have P = 0.0026 for the H values, see Table 5. This means that there is a 0.26 % probability that there exists not a true difference between the groups. If this chance is low enough, then we might provisionally say that the groups are different. Usually, if the probability equals to 0.05 or less, we could claim that the difference is statistically significant. The graphical representation of the solution differences also proves the claim (Figures 4a and 4b). As for the Y values analysed in Table 3, $F_{ratio} < F_{crit}$ and the P value is larger than the significance value α = 0.05 (McClave 2000), (Robouch 2003), (Van Diggelen 1997), (Walpole 1993). Therefore, while determining the Y values, an average value could be calculated by using the solutions of the four tests, or the repeatability value (r) given in Table 3 could be used as the repeatability of the RTK work. The variation due to the measuring system, whether as a percent of study variation or as a percent of tolerance, must be less than 10 %. The guidelines for acceptance of the repeatability are (URL 1) as follows:

- 10 % or less - The measurement system is acceptable
- 10 % - 30 % - Marginal
- 30 % or greater - The measurement system needs improvement

Here interpreting the test results for the X values is probably the most difficult task, see Table 4. Although the table indicates that $F_{ratio} < F_{crit}$ and P > α = 0.05, the repeatability value calculated by ANOVA shows a large value (i.e. falls into the category of "marginal" with >10 %, as indicated above). As a result, we would not accept the null hypothesis. Graphical representation also supports this decision as the mean values of the solution differences for the X values are far larger than zero (i.e. about 10 cm, see Figure 4). Although F is smaller than Fcrit, it is not close to 1 so that it could also be an indicator of the means significantly differing from each other. All the discussions above lead to the conclusion that there are significant differences between the five groups (R1-R2, 34140-R1, 34140-R2, 34802-R1, and 34802-R2).

The variance value is another indicator for detecting the outliers in all measurements. Figure 6 shows the variances from the ANOVA test for all RTK campaigns. The graphical representation reveals that variances for the X and H solution differences show greater scatter compared to the values of Y. On the other hand, the variances of some of the X and H values approach to zero, as indicated in the grey region of Figure 6. In this region all coordinate estimates derived from 5 different solutions match, i.e. prove with statistical significance that they could be used as position information. As noted above, in cases where graphical inspection becomes difficult, ANOVA would provide statistically significant information since it is based on statistical criteria.

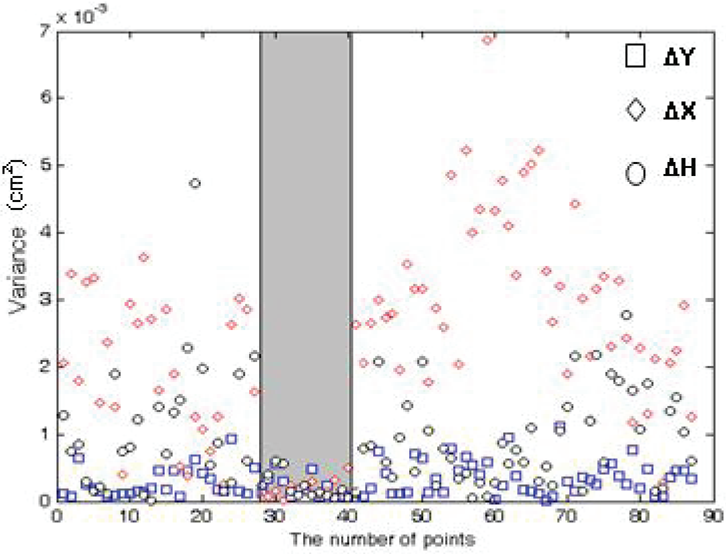

Figure 6. The graphic of the variance values of all measurements (Coordinate differences (Easting), Δ (Northing), ◊ (Height)) by using the different reference points

It has been demonstrated how ANOVA performs in distinguishing statistically significant solutions that could be used as position information from a group of observations obtained using multi-reference stations. In other words, ANOVA could also perform well in detecting solution for outliers which degrade the quality due to various factors (i.e. multipath, big DOP values, radio transmission, baseline lengths, etc.). Although inappropriate solutions encountered in this study could also be distinguished even by eye inspection, ANOVA is the best method since it relies on statistical criteria. Nowadays, multi-reference station approach is becoming standard in applying RTK GPS. One approach is to adjust corrections using the entire network of reference stations, and then send them from only one receiver to rover receivers (Raquet 2001). In case of receiving corrections from more than one reference station, ANOVA could be alternative approach in finding statistically significant results, thus improving the repeatability of RTK GPS in this sense.

5 CONCLUSIONS

RTK GPS differs from other GPS methods in that the positions could be resolved in seconds. However, this position information could easily be degraded due to various effects, such as limitations resulting from inter-receiver distances, multipath, poor satellite-receiver geometry, radio transmission, etc. (i.e. it usually results in poor ambiguity resolution). Hence, the application of the RTK GPS under different time intervals (or under varying satellite constellations) is recommended to increase the accuracy (repeatability) of the technique. Modern methods, such as multi-reference approach, suggest adjusting corrections from reference stations using the data of the entire network of reference stations. However, sending corrections from a single reference station or from a network of reference stations is still an ongoing issue and presents an ongoing work for the researchers. In case of receiving corrections from a network of reference stations, ANOVA could provide an alternative to multi-reference approach for detecting inappropriate solutions (outliers). Furthermore, ANOVA is a useful tool in increasing the repeatability of RTK GPS, as the measurements are taken in different time intervals under different satellite constellations. Being a statistical method and based on statistical testing, it would provide better judgment in choosing good solutions from different reference stations. Although there are difficulties in interpreting the results, one can take advantages of the statistical textbooks giving detailed information regarding ANOVA.

ACKNOWLEDGEMENT

Figure 2 is prepared by using Google-Earth.

REFERENCES:

El-Mowafy, A., (2000). Performance Analysis of the RTK Technique in an Urban Environment. The Australian

Surveyor, 45:1.

El-Mowafy, A., El-Maghraby, M. and Issa, M., (1997). Using Semi-Kinematic GPS Techniques for Establishing and Densifying Horizontal Control Networks in Egypt. Proceedings of International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation (KIS) Bannf, Canada, June 3-6, 1997.

Hoffmann-Wellenhof, B., Lichtenegger H., and Collins J., (2000). GPS Theory and Practice. Fifth revised edition, Springer-Verlag Wien-New York.

Langley, R., (1998). RTK GPS. GPS World, 9(9).

Lemmon, T.R. and Gerdan G., (1999). The Influence of the Number of Satellites on the Accuracy of RTK GPS Positions. The Australian Surveyor. 44(1).

McClave, T.J., Sinccich, T., (2000). Statistics, Eight Edition, Prentice Hall, Upper Saddle River, New Jersey, ISBN 0-13-022329-8.

Pirti, A. (2011). Evaluating the repeatability of RTK GPS, Survey Review Journal, Vol.43, No.320, April.

Robouch, P., Bulska, E., Duta, S., Lauwaars, M., Leıto, Majcen, I.N., Norgaard, J., Suchanek, M., Vassileva, E. and Taylor, P., (2003). TrainMiC - Training in Metrology in Chemistry, European Commission

Directorate General Joint Research Centre Institute for Reference Materials and Measurements, Belgium (tera.chem. ut.ee/˜ivo/metro/TrainMiC-EU-Rep.pdf)

Raquet, J., Lachapelle, G., (2001). RTK Positioning with Multiple Reference Stations. GPS World, April.

Satalich, J. and Rickeston, R., (1998). Field Test of Trimble 4000 Real-Time Kinematic GPS Survey System. Journal of Surveying Engineering (ASCE), 124(1).

Thales Navigation, (2002). Z Family (Ashtech) GPS Receivers Technical Reference Manual,(www.thalesnavigation. com/en/support/apptechnotes.asp#18), (ftp.thalesnavigation.com/Reference Manuals/Z-Max Ops&Aps Manuals)

Van Diggelen, F., (1997). "GPS and GPS+GLONASS RTK", The International Technical Meeting of the Institute of Navigation, ION-GPS 97, Kansas City, MO, September.

Walpole, R.E., Myers, R.H., (1993). Probability and Statistics for Engineers and Scientists. Fifth Edition, Macmillan Publishing Company, New York.

Wolf R.P., Ghilani D.G., (1997). Adjustment Computations Statistics and Least Squares in Surveying and GIS. 3rd Edition, Wiley Series in Surveying and Boundary Control, ISBN: 0-471-16833-5, USA.

URL1: www.minitabtraining.com/Downloads/TrainingSampleMeasurementSystems.pdf, www.bamr.co.za/randr.shtml (The guidelines for acceptance of Gauge Repeatability and Reproducibility, AIAG guidelines for the gage R&R table)

APPENDIX:

X_{GM}: The grand mean of a set of samples is the total of all the data values divided by the total sample size

SS(W): The variation due to differences within individual samples, denoted SS(W) for the sum of squares within groups.

SS(B): The variation due to the interaction between the samples is denoted SS(B) for the sum of squares between groups

SS(T): The total variation (not variance) is comprised of the sum of squares of the differences of each mean with the grand mean

MS(W): The variance due to the differences within individual samples is denoted MS(W) for the mean square within groups

MS(B): The variance due to the interaction between the samples is denoted MS(B) for the mean square between groups

df: Refers to the number of degrees of freedom in the particular SS defined on the same line

N-k: The degrees of freedom for the within group

k-1: Degrees of freedom

F: The F test statistic is found by dividing the between group variance by the within group variance. The degrees of freedom for the numerator are the degrees of freedom for the between group (k-1) and the degrees of freedom for the denominator are the degrees of freedom for the within group (N-k). The F variable is the ratio of two independent chi-square variables divided by their respective degrees of freedom. Also recall that the F test statistic is the ratio of two sample variances.

Received to publication: 25 July 2011
Accepted: 7 August 2012

Associated Prof. Dr. Atınç Pirti
Yıldız Technical University
Civil Faculty, Surveying Department
Davutpasa Campus
Esenler
Istanbul-TURKEY
E-mail: atinc@yildiz.edu.tr