VERIFICATION OF RAILROAD TRACK HEIGHT MEASUREMENT METHODS BASED ON STATISTIC ELABORATION OF THE RESULTS

Tomasz Owerko, Michał Strach
University of Mining and Metallurgy, Faculty of Mining, Surveying and Environmental Engineering

1. INTRODUCTION

The dynamic development of railroads is connected to the construction of new lines, as well as the modernization and reconstruction of existing rail tracks. This situation forces the verification of existing and an introduction of new measurement technologies on the geodesists. Geodetic activity connected to construction investment service is divided into realization measurements and afterward inventory. Depending on the used equipment and measurement techniques level coordinates and/or height point ordinates are appointed. Currently, because of the economical (time and work absorbing) and equipment (possibility of working with precise measuring devices) regards, the chosen methods are ones which allow elaboration of spatial coordinates. The currently dominant technology is 3D polar measurement with use of electronic tachometers. Lately, the geodesists use the GPS satellite measuring equipment. The decisive factors on if a given method can be employed are its preciseness coordinates. In case of railroad track axle measurements they cannot exceed the values included in Instruction D-19. To determine which of the measurement techniques is precise enough for defining the height of the track's levelling, it was necessary to verify it on the foundation of the statistic elaboration of the results. Analysis for the data comes from measurements executed on a 1020 meter segment of a railroad track.

2. TECHNOLOGY OF EXPERIMENTAL MEASUREMENTS AND OBSERVATION REDUCTION

The experimental studies were conducted on an active segment of a railroad track between Krakow Plaszow and Skawina nearby the Krakow Sidzina train station. Within the studied segment coordinates of 205 points were marked, representing the real route axel in profiles every 5 meters. The studies were done in one measuring cycle with both the GPS satellite method as well as the 3D polar method. During the studies real time kinematics measurements were made, using the Leica GPS SR530 receivers with choke ring AT503 antennas as well as a precise TCA 2003 electronic tachymeter of the same company. The tachymeter's characteristic is standard deviation of direction measurement by 0,5" ($1,5^{cc}$) and precise direction reading of $0,1^{cc}$. The distance measurement preciseness using this device is 1 mm + 1 ppm. The mentioned GPS satellite measurement device is characterized with precise measurement with the *stop and go* method on a level of 5÷10 mm + 1 ppm. The antennas for receiving GPS satellite signals and the prismatic reflector have been placed on a specially constructed measuring cart [Strach, 2003]. Thanks to this solution it was possible to execute studies using both systems at the same time. To be able to determine the quality of ordinates of

height points marked by satellite technique, it was necessary to compare them to coordinates of the same points, acquired the 3D polar measurement method. With this as an aim, it was necessary to do project the observed points (prism middle and centre of antenna) to the level determined by the rail tracks' progress. Geometric reductions were introduced to the coordinates of measured points, taking into account their heights in regard to the levelling of the stocked track.

Additionally, sample measurements were made, doing precise geometric levelling of points representing the track's axle. The received results were an analysis regard of preciseness of the RTK GPS method and the 3D polar method. The NA 3003 code leveller was used for these studies.

The measurements of point heights of axles with the GPS technique as well as the classic method were made in regard to different arrangement configurations. It was necessary to make another adjustment, taking the difference in undulation (distance between the geoid and the ellipsoid) into account, which was 4,6 cm for radical points of the studied track's length of 1020 m. The reduction corrections for the ellipsoid heights of specific points were determined from linear dependency accounting for distance between the points and the ends of the studied segment.

Finally, after introduction of reduction and corrections, three sets of height ordinates of the same points of the route's axle were acquired. Proceeding to the analysis of preciseness of the specific measurement methods the superiorities of every method were calculated according to the formulas (1).

[illegible] (1)

Next, from superiorities apportioned in this way, the difference between specific measurement methods was calculated. The formulas used here are:

[illegible] (2)

Among the calculated data for every point of the 1020 meter experimental segment, there have also been protruding observations. They concerned points determined by the RTK method, with the following numbers: 11405, 11425, 11480, 11660, 11700, 11705, 11735, 11745, 11750, 12105, 12135,12140,12145, 12225, 12405. For the disposed points differences in superiorities between the GPS method and geometric levelling were, in order: -262, 16, 27, 17, 17, -17, -18, -18, -20, 19, 21, -31, 21, -19, 17 millimetres. Based on the notes taken during the measurements it can be assumed that the influence on the decreasing quality of the coordinates were singular high terrain obstacles (trees, roofing of station).

3. STATISTIC ELABORATION OF THE RESULTS

3.1 Decomposition compatibility tests

They realize tasks tied to the results of geodetic measurements elaboration. We usually have to estimate the average value and its standard deviation or estimating the trust section for the above mentioned value. So the acquired results may be trustworthy, the

calculations have to be based on data, in which we eliminate the occurrence of thick and systematic mistakes.

But to be able to properly acquire statistic propositions concerning the results of experiment measurements and the mentioned above parameters describing the studied population, we must use proper statistic equipment.

Usage of commonly adapted statistic tests which allow estimation of average value and standard deviation is many times burdened by the assumption of the studied sample being from a population of a regular decomposition. There is a line of tests which allow verification of the above hypothesis. Overall, we can divide them into five basic types:

- **tests type χ_2**
- **tests type EDF (Empirical Distribution Function)**
- **tests based on the asymmetry and kurtosis coordinate**
- **regression tests**
- **others**

With regard to the large differences in the properties of each of the available tests, the decision which to use should be made based on its values and shortcomings.

The most frequently used compatibility test from the χ_2 test family is the Pearson test. The basic distinction of the Pearson test is a lack of direct connection with the chosen decomposition. Furthermore, we do not need to know the description parameters (average value, variance, and so on), meaning that we can estimate them based on the studied sample. In case of experimental studies the second trait is especially demanded. The test divides the observed sample into categories and compares the number of observations and awaited observations in specific categories. The categories can be chosen so that they have the exact same possibility or exact length.

During the adaptation of the Pearson χ_2 test, to check the hypothesis concerning regular decomposition, it is recommended to divide into categories of exact possibilities. If the studied decomposition is depending on only one parameter (simple hypothesis), we first use series test for signs of sequential differences C_i -N_i where:

- **C_i – number of observations counted in category i**
- **E_i – number of awaited observations in category i**

When the signs' random series hypothesis has been disposed of we are able to adapt the χ_2 test. We do not adapt the test for assembled hypothesis.

A different approach to the problem is presented by the EDF family of tests (empirical distribution function) – meaning being based on the empiric cumulative distribution function.

Different to the χ^2 type tests, the EDF tests do not divide the studied sample into categories. In aim of building an empiric cumulative distribution function, they arrange the sample growingly and then calculate the maximum difference between the theoretical and empiric cumulative distribution function.

Tests that base themselves on this rule can be divided into two categories (Ahmed M. Sultan, Hala Mahmoud Khaleel 2005). First is the supreme category, of which the most known representative is Kołmogorov-Smirnov. Among the basic values of the Kołmogorov- Smirnov test in proportion to the χ^2 test are:

– Possibility of adaptation with a smaller sample amount,

– With the same sample amount the Kołmogorov test is much more sensitive.

Its extension is a modification of Lilliefors, assuming that the descriptive parameters (average value and standard deviations) will be estimated from the sample and are previously unknown. This modification relied on calculation of the critical decomposition values using the Monte Carlo method. Thank to this the possibility of using this statistic in studies increased greatly – because we very rarely possess information about the true value of these parameters, which is assumed by the Kolmogorov-Smirnov statistic (as well as for the Pearson χ_2 test).

The next group of tests, which, in the following calculations, are represented by Cramer von Mises and Anderson Darling tests, can be inducted to the square category. An important trait of the Anderson-Darling test, in comparison to the Cramer von Mises statistic as well as the Komogołov-Smirnov test, is that observations differentiating from the average value are put under more emphasis (so called decomposition tails).

The test based on asymmetry and kurtosis coordinates as well as the previously mentioned Kołmogorov tests (the version modified by Lilliefors) or the Pearson χ_2 have two common traits: first, they are overall tests for compatibility, and second, they do not require complicated calculations.

Taking into account the fact that in geodesy we are usually interested in studying the compatibility with regular decomposition (as in the following article) as well as the general accessibility of computers, the above traits are not very valuable. The test which represents a different philosophy, meaning that it specializes in testing regular decomposition, but requires complicated calculations, is the Shapiro-Wilk test. Commonly noted as one of the best for this aim, gives good results for samples with a number of over 2000.

3.2 Behaviour of chosen tests regarding decomposition compatibility of measurement groups

In the described incident, there is a very interesting concurrence (or lack of) in the results achieved by using different measurement methods.

The calculations were performed for four data groups:

A. Comparison of height of points acquired by the GPS and TPS technologies – the entire group (disposed of only the thick mistakes, comp. formula 2)

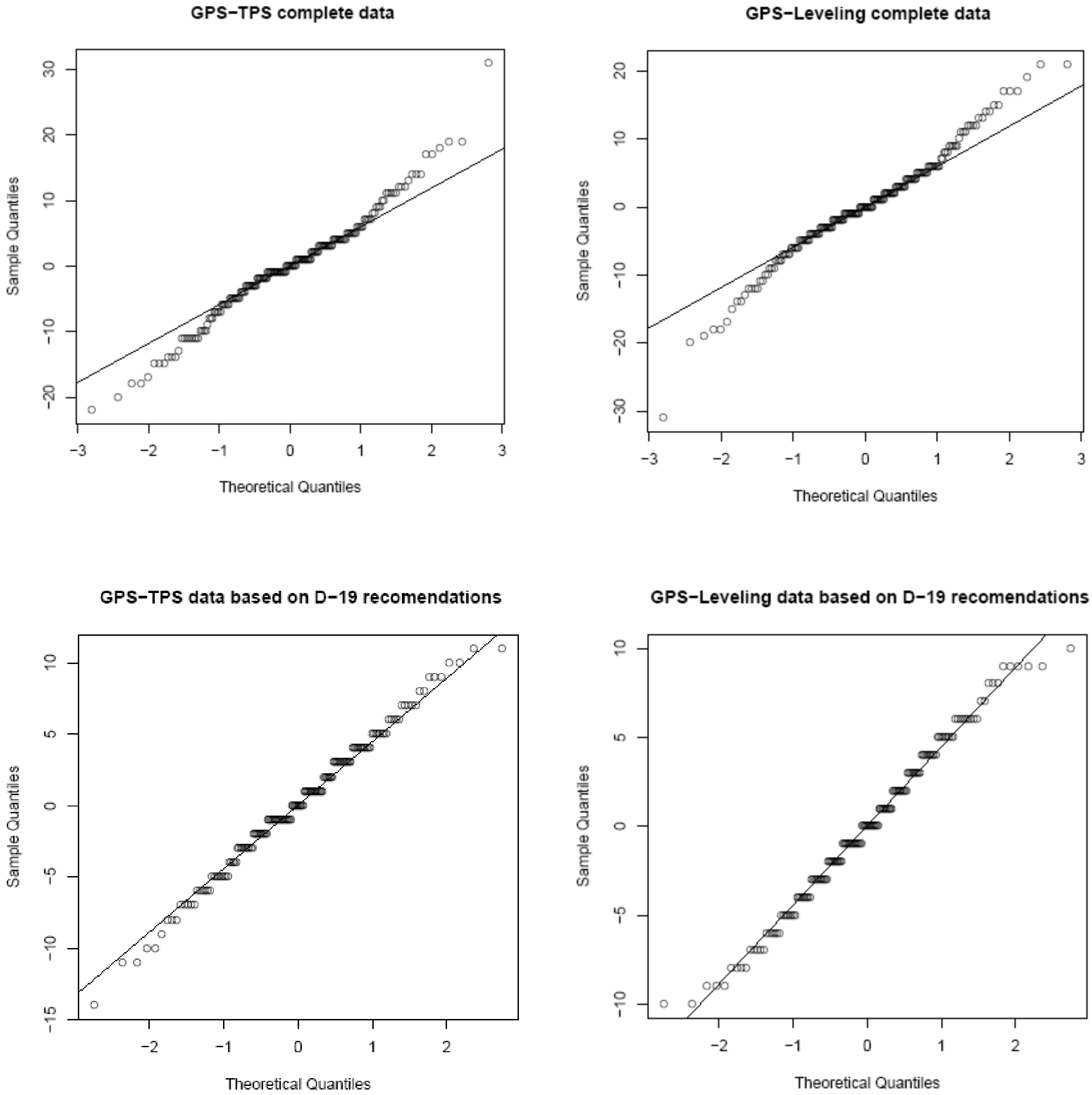
B. Comparison of height of points acquired by the GPS and levelling technologies – the entire group (disposed of only the thick mistakes, comp. formula 2)

C. Comparison of height of points acquired by the GPS and TPS technologies – group of points meeting the criteria included in instruction D-19 (deviance between point ordinates of the track from the tachometric and levelling measurement cannot exceed ±1cm) [§6 point 9] (comp. formula 2)

D. Comparison of height of points acquired by the GPS and levelling technologies – group of points meeting the criteria included in instruction D-19 [§6 point 9] (comp. formula 2)

Table 1 – Results of studying compatibility of samples with regular decomposition.

Group name	Shapiro - Wilk		Lilliefors		Pearson χ^2		Cramer-von Mises		Anderson-Darling	
	Statistic value.	p value	Statistic value.	p value	Statistic value.	p value	Statistic value.	p value	Statistic value.	p value
A	**0,976**	**0,002**	**0,0866**	**0,001**	**42,667**	**0,0001**	**0,310**	**0,001**	**1,638**	**0,001**
B	**0,976**	**0,002**	**0,079**	**0,004**	**31,333**	**0,005**	**0,2815**	**0,001**	**1,552**	**0,001**
C	**0,991**	**0,469**	**0,072**	**0,033**	**31,590**	**0,003**	**0,093**	**0,141**	**0,488**	**0,220**
D	**0,987**	**0,155**	**0,058**	**0,197**	**39,301**	**0,0001**	**0,077**	**0,224**	**0,479**	**0,231**



Drawing 1. Charts of quantile-quantile type for specific data groups.

Taking the results accumulated in table 1 it is advisable to notice the fact, that regardless of the test of choice for full observation groups TPS-GPS as well as GPS-levelling, the hypothesis that the studied variable is subject to regular decomposition is unequivocally disposed of. This of course has a direct impact on the remaining calculations.

The situation is not so obvious for analogical groups meeting the criteria included in instruction D-19 [§6 point 9]. Taking into consideration the calculated *p* value in specific tests, we notice that for the GPS-TPS group (signature C) three of five executed tests dispose of the zero hypothesis of compatibility with regular decomposition. The authors, taking into consideration the force of specific tests and quantity of the sample are leaning toward accepting the described hypothesis. The basic message toward this solution is the fact that with the sample's quantity in the described experiment (under 200) a very good result is achieved by the Shapiro – Wilk test. Additionally, the hypothesis is confirmed by the Anderson-Darling test, specifically sensitive to observations found in the so called decomposition tail. For the GPS-levelling group (signature D) four of five executed tests are for acceptance of the zero hypotheses. Below is the graphic of experimental data decomposition (in order, groups: A,B,C,D) illustrated with a quantile-regular chart.

3.3 Singular trial tests

Considering results from chart 1 with an assumed average value in trials for specific data groups the following calculations have been made:

- **For groups in which the compatible with regular decomposition hypothesis was disposed, estimation of the average was based on the Wilcoxon rank test.**
- **For groups in which the compatible with regular decomposition hypothesis was not disposed, estimation of the average was based on the *t*-Student test.**
- **In both the Wilcoxon and the *t*-Student test the following hypothesis was checked: average value of height differences appointed by the studied methods (Depending on the group its either TC-GPS or GPS-levelling) amounts to 0. The trust interval was set at a level of 95%.**

Table 2 – Testing the hypothesis: average value from trial $x_{śr}$=0 mm for specific data groups.

Data group	Wilcoxon rank test		*t*-Student test	
	Statistic value	p value	Statistic value	p value
A	9031,5	0,9574	---------------------	---------------------
B	8915,5	0,7650	---------------------	---------------------
C	---------------------	---------------------	0,1982	0,8432
D	---------------------	---------------------	-0,0864	0,9313

As seen in every data group there are no reasons to dispose of the average value hypothesis. During the description of the possibilities of the RTK-GPS technique for track regulation, based on the above drawing there is no way to make sure of the

systematic movement of the data group acquired by this method in comparison to the heights assumed as real (precise geometric or trigonometric levelling).

3.4 Tests for two trials

Result compatibility and homogeneous preciseness of acquired railroad track heights was studied based on the statistic comparison of chosen data groups.

Table 3 – Estimated average values and standard deviation of the trial.

Data group	Estimated average value	Estimated standard deviation
A	-0,04 mm	7,67 mm
B	0,11 mm	7,69 mm
C	0,07 mm	4,70 mm
D	-0,03 mm	4,49 mm

Statistic estimation of the homogeneous variance based on, respectably, the A and B as well as C and D, was executed with the popular F-Snedecör test. The tested hypothesis was that the variance ratio is equal to 1.

Table 4 – Results of the F-Snedecör test (variance comparison) as well as the t-Student test for two trials (comparison of averages in case of uncertainty in essential differences between variances).

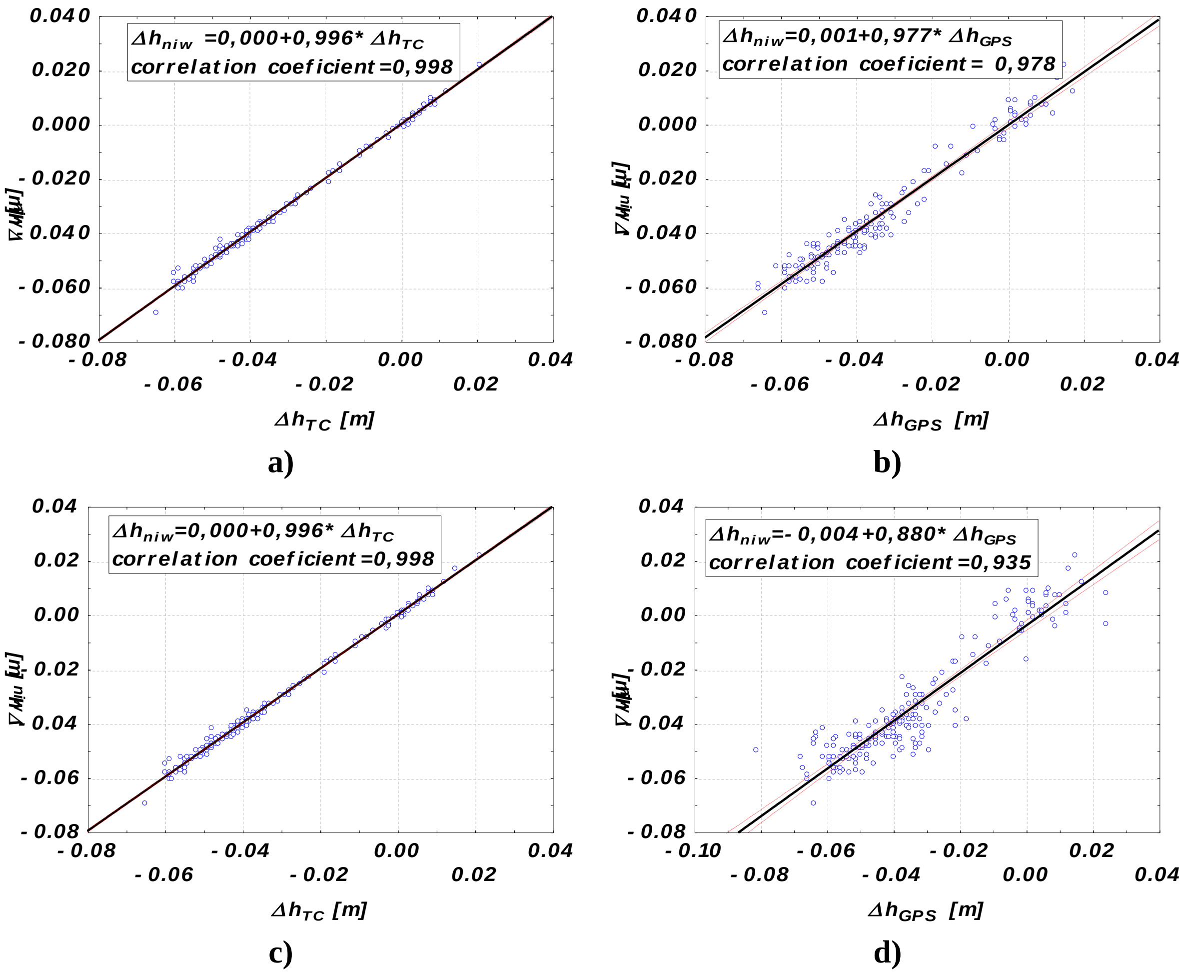
Data groups	F-Snedecör test	t-Student test
A i B	p = 0,97	p = 0,83
C i D	p = 0,56	p = 0,84

Considering the above results, it is seen that with both geometric and trigonometric levelling the RTK GPS method behaves similarly (drawing 4). The weakness of this technique concerning the determination of track ordinates is shown in drawings 2 and 3. A large variance and not being subjected to regular decomposition accounts for a lack of possible acquirement of compact data on the preciseness level anticipated in instruction D-19 [§6 point 9]. Simultaneously, based on the results of drawings 3 and 4, the activity of the systematic ordinate that causes negative influence on the acquired measurement values cannot be ascertained. The histograms of result decomposition for specific methods are symmetrical, which is proof of the above assumption.

3.5 Correlations between analyzed trials

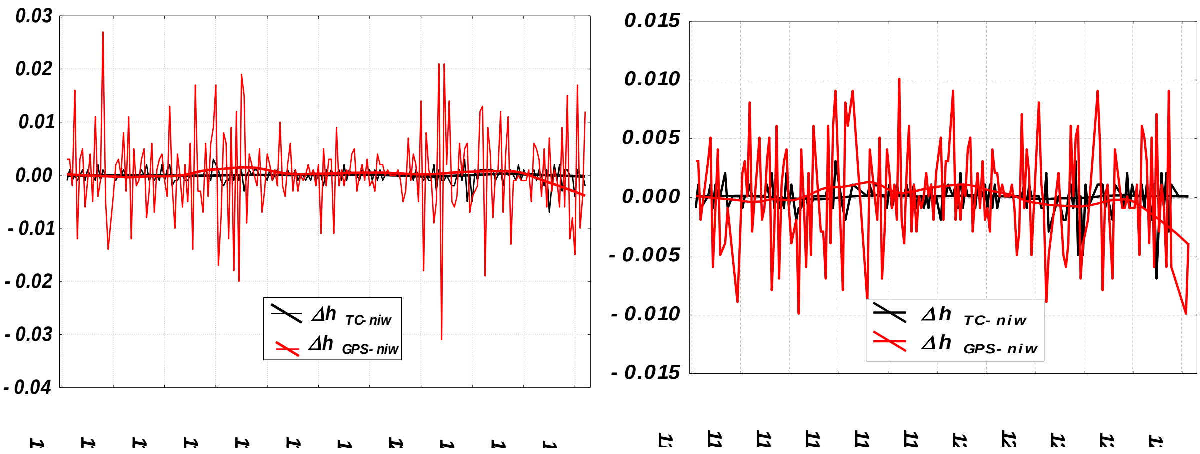
The further stage of statistical analysis was connected to dependency between specific variables, which are superiority values (Δh_{niw}, Δh_{TC} i Δh_{GPS}). Correlative dispersion charts of all analyzed data (drawing 2c and d) as well as that which met the criteria given in instruction D-19 (drawing 2a and b) were produced. These charts provided overall information of force and relations between calculated values of height differences. The points associated with specific values of superiorities in the dispersion chart (drawing 2a and c) are placed in close neighbourhood of regression lines acquired

from data Δh_{niw} i Δh_{TC}. This decomposition proves the almost full linear compatibility between the studied values. The data concerning the RTK GPS technique measurements is somewhat different. There is a clearly seen large observation group differing from the regression lines (drawing 2b and d).



Drawing 2. Correlative chart of the dispersion along with a 95% trust sector for the values of: a) Δh_{TC} i Δh_{niw} ≤10mm, b) Δh_{GPS} i Δh_{niw} ≤10mm, c) all Δh_{TC} i Δh_{niw}, d) all Δh_{GPS} i Δh_{niw}.

The strength of co-dependency of two variables can also be visualized numerically with the aid of the Pearson coordinate of linear correlation. The correlation signature of the coordinate provides information about the correlation's direction, and its emphatic value evidences the force of the relation. If the value of the coordinate is zero, it means that there is a complete lack of correlate relation between the studied variables. If the correlation value of the coordinate is one, then the correlate dependency changes to functional dependency (linear function). The calculated coordinates amounted to, respectively, 0,998 for analyzed superiority values Δh_{niw} and Δh_{TC} (drawing 2a and c) and for values Δh_{niw} and Δh_{GPS}: 0,935 and 0,978 in dependence of the analyzed data (drawing 2d and b). The acquired values of correlation coordinates confirm the assumptions regarding compatibility of acquired data. Therefore there is an almost full co-dependency (harmony) of measurements acquired by use of the polar method with geometric levelling. The linear correlation coordinate with a value of 0,935 also confirms almost full correlation. This description is usually used in statistic deduction (Christensen, 1996). However, it should be noted that in the analyzed case, preciseness on a millimetric level is important. With the objective of making final conclusions on the subject of measurement quality using the above methods, further analysis was executed. Superiority differences between data calculated by different measurement methods were assigned with this objective in mind. For this, formulas were adopted (2). Based on calculated values, charts of superiority differences $\Delta h_{TC\text{-}niw}$ i $\Delta h_{GPS\text{-}niw}$ (drawing 3) were produced. For data presented on charts the Lowess method of smoothing was used. In this method, there is a local dependency model adapted to every point by the method of polynomial regression, accounting for neighbouring points. This method is also called the immune weighed local regression. Smoothed data usually provide a clearer picture of overall dependency between two variables. On both charts, a very small dispersion of values $\Delta h_{TC\text{-}niw}$ is again noticed against a background of large hesitation of data $\Delta h_{GPS\text{-}niw}$.



a) b)

Drawing 3. Charts of superiority differences $\Delta h_{TC\text{-}niw}$ and $\Delta h_{GPS\text{-}niw}$ with smoothing function of the Lowess method for the values of: a) all, b) ≤10mm.

4. CONCLUSIONS

Verification of railroad track height measurement methods based on statistic elaboration of the results

In the research, a wide spectrum of possible use of equipment for statistic analysis was presented. In the herein research they were used for verification of railroad track height measurement methods. Based on the executed statistic analysis for data acquired from geometric levelling and GPS measurements meeting the preciseness criteria of ±10 mm, the following conclusions have been made:

1. Very visible difference in behaviour of χ^2 type test and Shapiro-Wilk test.
2. An essentially more sensitive K-S test(in comparison to Pearson χ^2) gives, in this example, results compatible with the Shapiro-Wilk test.
3. Lack of difference in results between the K-S test and the K-S test with Lilliefors' corrections (despite the fact that in the second case the description factors are estimated).

4. The studied decomposition does not show a decomposition character with so called ”heavy tails”, because the Anderson Darling test does not produce results far off (for example) the Lilliefors test, and is nominally sensitive to ”heavy tails”.

For data of acquired tachometric and GPS satellite measurements, meeting the preciseness criteria of ±10 mm, the following conclusions have been made:
1. The Shapiro-Wilk test does not dispose of normality hypothesis and, in light of following facts, may present a serious reason for making the decision.
2. There is a very big difference between specific tests based on the empirical distribution function, the Lilliefors test disposes of the zero hypothesis while (for example) the above-mentioned, sensitive to decomposition tails Anderson-Darling test does not.
3. The K-S test disposes of the zero hypotheses, which is rather no surprise to us – therefore two of five tests dispose of the zero hypotheses.

Analyses of correlation between three groups of analyzed data, as well as charts of superiority differences prove that there is a complete concurrence between levelling and tachymetry. Precise tachometric measurements are correct and economically valuable enough to continue work on three-dimensional regulation of railroad tracks including the service of high-speed rail construction. Unfortunately, this assumption as unacceptable following deeper analysis of height measurements using the GPS satellite technique. As far as the GPS system not being precise enough for correct rail height measurement, the experience of the authors allows to confirm its usefulness in situational measurements.

REFERENCES

Christensen R. (1996), ***Analysis of Variance, Design and Regression*****. Chapman and Hall, London.**

Esveld C. (2001), ***Modern Railway Track, 2nd Edition*****. Delft University of Technology, Koninklijke van de Garde BV.**

Gocał J., Lenda G., Strach M., Uznański A. (2003), ***Experimental Inventory Measurements of Railway Tracks by a RTK GPS Method*****. Proceedings of the Geodesy and Environmental Engineering Commission, Geodesy 40, str. 15–33. Polish Academy of Sciences the Kraków Section.**

Instruction D19 (2000), ***Instruction of organization and measuring in railroad geodesy*****. Enclosure to Bylaw number 144 of the PKP Administration Board from the day of 23.10.2000, Warsaw.**

Strach M. (2002), ***An assessment of the 3D polar method as applied to the adjustment of the railway track axis in the vertical plane*****. Proceedings of the Geodesy and Environmental Engineering Commission, Geodesy 39, pp. 51-63. Polish Academy of Sciences the Kraków Section.**

Strach, M., (2003). An assessment of usability the real time kinematics GPS technique to adjusting the railway track axis. Doctorate thesis. AGH, Kraków.

Kamys B., (2005), Statistic Methods of Measurement Elaboration – 1, Exposition for 1st year physics students”.

Ahmed M. Sultan, Hala Mahmoud Khaleel (2005): ***A New Modified Goodness of Fit Tests For Type 2 Censored Sample from Normal Population, Mathematical and Computational Applications,*** **Vol. 10, No. 2, pp. 165-177.**