Research Article

Dawid Kudas* and Agnieszka Wnęk

Validation of the number of tie vectors in post-processing using the method of frequency in a centric cube

https://doi.org/10.1515/geo-2020-0057
received August 6, 2019; accepted January 20, 2020

Abstract: The systems belonging to the European Position Determination System (EUPOS) are obliged to have a sub-service for automatic post-processing (APPS) of static satellite user observations based on observations registered by reference stations. The optimization of the number of tie vectors to reference stations is determined by the authors of APPS algorithms and is a key issue. The study validates the determination of the position of a point based on a smaller number of tie vectors than for the standard solution provided by a sub-service on an example of APPS of the Polish Active Geodetic Network (ASG-EUPOS). It was shown that the reduction of the number of tie vectors by one does not result in a significant deterioration of the accuracy of the obtained coordinates. The other objective of this study is to propose a method for comparing 3D sets. This method is based on 3D figures of assessment and a modification of a method for comparing text sets. The proposed method was validated by comparison with the Spearman correlation coefficient and Mahalanobis distance. The usefulness of the method for similarity and dissimilarity analyses of 3D sets has been demonstrated with an example of comparing the combination of reference vectors with a standard solution of APPS of ASG-EUPOS.

Keywords: EUPOS, POZGEO, post-processing, GPS signals, position frequency matrix, coordinate sets, cube

1 Introduction

European Position Determination System (EUPOS) was initiated in 2002 by the Berlin Senate Department for Urban Development and the European Academy of the Urban Development, Berlin. The goal of the project is to establish a uniform terrestrial system to support the Global Navigation Satellite System (GNSS) based on a multifunctional network of permanent reference stations that continuously register satellite signals. EUPOS encompasses Central European and Eastern European countries. These countries undertook to build and develop their national networks of reference stations under uniform standards and guidelines [1–5]: the network consists of permanent multifunctional reference Differential Global Navigation Satellite System (DGNSS) stations; the maximum distance between the stations does not exceed 70 km; the coordinates of the stations are determined with the highest precision in a geodetic reference frame, which is an implementation of ETRS'89, and in national coordinate reference frames; and in order to improve the geometry of the network and ensure sufficient accuracy near borders, the national networks use observations from stations of other positioning systems participating in EUPOS that are located in border areas. The institutions taking part in the project undertook to develop and make available in their national systems three sub-services [4]: EUPOS DGNSS, EUPOS Network Real Time Kinematic, and EUPOS Geodetic. EUPOS Geodetic is the sub-service for automatic post-processing (APPS) of code and carrier phase static or kinematic measurements based on RINEX 2.11 or 3.0 observation files sent to it. The sub-service should guarantee decimetre and sub-centimetre accuracy depending on the measurement equipment, duration of the measurement, and conditions of the measurement. The institutions partnering under EUPOS approached this obligation differently. Some national positioning systems use ready-made algorithms for APPS of static observations developed by leading geodesy companies such as

* **Corresponding author: Dawid Kudas,** Department of Land Surveying, University of Agriculture in Krakow, Krakow, Malopolska, Poland, e-mail: dawid.kudas@urk.edu.pl
Agnieszka Wnęk: Department of Land Surveying, University of Agriculture in Krakow, Krakow, Malopolska, Poland, e-mail: ag.wnek@urk.edu.pl

Trimble (e.g., SKPOS in Slovakia) or Leica (e.g., CZEPOS in Czech). Others based their EUPOS Geodetic sub-service on their own computational algorithms (e.g., Active Geodetic Network-EUPOS [ASG-EUPOS] in Poland). Accuracies possible thanks to real-time corrections in individual national positioning systems have been studied extensively [6–11]. Still, there are no publications on the determination of the accuracy, precision, and repeatability of the solutions offered by the APPS sub-services in individual national systems.

The ASG-EUPOS activated in 2008 in Poland is an example of a national network based on EUPOS standards [12]. Today, ASG-EUPOS consists of 103 reference stations located in Poland and 24 foreign reference stations (Figure 1). More information about spatial pattern analysis of the ASG-EUPOS stations in Poland can be found in ref. [13].

ASG-EUPOS stations make up the first-order and second-order horizontal geodetic networks. ASG-EUPOS and solutions that employ the system's sub-services are widely used, not only in geodesy. The data recorded at ASG-EUPOS stations are used in research such as geodynamic research [14–18] and ionosphere monitoring [19]. ASG-EUPOS provides the user with real-time (NAWGEO and KODGIS/NAWGIS) and post-processing (POZGEO and POZGEO-D) sub-services. The purpose of POZGEO-D is to make available to users historical GNSS observation data stored at ASG-EUPOS reference stations. POZGEO is dedicated to the adjustment of carrier phase Global Positioning System (GPS) observations from static measurements using APPS. The accuracy of positioning using POZGEO depends on the measurement conditions and duration. It ranges from 0.01 to 0.10 m [20].

The APPS POZGEO module provides relative positioning based on a network of permanent reference stations of ASG-EUPOS; it offers local support. APPS requires not less than 900 epochs in the user's observation file. POZGEO uses a completely Polish algorithm [21] linked to a sub-service of Trimble. To determine the position of a point, APPS POZGEO selects six optimum base vectors from among the ten shortest vectors between the point and the nearby reference stations [22]. Finally, the base vector is determined using

Figure 1: Distribution of the ASG-EUPOS system stations in Poland (www.asgeupos.pl, state on 15 March 2018).

one of the two methods depending on the duration of the measurement session. For a 1.5 h (optimally 2 h) measurement session, the double carrier-phase differential method [23] is used. Otherwise, the triple carrier-phase differential method is used, which yields results with centimetre accuracy [24,25]. While there are quite many publications on real-time sub-services of ASG-EUPOS [26–29], the evaluation of the functioning of POZGEO has not been a popular research topic.

The current level of development of active geodetic networks and the availability of methods to determine positions based on reference stations allow us to obtain accurate and precise coordinates of point positions. The important aspect of each of the position determination methods is its reliability. It is related to the resistance of the infrastructure to failures. In assessing the correctness of the positioning methods, accuracy measures are commonly used for both horizontal and three-dimensional (3D) coordinates. The circular error probable (CEP) is used to determine the accuracy of horizontal positions appointed by navigation and guidance systems such as GPS as well as for error analysis in geopositional assessments [30,31]. In the case of the analysis of 3D positions, the spherical error probable (SEP) is used [32–34]. The CEP refers to the radius of a circle meanwhile the SEP refers to the radius of a sphere in which 50% of the values occur. Additionally, appropriate probability levels are assigned to determine the spatial position-error bounds [35]. For the 3D positions, the following measures can also be calculated: mean radial spherical error (MRSE), 90% spherical accuracy standard (SAS90), and 99% spherical accuracy standard (SAS99). These measures specify the accuracy of the 3D positions compared with the reference positions at 61%, 90%, and 99% probability levels, respectively. The proper operation and accuracy of the point positioning methods can be determined on the basis of analyses of time series of multiple determinations of the position of the same point. For this purpose, statistical analyses can be used. When comparing sets of coordinates obtained by different methods or under different conditions, their mutual correlations can be determined. However, in the case of 3D coordinates, we are forced to determine many correlation coefficients and as a result their interpretation is difficult.

The comparison of the sets described using multi-dimensional variables can also be carried out using multivariate analysis, for more information see ref. [36]. The dissimilarity between two 3D sets can be determined using measures of distance, for example, Euclidian distance or Mahalanobis distance [37]. However, Euclidian distance is very sensitive to the scales of the variables and does not include correlated variables. In the case of Mahalanobis distance the covariance matrix of variables is taken into account. Therefore, it can be problematic to compute the inverse of the covariance matrix if the variables are highly correlated or if the number of data in sets is less than the number of variables [38]. The reliability of Mahalanobis distance results is also questioned [39,40]. However, Mahalanobis distance is included in many statistics packages and used in various scientific studies, for example, in the field of image processing [41]. The similarity of sets can also be determined based on the frequency of their elements or groups of elements. Methods based on the frequency of elements in sets have found application in comparing text documents [42]. In most of these methods, text is represented as containing unigrams or bigrams, which are collocated into a term-document matrix (TDM). In this case, the cosine distance is mainly used as a measure of distance between text representations [43].

The aim of this study is to present and validate a method for comparing sets of *XYZ* geocentric Cartesian coordinates. The presented method used 3D figures of assessment and it is a modification of the method of comparing text sets. The proposed method is based on the assessment of the frequency of the occurrence of resulting positions in the area around the positioned point after its delimitation by division into 3D figures of assessment. The proposed method was confronted with correlation coefficient and other methods of determining dissimilarity such as Mahalanobis distance. Another objective of this study is to validate the number of tie vectors used in the ASG-EUPOS APPS sub-service. For this purpose, the proposed method for comparing 3D sets was applied. The obtained results were also compared with the accuracy measures of position, which are used by GNSS users. The analyses will show whether a reduced number of tie vectors allow determination of the same accuracy of position determination in post-processing. It is of particular importance near state borders and in coastal areas where there are no EUPOS reference stations nearby. Validation of the correctness of APPS in the case of the excluded parts of stations located near the place of observation provides information about the reliability of this sub-service.

2 Methods

The study used dual-frequency carrier-phase observations recorded by the research reference station of the

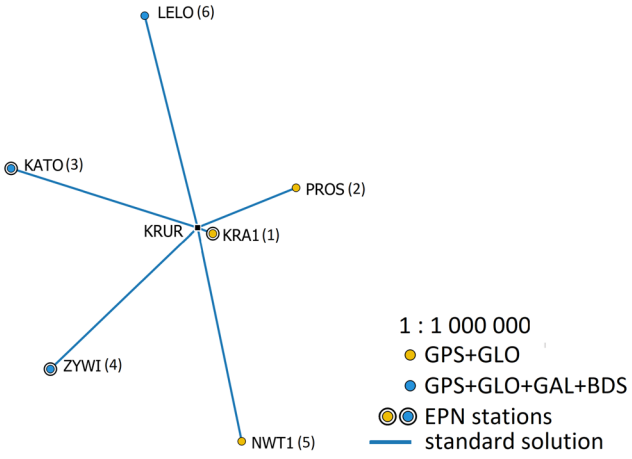

Figure 2: The geometry of the tested tie vector configuration.

University of Agriculture in Kraków, KRUR. The reference station was at that time equipped with a TRM55971.00 TZGD antenna and a Trimble NetR9 receiver. The observations were made from 1.01.2017 to 29.04.2017 (1-119 DOY 2017). ASG-EUPOS's POZGEO received 2734 RINEX 2.11 files with 1 h of static GPS observations with a 1 s sampling interval, which is 96% of the observations that can potentially be recorded in this period. The missing 4% was due to power outages and upgradation work in the facility. The authors selected 2,190 cases of position determined based on six reference stations indicated by POZGEO's algorithm, which were located optimally as regards the determined point (the standard solution) based on reports from the double carrier-phase difference APPS (Figure 2). These 2,190 positions created the set V6, which was used to validate the operation of the APPS algorithm of POZGEO under typical operation.

In 544 cases, the position was determined based on spatial vectors to a smaller number of reference stations or stations other than the six optimally located stations (alternative solutions). These cases were excluded from the analysis.

Linear deviations from the reference coordinates of KRUR were calculated for the positions obtained using the standard solution and alternatives (Figure 3).

Three variants of positioning were simulated with the number of tie vectors different than that employed in POZGEO: five, four, and three (Figure 4). The authors used data from reports from APPS to simulate the coordinates of the KRUR station position. Especially, components of tie vectors between the determined point and particular reference stations as well as reference coordinates of particular stations were used. The discussion was based on a juxtaposition of the individual simulations with the standard solution that employed six reference stations. In each variant, 6, 15, and 20 combinations, respectively, could be generated.

The procedure resulted in sets of coordinates determining the position of the point depending on the variant and combinations of the possible tie-ins for each variant. The coordinate notation accuracy was set to 0.001 m. These sets were used to validate the operation of APPS after excluding one (equation (1)), two (equation (2)), and three (equation (3)) of the reference stations typically used in positioning operations for the selected measurement point:

$$\begin{aligned} V5(c) = \{(X, Y, Z) : X_i &= (X_i(m) + X_i(n) + X_i(o) + X_i(p) + X_i(q))/5, \\ Y_i &= (Y_i(m) + Y_i(n) + Y_i(o) + Y_i(p) + Y_i(q))/5, \\ Z_i &= (Z_i(m) + Z_i(n) + Z_i(o) + Z_i(p) + Z_i(q))/5, \\ & m \neq n \neq o \neq p \neq q, \quad m, n, o, p, q \in \{1, 2, 3, \ldots, 6\}, \\ & c \in \{1, 2, 3, \ldots, 6\}, \quad i \in \{1, 2, 3, \ldots, 2190\}\} \end{aligned} \tag{1}$$

$$\begin{aligned} V4(c) = \{(X, Y, Z) : X_i &= (X_i(m) + X_i(n) + X_i(o) + X_i(p))/4, \\ Y_i &= (Y_i(m) + Y_i(n) + Y_i(o) + Y_i(p))/4, \\ Z_i &= (Z_i(m) + Z_i(n) + Z_i(o) + Z_i(p))/4, \\ & m \neq n \neq o \neq p, m, n, o, p \in \{1, 2, 3, \ldots, 6\}, \\ & c \in \{1, 2, 3, \ldots, 15\}, \quad i \in \{1, 2, 3, \ldots, 2190\}\} \end{aligned} \tag{2}$$

$$\begin{aligned} V3(c) = \{(X, Y, Z) : X_i &= (X_i(m)+X_i(n)+X_i(0))/3, \\ Y_i &= (Y_i(m) + Y_i(n) + Y_i(o))/3, \\ Z_i &= (Z_i(m) + Z_i(n) + Z_i(o))/3, \\ & m \neq n \neq o, \quad m, n, o \in \{1, 2, 3, \ldots, 6\}, \\ & c \in \{1, 2, 3, \ldots, 20\}, \quad i \in \{1, 2, 3, \ldots, 2190\}\}, \end{aligned} \tag{3}$$

where V is the number of the variant; c is the number of the combination; i is the number of positions in simulated sets; m, n, o, p, and q are the stations used in the position determination selected from the set of stations used in the standard solution; and $X_{i,m}$, $Y_{i,m}$, $Z_{i,m}$, $X_{i,n}$, $Y_{i,n}$, $Z_{i,n}$, $X_{i,o}$, $Y_{i,o}$, $Z_{i,o}$, $X_{i,p}$, $Y_{i,p}$, $Z_{i,p}$, $X_{i,q}$, $Y_{i,q}$, and $Z_{i,q}$ are the coordinates of the selected point determined on the basis of the tie vector to the single station.

The sets of the coordinates of standard solution and simulated combinations were transformed into sets of spatial pairs of linear deviations (equations (4)–(7)) in such a way that for each coordinate the reference coordinates of the KRUR station (X_R, Y_R, and Z_R) were subtracted:

$$\begin{aligned} \Delta V6 = \{(\Delta X, \Delta Y, \Delta Z) : \Delta X_i &= X_i - X_R, \Delta Y_i = Y_i - Y_R, \\ \Delta Z_i &= Z_i - Z_R, (X_i, Y_i, Z_i) \in V6, \\ & i \in \{1, 2, 3, \ldots, 2190\}\} \end{aligned} \tag{4}$$

$$\begin{aligned} \Delta V5(c) = \{(\Delta X, \Delta Y, \Delta Z) : \Delta X_i &= X_i - X_R, \Delta Y_i = Y_i - Y_R, \\ \Delta Z_i &= Z_i - Z_R, (X_i, Y_i, Z_i) \in V5(c), \\ & c \in \{1, 2, \ldots, 6\}, i \in \{1, 2, 3, \ldots, 2190\}\} \end{aligned} \tag{5}$$

$$\begin{aligned} \Delta V4(c) = \{(\Delta X, \Delta Y, \Delta Z) : \Delta X_i &= X_i - X_R, \Delta Y_i = Y_i - Y_R, \\ \Delta Z_i &= Z_i - Z_R, (X_i, Y_i, Z_i) \in V4(c), \\ & c \in \{1, 2, \ldots, 15\}, i \in \{1, 2, 3, \ldots, 2190\}\} \end{aligned} \tag{6}$$

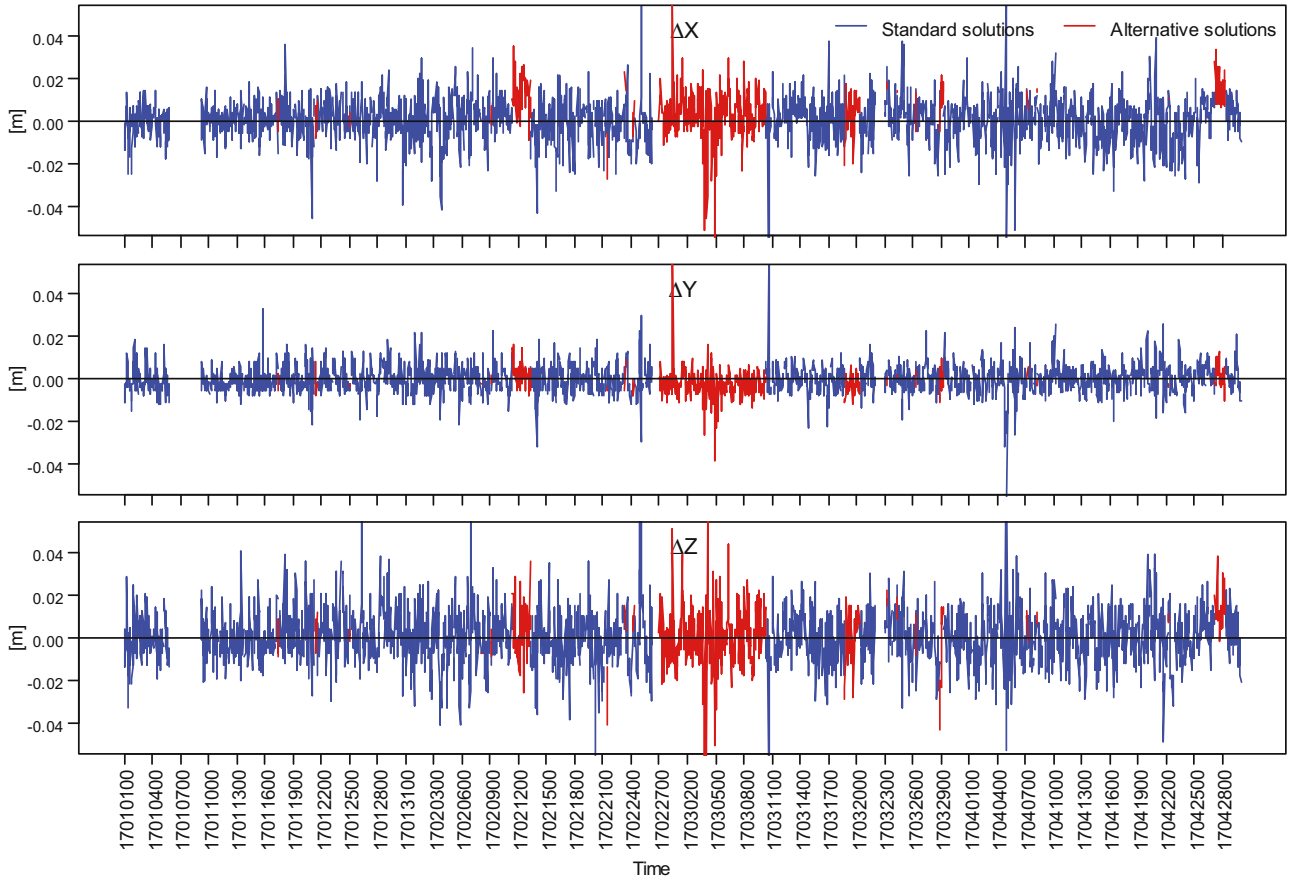

Figure 3: Linear deviations of coordinates from reference coordinates.

$$\begin{aligned} \Delta V3(c) = \{(\Delta X, \Delta Y, \Delta Z) : \Delta X_i = X_i - X_R, \Delta Y_i = Y_i - Y_R, \\ \Delta Z_i = Z_i - Z_R, (X_i, Y_i, Z_i) \in V3(c), \\ c \in \{1, 2, \ldots, 20\}, i \in \{1, 2, 3, \ldots, 2190\}\} \end{aligned} \quad (7)$$

where V is the number of the variant, c is the number of the combination, and i is the number of positions in simulated sets.

Based on the standard deviations of the simulated coordinates compared to the reference coordinates, accuracy measures were calculated: MRSE, SAS90, and SAS99. The accuracy measures facilitated the assessment of the accuracy of each variant of the tie vector scheme.

The comparison of the sets of positions resulting from changed numbers of tie vectors and their mutual arrangement with the set of standard solutions was based on the dissimilarity measures for the sets. A method for comparing text in terms of the frequency of occurrence of selected words, term frequency matrix (TDM) [42], was used after adjustments to handle linear deviation analysis. The resulting sets of linear deviations were expressed as a frequency of occurrence of the positions in 3D figures of assessment regularly distributed around the reference position. This way, the comparison of results was reduced to the comparison of the frequency of positions with the original results determined by POZGEO after spatial ordering. The cube was selected as the 3D figure because of the ease of construction and because the cube demonstrates well the simultaneous change of coordinates by the value of the length of its edge when aligned with the axes of the Cartesian coordinate system. The analysed space used to compare the sets had to be identical. Its extent was set based on the maximum value among the calculated absolute values of linear deviations $|\Delta X|$, $|\Delta Y|$, and $|\Delta Z|$ rounded up to the nearest value expressed with 0.01 m accuracy (equation (8)). This generated a bounding cube, the geometric centre of which coincided with the origin of the coordinate frame and the reference position (equation (9)), while 2R edges delimited the clustering area regarding the 3D figures of assessment (Figure 5).

This bounding cube H_L with vertices at $(\pm R, \pm R, \pm R)$ contains elements of all the compared sets of linear deviations:

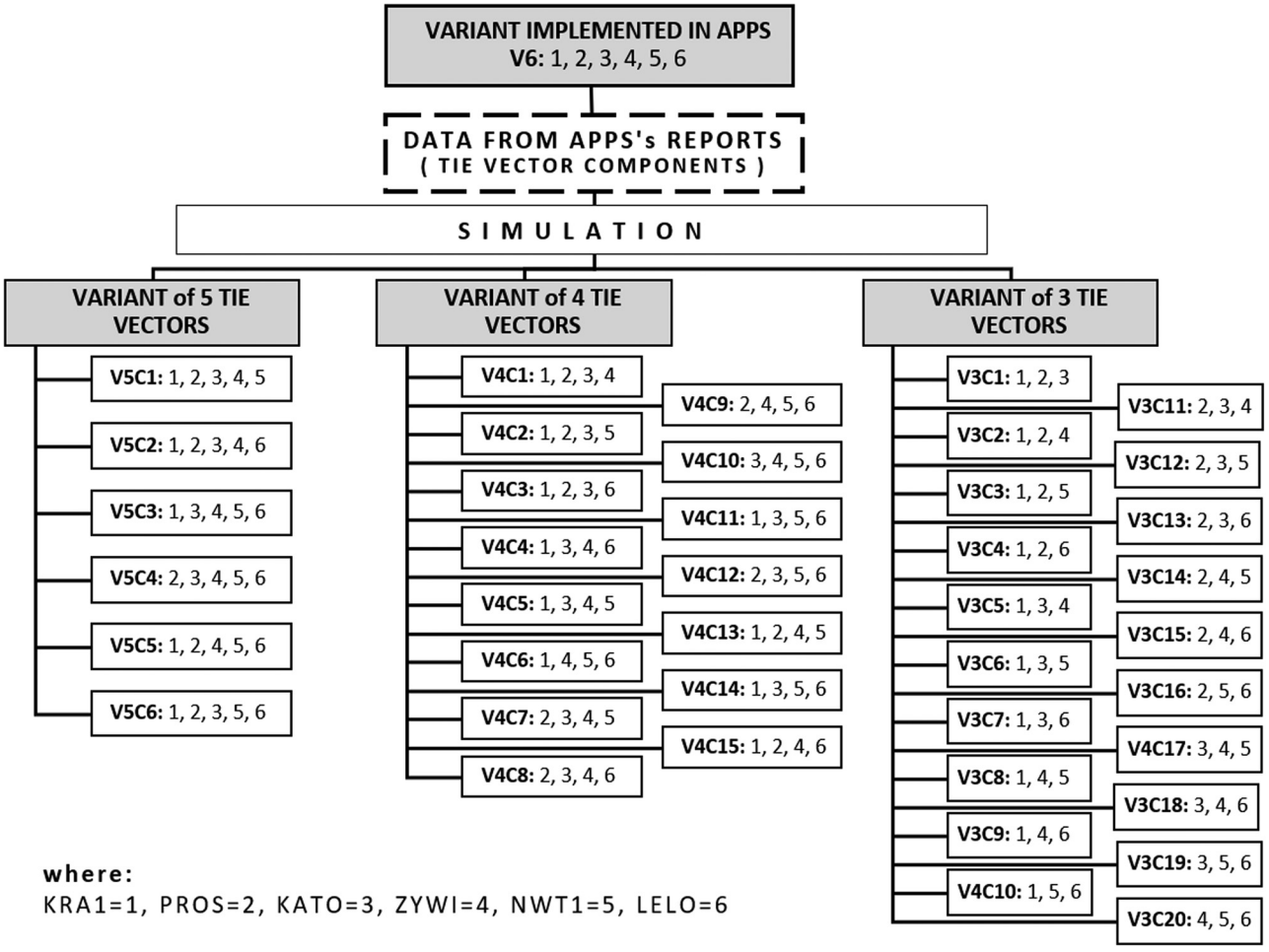

Figure 4: Diagram of the analysed tie-in vectors.

$$R = \max\{\max|\Delta X_i|,\ \max|\Delta Y_i|,\ \max|\Delta Z_i|, \Delta X_i, \Delta Y_i, \Delta Z_i \in \Delta V6 \cup \Delta V5(c) \cup \Delta V4(c) \cup \Delta V3(c)\} \quad (8)$$

$$H_L = \{(X, Y, Z) :\ \max\{|X_i - X_0|, |Y_i - Y_0|, |Z_i - Z_0|\} = R, (X_0, Y_0, Z_0) = (0, 0, 0)\}. \quad (9)$$

The space inside the bounding cube H_L was filled with 3D figures of assessment, i.e. unit cubes H_U. In the described bounding cube H_L, the following number of unit cubes H_u with edge r can be identified and then numbered:

$$j = (8R^3)/r^3. \quad (10)$$

In order to compare the sets on various accuracy levels, the following dimensions of the unit cube were used: H_{U10}: $0.010 \times 0.010 \times 0.010$ m, H_{U05}: $0.005 \times 0.005 \times 0.005$ m, and H_{U02}: $0.002 \times 0.002 \times 0.002$ m. The positions resulting from the analysed combinations were clustered using 1,25,000,

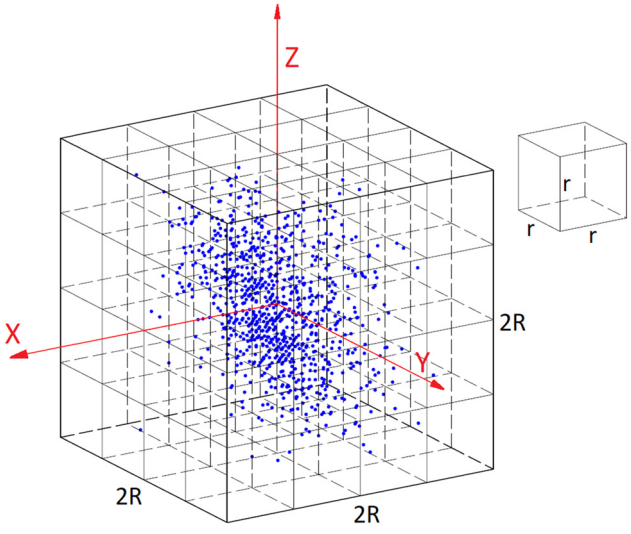

Figure 5: Example of a bounding cube and unit cubes.

10,00,000, and 1,56,25,000 unit cubes, respectively. The number of positions of each set of linear deviations was determined for each unit cube. The data were used to construct the frequency of positions in a centric cube matrix (FCC):

$$\begin{aligned} &\mathrm{FCC}(v, c) \\ &\quad = [f_1(\Delta X, \Delta Y, \Delta Z), f_2(\Delta X, \Delta Y, \Delta Z), \ldots, f_j(\Delta X, \Delta Y, \Delta Z)], \end{aligned} \tag{11}$$

where (ΔX, ΔY, ΔZ) are the spatial pairs of linear deviations from a set of selected variant's combinations and f is the frequency of spatial pairs of linear deviations in selected unit cubes numbered from 1 to j.

Based on the FCC matrices, the measure of similarity between the sets of linear deviations of coordinates obtained in the simulated combinations and the standard solution was calculated. The authors used a similarity measure based on cosine amplitude (equation (12)), which yields accurate similarity measures [43]. The FCC matrix does not include negative values. Thus, it is possible to compute cosine distance without the need for data normalization:

$$r_{\mathrm{ca}}(\mathrm{FCC}(6), \mathrm{FCC}(v, c)) = \left| \sum_{i=1}^{j} f_{\mathrm{FCC}(6),i} f_{\mathrm{FCC}(v,c),i} \right| \Bigg/ \sqrt{\left(\sum_{i=1}^{j} f_{\mathrm{FCC}(6),i}^2 \sum_{i=1}^{j} f_{\mathrm{FCC}(v,c),i}^2 \right)}. \tag{12}$$

The cosine similarity measure was transformed into the cosine distance between FCC matrices (equation (13)), which is a dissimilarity measure:

$$d_{\mathrm{fcc}}(\mathrm{FCC}(6),\mathrm{FCC}(v, c)) = 1 - r_{\mathrm{ca}}(\mathrm{FCC}(6),\mathrm{FCC}(v, c)). \tag{13}$$

The results of the presented FCC method were verified by calculating another dissimilarity measure in the form of the Mahalanobis distance using the sets of linear deviations directly and the inverse covariance matrix (C^{-1}) (equation (14)).

$$\begin{aligned} &d_{\mathrm{M}}(\Delta V6, \Delta Vn(c)) \\ &\quad = \sqrt{((\Delta V6 - \Delta Vn(c))^T C^{-1} (\Delta V6 - \Delta Vn(c)))}, \end{aligned} \tag{14}$$

where $\Delta V_n(c)$ is the set of spatial pairs of linear deviations of selected variant's combination using n tie vectors.

The statistical data analysis was carried out using the R environment [44].

3 Results

Figure 6 shows the mean linear deviations and their standard deviations for each possible combination in particular variants (Figure 6). The mean value of linear deviation for all three coordinates is 0.000 m. A comparison of the mean values and standard deviations for individual variants and combinations with values from the standard solution demonstrated that similar values were obtained for V5C3–V5C6, V4C1, V4C2, V4C6, V4C9–V4C12, V3C2, V3C10, V3C11, and V3C19.

The authors determined the values of the radius of a sphere, the geometric centre of which was located at the reference position, which contained positions determined for each combination in the 3D space at 61%, 90%, and 99% probability for the standard solution and each proposed combination (Figure 7). Combinations V5C5, V5C6, V4C2, V3C2, and V3C10 achieve similar MRSE, SAS90, and SAS99 compared to the standard solution for which the MRSE is 0.018 m, SAS90 is 0.025 m, and SAS99 is 0.034 m. The maximum values of MRSE, SAS90, and SAS99 of 0.036, 0.049, and 0.066 m, respectively, were observed for V3C17. For solutions that used five and four tie vectors, the MRSE took values from 0.02 to 0.03 m, while for solutions with three vectors, from 0.03 to 0.04 m. MRSE did not exceed 0.040 m for any combination and SAS90 for 80% of the combinations. Therefore, it can be stated with a probability of 61% that the positions determined for these combinations will not be further from the true value than 0.04 m, which is also true for 80% of cases at a probability of 90%. For solutions that used five and four tie vectors, the mean SAS90 was 0.03 m, while for solutions with three vectors, 0.04 m. For solutions that used five and four tie vectors, the mean SAS99 was 0.04 m, while for solutions with three vectors, 0.05 m. SAS90 did not exceed 0.05 m for 95% of the simulated combinations. Therefore, it can be stated with a probability of 90% that the positions determined for these combinations are not further from the true value than 0.05 m regardless of the number of tie vectors. This statement is true also at 99% probability for 66% of the simulated combinations.

The authors conducted the Shapiro–Wilk test of normality for the set of linear deviations of each possible combination of tie vectors in the proposed simulation variants. The test demonstrated that the hypothesis of the sample originating from a normal distribution population should be rejected at a significance level of 0.05 for all cases. Therefore, Spearman's rank correlation coefficients were determined between linear deviations of coordinates in the standard solution and individual simulated combinations, separately for each coordinate. Significance tests of the estimation of correlation coefficients at a significance level of 0.05 have demonstrated that the Spearman's rank correlation coefficients were statistically significant and positive for all the analysed cases (Figure 8). The lowest

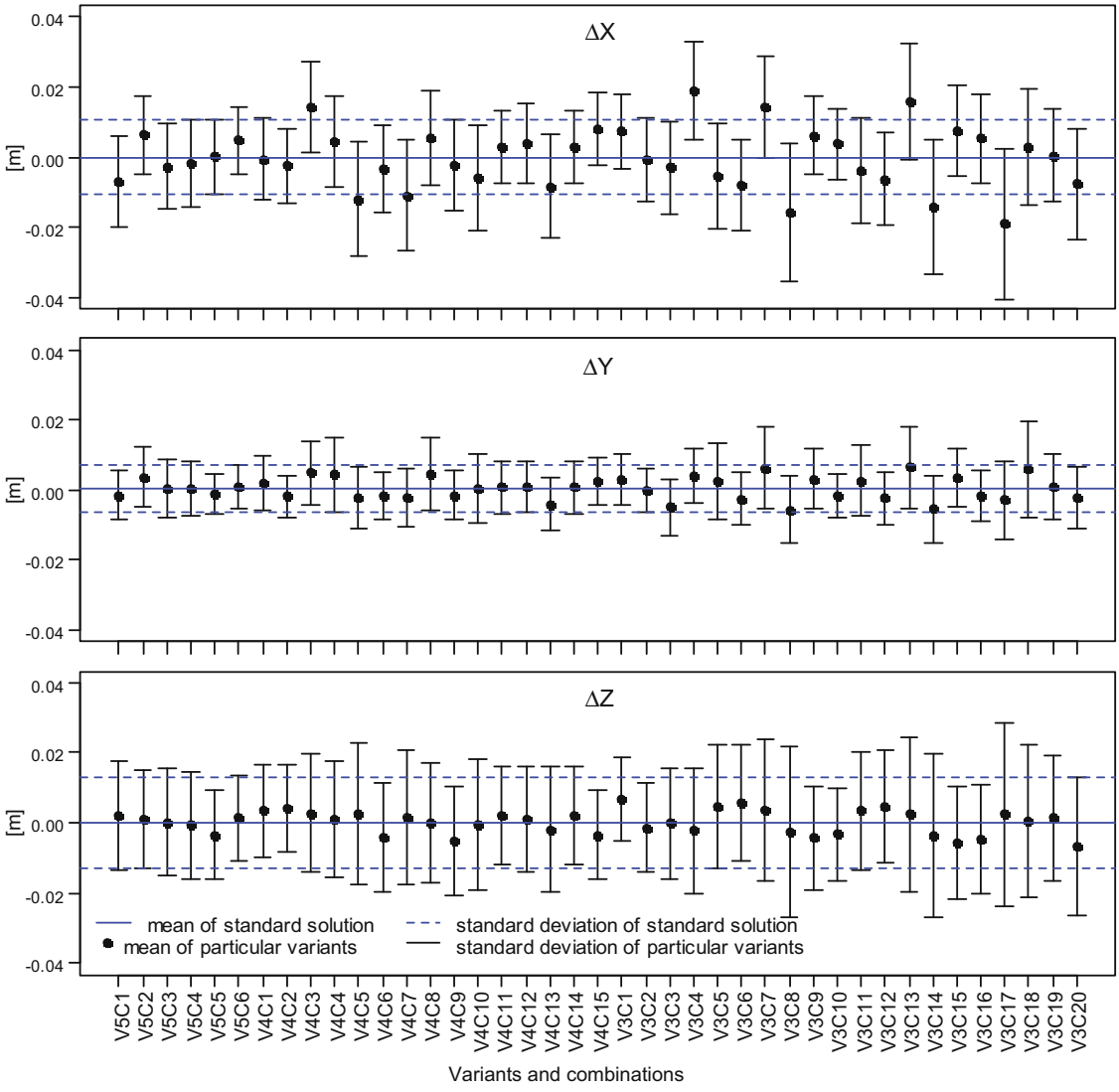

Figure 6: Mean values of deviations Δ*X*, Δ*Y*, and Δ*Z* with values of standard deviations for individual combinations in the proposed variants of solutions.

correlation coefficients were found in combinations V4C3, V3C4, V3C7, and V3C13 for linear deviations of coordinates *X* and *Z*, and in combinations V4C13, V3C3, and V3C14 for deviations of the *Y* coefficient. High values of the correlation coefficient for linear deviations of all coordinates were found for V5C3, V5C4, V5C5, and V4C10. According to a classification by J. Guilford, very high correlations were found for all combinations of the variant with five tie vectors, 67% of combinations of the solutions with four tie vectors, and 30% of the combinations with three tie vectors.

The calculated Spearman's rho correlation coefficients were determined individually for each component of the position. This resulted in a challenge involving the indication of the level of correlation of specific combinations with the standard solution in the 3D space. The allocation of ranks to linear deviations is not identical within each coordinate and combination. This prevents any direct comparison of the analysed combinations with each other. In order to obtain a synergistic coefficient of similarity describing the spatial relations between the sets, the authors decided to determine the frequency of positions in selected 3D figures of assessment. Sets of pairs of linear deviations were represented by matrices. This approach facilitated the retention of georeferencing of the reference position because the location of unit cubes in relation to the reference position was strictly defined. The dissimilarity measure in the form of cosine distance was calculated between the FCC matrix of the standard solution and the FCC matrices of proposed combinations (Figure 9). Thus, the

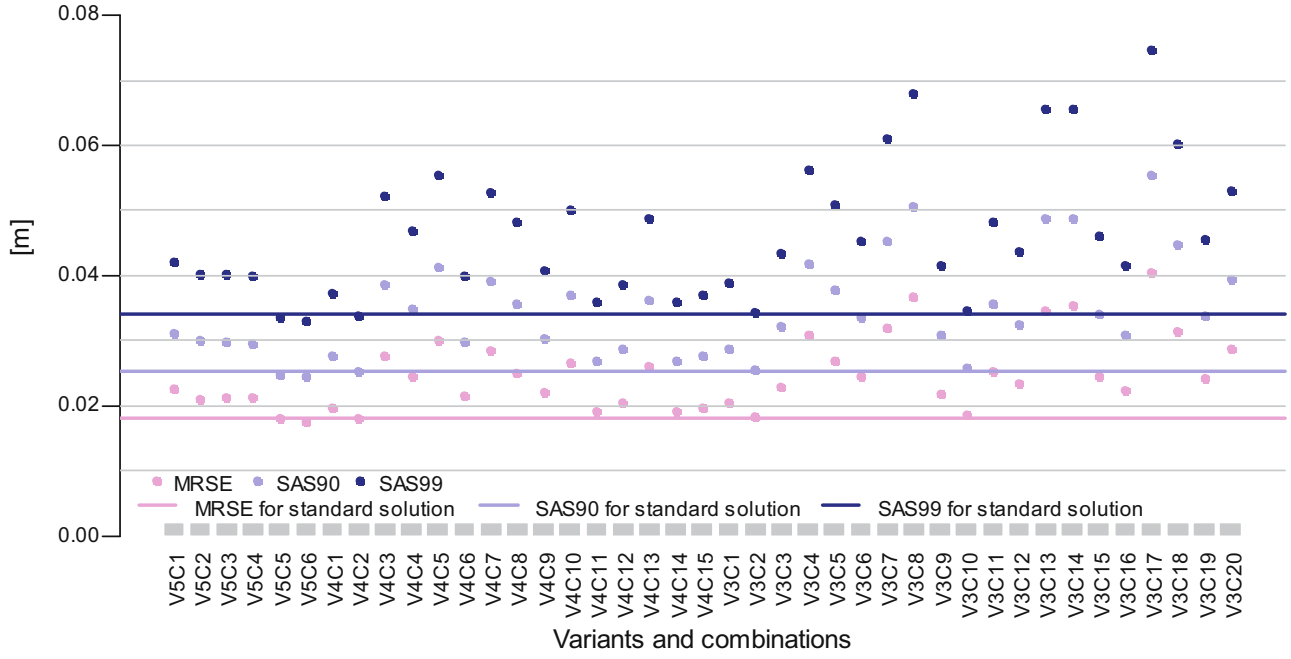

Figure 7: MRSE, SAS90, and SAS99 for the standard solution and the simulated combinations.

FCC method allows classifying particular combinations with respect to standard solutions.

The distance in the FCC method has demonstrated similarity of solutions using five tie vectors and the set of the standard solutions especially as regards the H_{U10} accuracy of grouping. The values of distance in the FCC method in these combinations do not exceed 0.20. The distance in the FCC method was below 0.20 in the

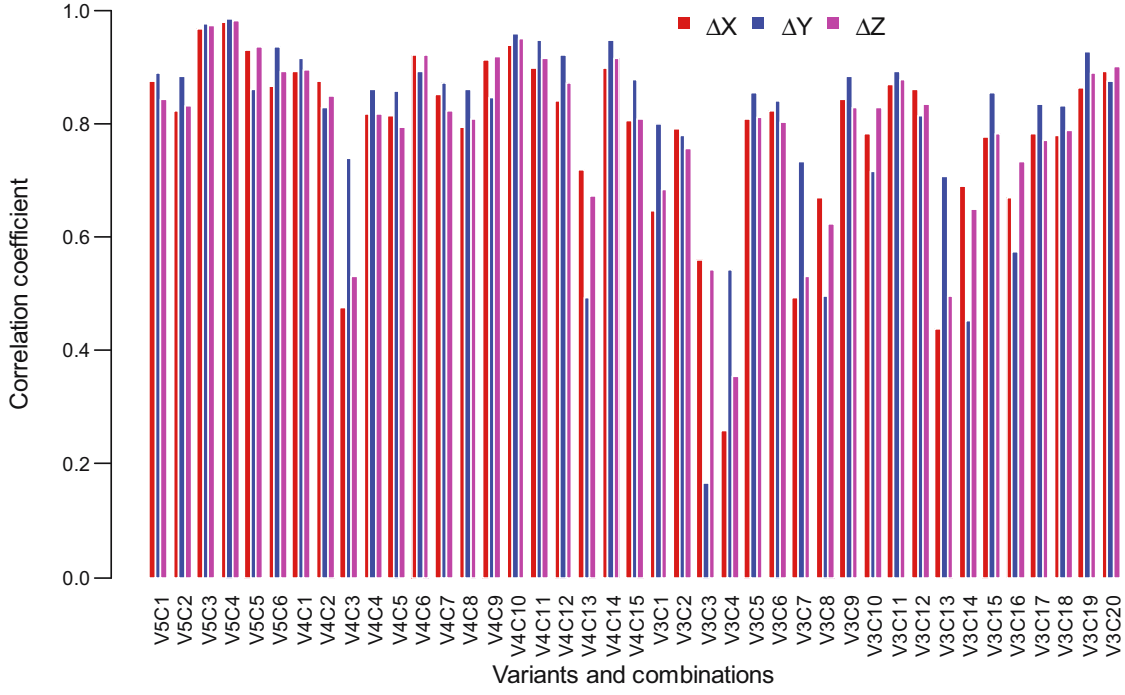

Figure 8: Spearman's rho correlation coefficients between the standard solutions and individual simulated combinations based on linear deviations of coordinates expressed at 0.010 m accuracy.

4-vector variant in 66% of the possible combinations, and in the 3-vector variant, in 35% of the combinations. Therefore, the sets should be considered similar because of the distance in the FCC method in the position frequency matrix of 0.01 m unit cubes. The reduction of the edge of the unit cube increased the coefficient of dissimilarity between the analysed combinations and the set of standard solutions. For unit cubes H_{U05}, the distance in the FCC method between combinations in the 5-vector variant and the reference set was 0.20 for 83% of the combinations. For the 4-vector variant, it was 46% and for the 3-vector variant, 10%. In the case of unit cubes H_{U02}, no possible combination was located at a distance of 0.20 from the reference set. The minimum value of distance in the FCC method was 0.28 for combination V4C11. The smallest values for unit cubes H_{U10}, H_{U05}, and H_{U02} were found for V5C3, V5C4, V5C6, V4C11, V4C12, V4C14, V3C2, and V3C19. Generally, the mean distance in the FCC method increases in all cases when changing the comparison level from H_{U10} to H_{U05}. The mean increase of the distance in the FCC method is 86% for all combinations of five vectors and 69% and 58% for combinations of four and three tie vectors, respectively. Due to the change in the comparison level of the sets from H_{U10} to H_{U05}, the smallest change of distance by 9% was noted for the combination of V3C4 and the largest by 269% for V3C2. For the change of the comparison level from H_{U10} to H_{U02}, there was an average increase in the measure of dissimilarity by 128% for the combination of five tie vectors, 131% and 139% for four and three tie vectors, respectively. The smallest value of change by 69% was noted for the combination of V3C4 and the largest by 222% for V3C18. In the case of changing the comparison level from H_{U05} to H_{U02} a mean increase of dissimilarity amounting to 181% for the combination of five tie vectors and 152% and 76% for four and three tie vectors is noticed, respectively. The highest increase of dissimilarity was found for V4C14 which reached 404% and the smallest for V3C4 which in turn takes a value of 10%. Different values of distance in the FCC method were achieved depending on the level of comparison, indicating the spatial diversity of the distributions of the compared sets and the set of standard solutions.

For verification of the obtained dissimilarities of sets, the distance in the FCC method is compared with Mahalanobis distance. Although the Mahalanobis distance calculated between sets of linear deviations of the combinations and linear deviations of the standard solution assumed values close to 0, the values varied significantly (Figure 10). The results of Mahalanobis distances are compatible with the values of distance in the FCC method at the different levels of comparison. The smallest values of the Mahalanobis distance were

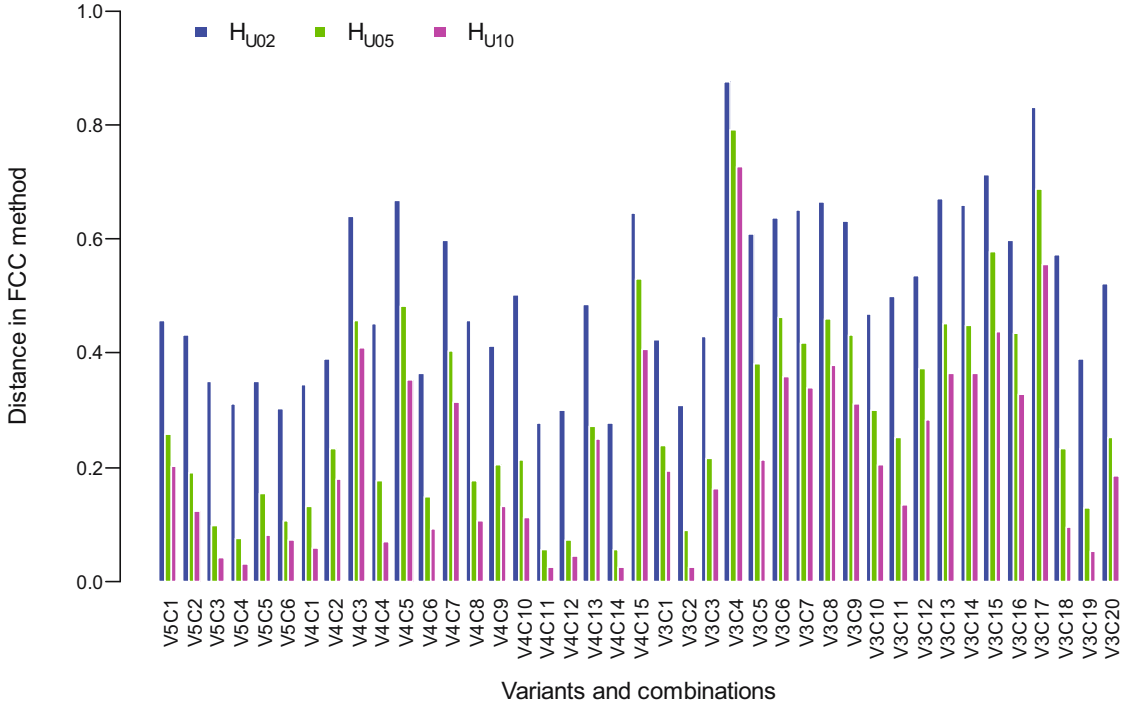

Figure 9: The distance in the FCC method between the standard solution and the simulated combinations.

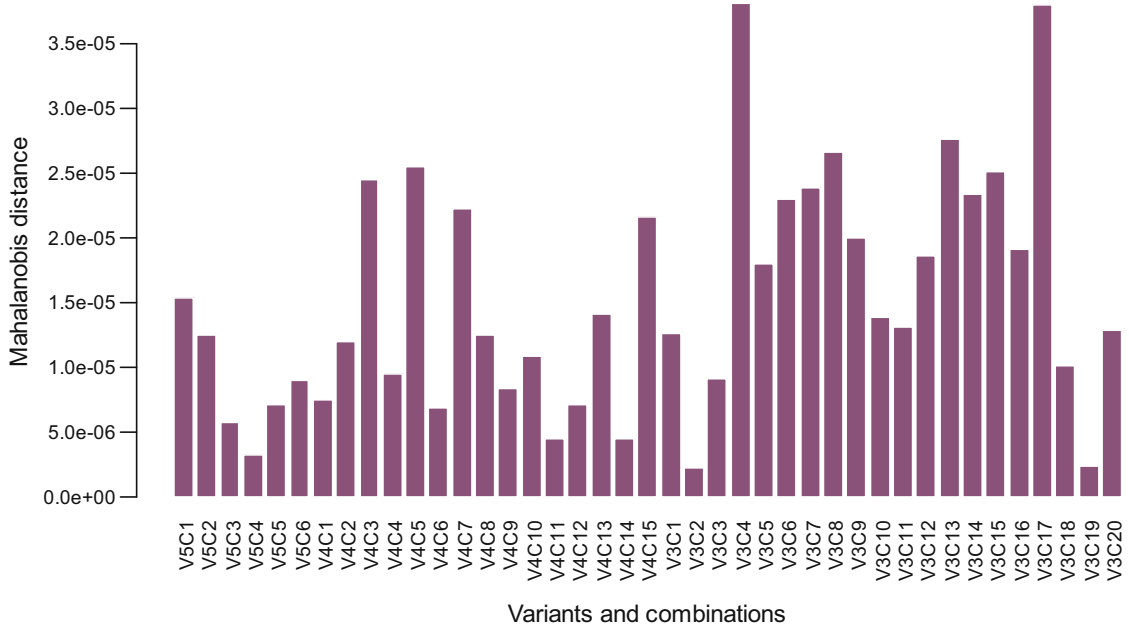

Figure 10: The Mahalanobis distance between the set of linear deviations of the standard solution and the simulated combinations.

found for the same combinations, for which the dissimilarity with the reference set exhibited a small value of the distance in the FCC method (Figure 10). Extreme values indicative of no similarity were found for combinations V3C4 and V3C17.

The measures of dissimilarity were compared to the value of Spearman's rho correlation coefficient for linear deviations and Mahalanobis distance (Figures 8 and 10). If Spearman's rho correlation coefficient reaches values close to 1, the distances in the FCC method at the levels of comparison H_{U02}, H_{U05}, and H_{U10} as well as Mahalanobis distance have the smallest values. It is especially visible in the case of V5C3, V5C4, V4C11, V4C14, V3C19, and V3C20. In the analysed measures there is also compliance as to the lack of similarity, which can be seen in the case of V3C4.

In order to properly compare the obtained values of dissimilarity measures, the unitization with zero minimum method was applied. In this way, they adopted values from 0 to 1, where 0 means total similarity and 1 means no similarity (Figure 11).

In most cases of variants and combinations, the Mahalanobis distance and distance in the FCC method at levels H_{U02}, H_{U05}, and H_{U10} show comparable values. However, it can be seen that in many cases the Mahalanobis distance achieves slightly higher values. The Mahalanobis distance is the closest to the distance in the FCC method for comparison at the level of H_{U10} in the case of 43% of analysed combinations, 32% combinations for the level of H_{U05}, and 25% combinations for H_{U02}.

4 Discussion

The proposed FCC method of the frequency of positions in a centric cube based on the clustering of positions in unit cubes with an edge length of 0.010, 0.005, and 0.002 m facilitates the construction of a position frequency matrix. The cosine distance calculated for the FCC matrix is then used to compare spatial dependencies among sets of positions appropriately and to identify similar sets. It can be noticed that the variants and combinations for which the MRSE, SAS90, and SAS99 measures have reached values very similar to those for standard solution are also characterized by low coefficients of dissimilarity obtained after application of the FCC method for comparing 3D sets of *XYZ* Cartesian coordinates. It can be particularly noticed in cases of combinations V5C5, V5C6, V4C2, V3C2, and 3C10. Verification of the FCC method with respect to the commonly known Mahalanobis distance measure or Spearman correlation coefficients also showed its correctness. The measure of dissimilarity obtained on the basis of the FCC method for comparing 3D sets proved to reach similar values. The undoubted advantage of the FCC method is the possibility of using it for comparing sets of 3D coordinates as well as other multidimensional sets. The FCC method is also easy to use and to implement. However, special attention should be paid to the selection of the appropriate edge length of the unit cube. The selection of this magnitude will depend on the type of data being analysed as well as the level at which information about data similarity is expected. The analyses carried out in this

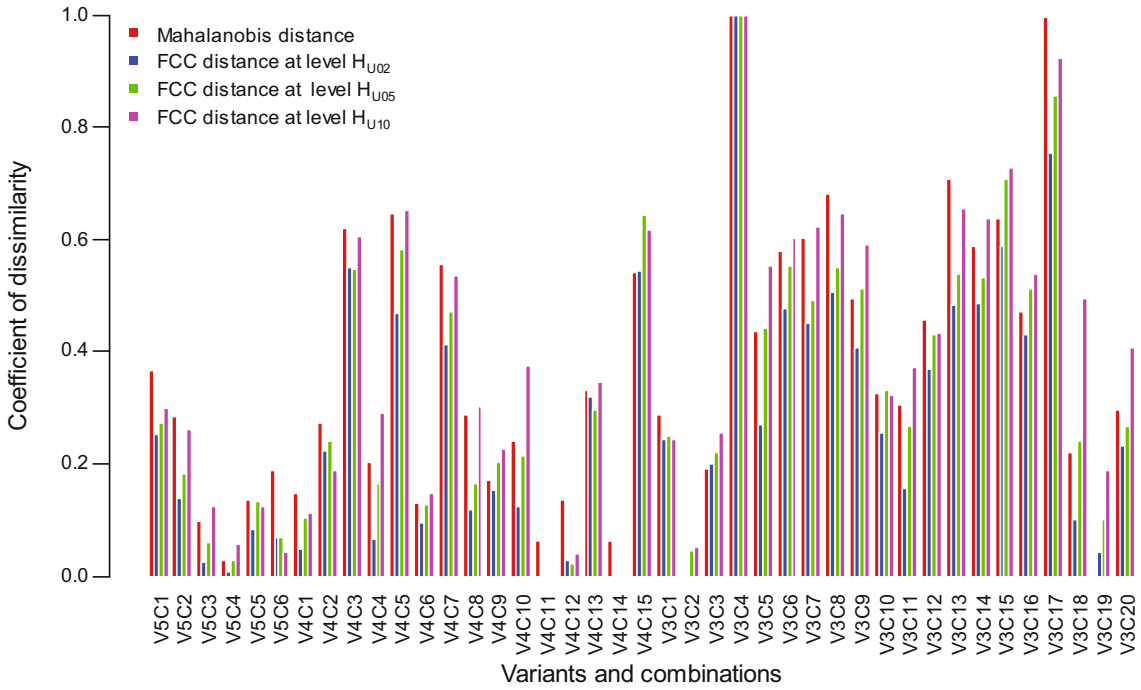

Figure 11: The Mahalanobis and the distance in the FCC method after its unitization.

study show that with the increase in the length of the cube edge, the coefficient of similarity of sets increases. Thus, the assumption of an inadequate size of the length of the cube edge may result in unreliable results. In the case of analysing less accurate positioning methods larger edge length of the unit cube should be used, which will be adequate to the accuracy of the input data sets. Furthermore, the proposed method allows assessing more accurately the similarity of solutions than inference based on a comparison of commonly used measures of accuracy.

The FCC method for comparing 3D sets allowed verifying whether it is possible to determine the position using APPS based on a smaller number of tie vectors with accuracy close to the value held in relation to the six vectors. The positioning results obtained with the post-processing algorithm of ASG-EUPOS are consistent with the reference data below 0.020 m at 61% probability, 0.030 m at 90% probability, and 0.040 m at 99% probability when six tie vectors to nearby reference stations and 1 h GPS observations with a 1 s sampling interval are used. A reduction of the number of tie vectors to five stations does not result in a significant deterioration of the accuracy of coordinates in most cases. It can be stated with a probability of 90% that the positions determined in the ETRS89 Cartesian coordinates for 96% simulated combinations based on a typical solution are not further from the true value than 0.05 m regardless of the number of tie vectors. ASG-EUPOS's POZGEO is useful for determining points in geodetic control networks, which is consistent with its intended purpose [45,46].

It has been demonstrated that such configurations of three tie vectors can be selected from the currently used six tie vectors in the APPS algorithm of ASG-EUPOS and the results feature similar positioning accuracy and consistency of solutions to the results currently provided by the system. This proves the hypothesis of other researchers who question using six tie vectors in the POZGEO algorithm [47].

It should be considered whether the APPS algorithm should base the solution on five or four tie vectors upon failure of a nearby reference station without employing an additional reference station so that the result can be confronted with previously determined points through the elimination of the vector to the failed reference station. The proposed FCC method can be useful for evaluating the work of a reference station in comparison to other stations belonging to a selected active geodetic network. Further research will focus on determining the average frequency distribution pattern of positions for the FCC method for selected reference station systems.

Acknowledgments: This research was financed by the Ministry of Science and Higher Education of the Republic of Poland, grant numbers BM-2308./KG/2018 and DS-3356/KG/18.

Conflict of interest: The authors of this manuscript declare no conflict of interest.

References

[1] EUPOS Technical Standards, rev. 3, 2013. http://www.eupos.org/techsd.

[2] EUPOS Guidelines For Cross-Border Data Exchange, 2006. http://www.eupos.org/node/8.

[3] EUPOS Guideline for EUPOS Reference Frame Fixing, 2007. http://www.eupos.org/node/8.

[4] EUPOS Technical Standards, rev. 2, 2008. http://www.eupos.org/techsd.

[5] Graszka W, Rosenthal G, Śledziński J. Ten years of establishment of the satellite reference station system EUPOS in Central and Eastern Europe. Rep Geodesy Geoinform. 2012;93:11–24.

[6] Parseliunas E, Buga A, Marozas L, Petniunas M, Urbanas S. LitPOS – a part of EUPOS. Geodesy Cartograph. 2008;34:2. doi: 10.3846/1392-1541.2008.34.50-57.

[7] Dobelis D, Zvirgzds J. Network RTK performance analysis: a case study in Latvia. Geodesy Cartograph. 2016;42(3). doi: 10.3846/20296991.2016.1226383.

[8] Droščák B, Smolík K. Analysis of the SKPOS users initialisation times. Slovak J Civil Eng. 2014;22(3):13–20. doi: 10.2478/sjce-2014-0013.

[9] Raška M, Pospíšil J. Minimal detectable displacement achievable by GPS-RTK in CZEPOS Network. Geoinform FCE CTU. 2015;14(1):29–38. doi: 10.14311/gi.14.1.2.

[10] Štroner M, Urban R, Královič J. Testing of the relative precision in local network with use of the Trimble GEO XR GNSS receivers. Rep Geodesy. 2013;94:27–36. doi: 10.2478/rgg-2013-0004.

[11] Metsar J, Kollo K, Ellmann A. Modernization of the Estonian national GNSS reference station network. Geodesy Cartograph. 2018;44(2):55–62. doi: 10.3846/gac.2018.2023.

[12] Bosy J, Graszka W, Leończyk M. ASG-EUPOS – a multifunctional precise satellite positioning system in Poland. Eur J Navigat. 2007;5(4):371–74.

[13] Calka B, Bielecka E, Figurski M. Spatial pattern of ASG-EUPOS sites. Open Geosci. 2017;9(1):613–21. doi: 10.1515/geo-2017-0046.

[14] Bogusz J, Figurski M, Kontny B, Grzempowski P. Horizontal velocity field derived from EPN and ASG-EUPOS satellite data on the example of south-western part of Poland. Acta Geodynam Geomater. 2012;3(167):349–57.

[15] Bogusz J, Figurski M, Kontny B, Grzempowski P. Unmodelled effects in the horizontal velocity field determination: ASG-EUPOS case study. Artificial Satellites. 2012;47(2):67–79. doi: 10.2478/v10018-012-0014-x.

[16] Kuczynska-Siehien J, Lyszkowicz A, Sideris MG. Evaluation of altimetry data in the Baltic Sea Region for computation of new Quasigeoid models over Poland. In: Freymueller J, Sánchez L. editors. International Symposium on Advancing Geodesy in a Changing World, Proceedings of the IAG Scientific Assembly. Springer; 2019. p. 149. doi: 10.1007/1345_2018_35.

[17] Kowalczyk K. The creation of a model of relative vertical crustal movements in the Polish territory on the basis of the data from active geodetic network EUPOS (ASG EUPOS). Acta Geodyn Geomater. 2015;12(3179):215–25. doi: 10.13168/AGG.2015.0022.

[18] Szafranek K, Bogusz J, Figurski M. GNSS reference solution for permanent station stability monitoring and geodynamical investigations: the ASG-EUPOS case study. Acta Geodyn Geomater. 2013;10(1):67–75.doi: 10.13168/agg.2013.0006.

[19] Krypiak-Gregorczyk A, Wielgosz P, Krukowska M, A new ionosphere monitoring service over the ASG-EUPOS network stations. In Environmental Engineering. Proceedings of the International Conference on Environmental Engineering. ICEE (Vol. 9, p. 1). Vilnius Gediminas Technical University, Department of Construction Economics & Property. doi: 10.3846/enviro.2014.224.

[20] ASG-EUPOS services http://www.asgeupos.pl/index.php?wpg_type=serv&sub=gen.

[21] Kadaj R, Świętoń T. Algorytm i oprogramowanie modułu automatycznego postprocessingu w polskim systemie satelitarnych stacji referencyjnych ASG-EUPOS [Algorithm and software of automatic postprocessing module (APPS) in Polish precise satellite positioning system ASG-EUPOS], ZN Politechniki Rzeszowskiej. Budownictwo i Inżynieria Środowiska. 2009;262(51):37–57, (in Polish with English summary).

[22] Kadaj R, Świętoń T. Theoretical and applied research in the field of higher geodesy conducted in Rzeszow. Rep Geodesy Geoinform. 2016;100(1):79–100. doi: 10.1515/rgg-2016-0008.

[23] Kadaj R. Pół-analityczne rozwiązanie bazowe typu “float” postprocessingu GPS. [Half analytical base “float” solution of GPS postprocessing]. Zeszyty Naukowe Budownictwo i Inżynieria Środowiska. 2012;283(59):169–80, (in Polish with English summary).

[24] Kadaj R. New algorithms of GPS post-processing for multiple baseline models and analogies to classic geodetic networks. Geodesy Cartograph. 2008;57(2):61–79.

[25] Kadaj R. Zastosowanie różnicowego układu obserwacyjnego typu Schreibera do opracowania sesji pomiarów statycznych GPS [Application of Schreiber’s type diff erence observation system for elaboration of a session of static GPS measurements]. Biuletyn WAT. 2010;LIX 2(658):85–106, (in Polish with English summary).

[26] Bakuła M, Pelc-Mieczkowska R, Walawski M. Reliable and redundant RTK positioning for applications in hard observational conditions. Artificial Satellites. 2012;47(1):23–33. doi: 10.2478/v10018-012-011-0.

[27] Krzyżek R, Skorupa B. Analysis of accuracy of determination of eccentric point coordinates of the KRAW permanent geodetic station in RTK GPS measuring mode with the application of the NAWGEO service of the ASG-EUPOS system. Geomat Environ Eng. 2012;6(4):35–46. doi: 10.7494/geom.2012.6.4.35.

[28] Rogowski J, Specht C, Weintrit A, Leszczyński W. Evaluation of positioning functionality in ASG EUPOS for hydrography and off-shore navigation. TransNav Int J Mar Navigat Safety Sea Transport. 2015;9(2):221–27. doi: 10.12716/1001.09.02.09.

[29] Bakuła M. Constrained and network multi-receiver single-epoch RTK positioning. Survey Rev. 2018:1–10. doi: 10.1080/00396265.2018.1547532.
[30] Wang Y, Jia XR, Yang G, Wang YM. Comprehensive CEP evaluation method for calculating positioning precision of navigation systems. Appl Mech Mater. 2013;341:955–60. doi: 10.4028/www.scientific.net/AMM.341-342.955.
[31] Wang Y, Yang G, Yan D, Wang YM, Song X. Comprehensive assessment algorithm for calculating CEP of positioning accuracy. Measurement. 2014;47:255–63. doi: 10.1016/j.measurement.2013.08.062.
[32] Singh N. Spherical probable error. Nature. 1962;193(4815):605. doi: 10.1038/193605a0.
[33] Singh N. Spherical probable error (SPE) and its estimation. Metrika. 1970;15(1):149–63. doi: 10.1007/BF02613568.
[34] Jiang R, Wang YY, Yang GL, Wang YM. Time-SEP comprehensive assessment of methods for calculating positioning accuracy of navigation systems. Adv Mater Res. 2013;765:2026–30. doi: 10.4028/www.scientific.net/AMR.765-767.2026.
[35] Ignagni M. Determination of circular and spherical position-error bounds in system performance analysis. J Guid Control Dynam. 2010;33(4):1301–05. doi: 10.2514/1.47573.
[36] Olkin I, Sampson AR. Multivariate analysis: overview. Int Encyclopedia Soc Behav Sci. 2001. doi: 10.1016/B0-08-043076-7/00472-1.
[37] De Maesschalck R, Jouan-Rimbaud D, Massart DL. The Mahalanobis distance. Chemomet Intell Lab Syst. 2000;50(1):1–18. doi: 10.1016/s0169-7439(99)00047-7.
[38] Varmuza K, Filzmoser P. Introduction to multivariate statistical analysis in chemometrics. In: Weiner IB, editor. Handbook of Psychology, Research Methods in Psychology, vol. 2. Hoboken, New Yersey, USA: CRC Press, John Wiley & Sons; 2003.
[39] Hadi A, Simonoff J. Procedures for the identification of multiple outliers in linear models. J Am Stat Assoc. 1993;88:1264–72.
[40] Egan W, Morgan S. Outlier detection in multivariate analytical chemical data. Analytical Chemistry. 1998;70:2372–79.
[41] Meyer-Baese A, Schmid V. Statistical and syntactic pattern recognition. In: Pattern Recognition and Signal Analysis in Medical Imaging, 2nd ed. Oxford, UK: Academic Press, Elsevier Inc.; 2014. doi: 10.1016/B978-0-12-409545-8.00006-6.
[42] Turney PD, Pantel P. From frequency to meaning: vector space models of semantics. J Art Intell Res. 2010;37:141–88.
[43] Bullinaria J, Levy J. Evaluating wordnet-based measures of semantic distance. Comput Linguist. 2007;32(1):13–47.
[44] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2018, https://www.R-project.org/ (archived on 15 March 2019).
[45] Bakuła M, Kaźmierczak R, Grunwald G. Analysis of the possibilities for applying the ASG-EUPOS system services for establishing the detailed control networks. Tech Sci. 2011;142:217–28.
[46] Dawidowicz K, Krzan G, Świątek K. Urban area GPS positioning accuracy using ASG-EUPOS POZGEO service as a function of session duration. Artificial Satellites. 2014;49(1):33–42. doi: 10.2478/arsa-2014-0003.
[47] Pażus R. Spojrzenie na ASG-EUPOS od strony użytkownika serwisu POZGEO, cz. III. Geodeta. 2009;5:25–8.