# Evaluation of accuracy of digital map data via multiple comparisons

## A. DOSKOCZ[*1] and W. REJCHEL[2]

[1]Faculty of Geodesy, Geospatial and Civil Engineering, University of Warmia and Mazury, 15 Prawochenskiego St., 10-720 Olsztyn, Poland

[2]Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, 12/18 Chopina St., 87-100 Torun, Poland

**Abstract.** Digital map data sets (or geo-databases) are an important part of the spatial data infrastructure (SDI) of the European Community. Different methods of producing large-scale map data are described in the paper, and the aim is to compare the accuracy of these methods. Our analysis is based on statistical tools belonging to the multiple comparisons theory. The first method is the well-known analysis of variance (ANOVA), and the second one is the rank-based method. The latter approach, which is rarely used in geodetic research, allows us to determine the order of the considered methods with respect to the positional accuracy of digital map data that they produce. Using this approach, one can identify the least accurate set of map data or a fragment of a map that should be updated by a new direct survey. The rank-based methods can also be rather easily applied to other technical (engineering) disciplines, e.g. geodesy and cartography.

**Key words:** digital map data, evaluation of positional accuracy, analysis of variance (ANOVA), multiple comparisons, rank-based methods.

## 1. Introduction

Different methods of producing digital map data are described in the paper. Large-scale digital map data, especially of urbanized territory (e.g. data sets acquired for producing cadastral maps and base maps), are an important part of the spatial data infrastructure (SDI) of the European Community (so-called INSPIRE data) [1]. The national geodetic and cartographic resources in Poland were obtained in recent decades using different coordinate systems and methods of data acquisition [2–4]. All data are currently integrated and harmonized in the national spatial reference system [5]. The key aspect to integrating geographic data and their interoperability in the spatial data infrastructure is the accuracy of digital databases [6]. The aim of the paper is to compare distinct methods of producing digital map data and to rank them from the most accurate to the least. Obviously, knowing the least accurate fragment of a map, one can update it by performing a new direct survey. The updated and accurate map data should be used for land-use planning in realization of investments, as well as for the risk assessment of decision-making about the localization of objects (e.g. [7]).

It is well-known that a quality digital data description includes: genealogy, completeness, logical consistency, and also thematic, temporal, and positional accuracy [8]. The positional accuracy of data is crucial in the modern economy, in particular to support decision-making processes for planning and realization of investments supported by large-scale base maps. The base maps are created as standard cartographic products in 1:500–1:5000 scales, and are developed from data sets and data bases of reference data, e.g.: topographic objects (BDOT500 database), land and buildings registries (EGiB database), geodetic registries of utilities (GESUT database). These data bases were obtained in recent decades using different methods of data acquisition.

The INSPIRE directive [1] requires the Commission to establish a geoportal, and the member states of the European Union shall provide access for this geoportal to their infrastructures, as well as to any access points they themselves decide to operate. A release of the INSPIRE geoportal has been published for enhancing access to European spatial data [9]. It enables searching through spatial data sets and spatial data services from the EU member states within the framework of the INSPIRE directive.

Furthermore, the European location framework (ELF) BaseMap project [10] provides platforms for accessing INSPIRE-based reference data [11]. This project comprises the datasets provided by national mapping and cadastral authorities (NMCAs) and (in the future) by other data providers [12]. The pilot-phase ELF BaseMap service began in spring 2015 for the Nordic cluster and the Netherlands. Meanwhile, more countries (Denmark, Poland, Czech Republic, and Belgium) have started producing national ELF BaseMap contributions. The recommended production process seems to be feasible, but the map concept needs further revision. In addition, the INSPIRE data models are generally too complex, and not adapted for mapping purposes: overlaps between INSPIRE themes, potential presence of deprecated features, and sometimes attributes of properties are not directly attached to geometry [13]. According to the ELF BaseMap project, the datasets prepared by the NMCAs have to be compatible with INSPIRE data specifications [14].

An interesting programme was realized in Great Britain by Ordnance Survey (OS). OS defined the "absolute positional accuracy" as a comparison between the coordinates of a point on the map and the coordinates of the same point on the ground (in the British National Grid Reference System). In

years 2001–2006, the positional accuracy improvement (PAI) programme was realized, checking and improving the absolute positional accuracy of all maps covering rural areas of Great Britain [15]. The result of the PAI Programme was: the root mean square error (RMSE) of the absolute positional accuracy of ±0.4 m for the re-survey of rural towns, and the RMSE of the overall absolute positional accuracy of ±1.1 m for all other 1:2500 scale rural areas.

The positional accuracy of different methods of producing digital map data plays a key role in the current paper. The statistical analysis basing on multiple comparisons (multiple hypothesis testing) is used to compare the quality of these methods. The proposed approach can be rather easily applied to other branches of geodesy and cartography.

Large-scale maps are exploitable in many economic issues, for example in detailed localization of objects for land-use planning. Computer technology development enables converting them to digital maps. To produce large-scale digital map data, one has to plan a geodetic control network and a survey of field data (with total station, RTK/RTN GNSS, or smart station) or acquire data using other methods (digitalization or scanning of analogue maps). There are many methods of producing large-scale map data [16]. They are briefly presented on Fig. 1.
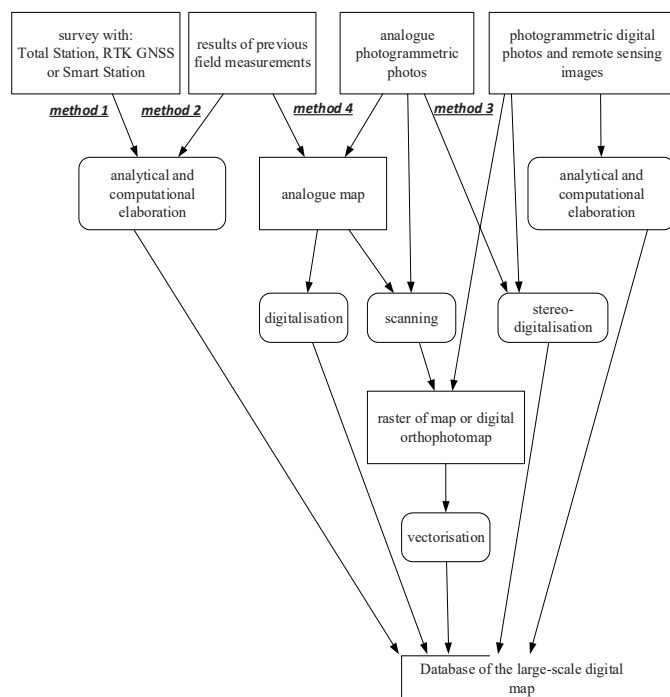


Fig. 1. Methods of situational data acquisition for large-scale maps [27]

The positional accuracy of digital map data is investigated by many researchers [16–18] on the basis of statistical analysis [19–22], as well as other methods [23–27]. They usually concern medium- and small-scale maps. However, in Poland there also exist large-scale maps of the entire country in the form of base maps that can be investigated as well.

In the paper four methods of producing digital map data that are used to create large-scale maps are considered:

- Method 1: a digital map in 1:500 scale was produced on the basis of direct survey performed in "Kortowo" campus using a total station,
- Method 2: a digital map was produced using existing results of surveys that were performed using the orthogonal measurements method and, in recent years, the polar method using a total station (in relation to the restorable third-order network). The data was obtained from the Municipal Surveying Documentation Centre (MSDC) of Zielona Góra,
- Method 3: a digital orthophoto map of Olsztyn was produced basing on 1:5000 aerial photographs. Aerial photographs were processed to a digital form using a matrix scanner with the resolution of 1000 dpi. This source data was provided to photogrammetric products with ground sample distance (GSD) 10–15 cm. Then, the digital orthophoto map was developed in scale of 1:2000. The data was obtained from the MSDC of Olsztyn,
- Method 4: a digital map was produced using a graphical-and-digital processing method (by vectorization) of analogue base maps at a scale of 1:500 with layers of utilities at scales of 1:500 and 1:1000. Layers of vector digital maps were obtained from the MSDC of Olsztyn.

For every considered method, there is a sample containing its errors. Namely, for every procedure, control points within investigated areas were randomly selected. Then, the coordinates $(X, Y)$ of every control point were determined by means of new field surveys that were performed with a total station. Besides, the coordinates $(X', Y')$ of every control point were also collected using database reports for methods 1, 2, and 4. For the third method, they were obtained by manual vectorisation of the raster image of the orthophoto map. The error of a given control point is the Euclidean distance between $(X, Y)$ and $(X', Y')$:

$$Z = \sqrt{(X'-X)^2 + (Y'-Y)^2}.$$ (1)

For the first method, there were $n_1 = 484$ control points covering the area of approximately 200 ha, and for the second one, $n_2 = 1636$ control points covering the area of 330 ha (out of 5800 ha of the total area of the city) were considered. Besides, $n_3 = 311$ errors from the area of 115 ha were calculated for the third procedure, and $n_4 = 2287$ errors for the last one, for an area of 355 ha (out of 8800 ha of the total area of the city).

As it has been already mentioned, comparing the accuracy of digital map data is challenging, but nevertheless, that is the aim of the current paper. The standard statistical analysis of these four data sets containing errors of the methods did not bring the expected results [28]. The main reason was that the considered data sets contain outliers and are not normally distributed, which is quite typical while working with real data sets, such as geodetic surveys. Thus, the argumentation

used in the current paper is based on rank methods [29] that are known in geodetic research, e.g. [30]. These methods are based on replacing each element of the sample by its rank, i.e. the number of this element in the sorted sample from least to greatest. Obviously, doing that, some information on the observed phenomena is lost, but this approach makes the further reasoning free of inconvenient assumptions and more robust. Preliminary (and encouraging) results are contained in [31]. Using rank methods, confidence intervals were constructed and hypothesis tests concerning the errors of the methods were performed. But the weakness of these results is that they are "individual". They do not take into consideration the fact that one compares four methods simultaneously, which needs more involved argumentation. In this paper, multiple comparisons are used to investigate the accuracy of the considered methods. They are briefly introduced in the next section. In [32] or [33], one can find more comprehensive description of this problem. In Section 3, the analysis of variance (ANOVA) is applied. This method is probably the most popular, if one chooses to compare several methods. However, this approach can be applied only if some "regularity conditions" for the considered methods (via data sets containing errors of these methods) are satisfied. Unfortunately, for the considered data sets these assumptions are violated, and that makes the inference based on ANOVA deceptive. To overcome this difficulty, methods based on ranks are applied in Section 4.

The main contribution of the current paper to geodetic research is the application of rank methods to multiple comparisons and demonstrating the supremacy of this approach over the widely-used ANOVA. Indeed, using rank methods, one can unambiguously determine the ordering between the accuracy of considered methods. It means that one can easily indicate the method producing the digital map with the highest accuracy, as well as the second one in this ranking, and so on. In the paper the positional accuracy of large-scale digital map data is compared, but problems that are met, advantages and shortcomings of the described statistical methods are often found in other engineering sciences.

Finally, all calculations contained in the paper are obtained using "R" software [34].

## 2. Multiple comparisons

If the aim is to compare two methods (their means, medians etc.), then one can use the parametric Student's t-test or the non-parametric Wilcoxon's rank sum test to find out whether there is a significant difference between methods, and which one is "better". In case of several, say $K$, competitive approaches, one can apply analogs of the aforementioned tests (the $F$-test or the Kruskal-Wallis test), but these tests only tell us whether all procedures give the same results. If this hypothesis should be rejected, they are not able to indicate which one is better or worse than others. To do it, the methods can be compared in pairs, but the first one should define the significance level of this multiple comparison, i.e. $K(K-1)/2$ tests. The experiment-wise significant level is the probability of making one or more type 1

errors (false discoveries, rejecting a true null hypothesis) among all the hypotheses while performing multiple hypothesis tests. Therefore, $K(K-1)/2$ individual Studen's or Wilcoxon's tests on the standard significance level (say $\alpha = 0.05$) cannot be used, because the experiment-wise significant level would increase, and too many true null hypotheses would be rejected. Methods of solving this problem have been developed since the fifties of the twentieth century [32, 33]. The simplest approach is the Bonferroni correction that is based on the fact that to have the experiment-wise significant level $\alpha$, it is enough to perform every individual test on the significance level $\alpha/M$, where $M$ is the number of performed tests. The weakness of this procedure is its "conservativeness" which means that it rarely rejects null hypotheses. In the literature, there are many improvements of this procedure, for instance the Tukey's honest significant difference (HSD) test [35], the Scheffé's test [36], or the Holm's procedure [37]. The first two methods relate to ANOVA and the former of the two is applied in the next section. In Section 4, the approach based on rank methods is introduced. It significantly improves statistical inference, especially if one works with survey data sets that often contain outliers and violate the assumptions of ANOVA.

## 3. Analysis of variance (ANOVA)

The task is to compare four data sets containing errors of methods that were used to produce digital maps.

Table 1 Descriptive statistics

|  | *m* [m] | *med* [m] | *sd* [m] | *iqr* [m] |
|---|---|---|---|---|
| method 1 | 0.06 | 0.04 | 0.08 | 0.04 |
| method 2 | 0.20 | 0.12 | 0.54 | 0.14 |
| method 3 | 0.18 | 0.16 | 0.11 | 0.15 |
| method 4 | 0.34 | 0.28 | 0.27 | 0.27 |

In the preliminary phase of this analysis, elementary descriptive statistics are calculated. They are contained in Table 1, where: $m$ denotes the mean, similarly $med$ – median, $sd$ – standard deviation, $iqr$ – interquartile range. Histograms and boxplots of errors are also presented in Fig. 2 and Fig. 3, respectively.

This rough analysis suggests that there are differences in the accuracy of the considered methods: the first one seems to be the best one, while the last one gives relatively poor results. Besides, one can notice that distributions of errors are positively skew, not distributed normally, and that samples probably contain outliers (especially the second and the fourth one). These facts influence the further analysis, which is based on multiple comparisons.

The analysis of variance (ANOVA) is very popular and facilitative in solving problems similar to the considered ones. The number of elements in the sample is denoted by $n_1$, $n_2$, and
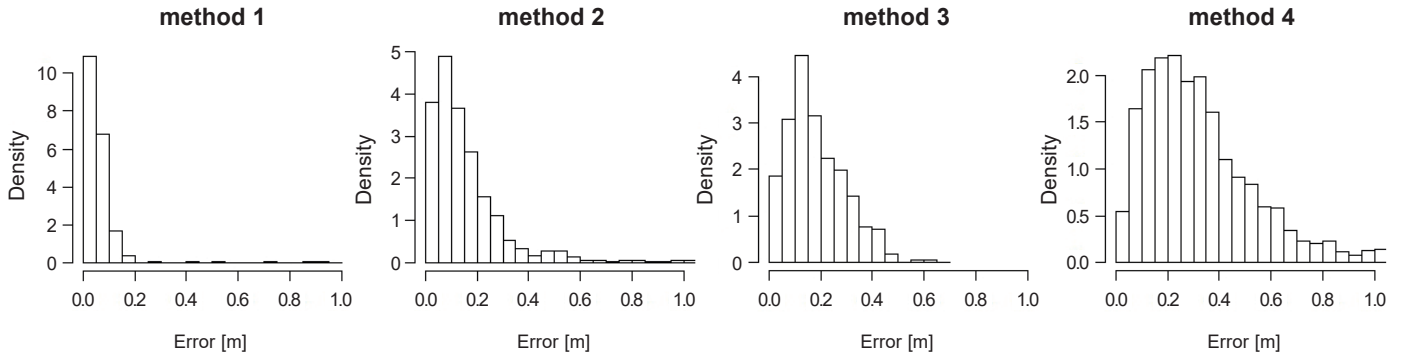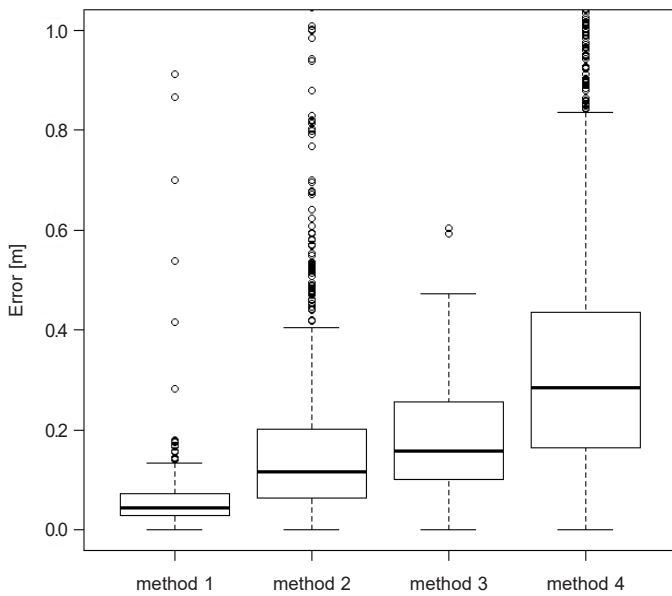
Fig. 2. Histograms of errors



Fig. 3. Boxplots of errors

$n_3$ correspondingly, and $n = n_1 + n_2 + n_3 + n_4$ is the overall number of elements (in our case $n = 4718$). The following model is considered

$$Z_{ij} = \mu + \mu_j + \varepsilon_{ij}, \quad i = 1, \ldots, n_j, \quad j = 1, 2, 3, 4, \quad (2)$$

where $Z_{ij}$ is the $i$-th element of the sample related to the $j$-th method. Besides, $\mu$ is the (unknown) overall mean, $\mu_j$ is the (also unknown) effect of the $j$-th method and $\sum_{j=1}^{4} \mu_j = 0$. The last equation makes the values $\mu, \mu_1, \mu_2, \mu_3, \mu_4$ uniquely determined. Notice that in (1), the error $Z_{ij}$ is decomposed into two parts: the first one $(\mu + \mu_j)$ is a nonrandom, "systematic" error that describes the accuracy of $j$-th method, while the second component $\varepsilon_{ij}$ is a random error. This additive model is natural, and seems to reflect the real problems quite well.

The main assumption in ANOVA is that random errors $\varepsilon_{ij}$ are independent random variables having the normal distribution $N(0, \sigma^2)$ In particular, it means that errors $\varepsilon_{ij}$ have the same variance $\sigma^2$, regardless of the sample they come from.

The accuracy of methods is compared by testing hypotheses concerning their $\mu_j$ effects. But first one needs a few equations:

$$\bar{z}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} z_{ij}, \qquad \bar{z} = \frac{1}{n} \sum_{j=1}^{4} \sum_{i=1}^{n_j} z_{ij},$$

$$SSB = \sum_{j=1}^{4} n_j (\bar{z}_j - \bar{z})^2, \qquad SSW = \sum_{j=1}^{4} \sum_{i=1}^{n_j} (z_{ij} - \bar{z}_j)^2. \quad (3)$$

Therefore, $\bar{z}_j$ is the empirical mean of errors in the $j$-th sample, while $\bar{z}$ is the overall empirical mean. Besides, $SSB$ is the sum of squares between methods, and $SSW$ is the sum of squares within methods. They both (after norming) estimate the variance $\sigma^2$. The first step in ANOVA is checking if there is any difference between the considered methods (their $\mu_j$ effects). Thus, the well-known $F$-test is performed:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,$$

$$H_1 : \textit{at least two of } \mu_j \textit{ effects are different.} \quad (4)$$

This elementary test is based on comparing values of $SSB/3$ and $SSW/(n-4)$. Denominators 3 and $(n-4)$ are degrees of freedom of $SSB$ and $SSW$ correspondingly, if the hypothesis $H_0$ is true and variables $\varepsilon_{ij}$ satisfy the aforementioned assumptions. In the considered problem, the $p$-value of the $F$-test is less than $10^{-5}$, so the considered samples are significantly distinct. However, to find out which of them differ, one should perform tests of pairs of methods, comparing their effects, this being the place where multiple testing methods are involved. Thus, the problem of testing six null hypotheses simultaneously is considered:

$$H_0^{jk} : \mu_j = \mu_k \qquad H_1^{jk} : \mu_j \neq \mu_k, \quad (5)$$

where $j < k, j, k = 1, 2, 3, 4$. The analysis is based on the widely-used Tukey's test [36]. This procedure rejects the hypothesis $H_0^{jk}$, if:

$$|\bar{z}_j - \bar{z}_k| > q^* \sqrt{\frac{SSW}{2} \left( \frac{1}{n_j} + \frac{1}{n_k} \right)}, \quad (6)$$

where $q^*$ is the critical value of the test [35]. In Table 2, $p$-values of Tukey's tests are given, while the accuracy of four methods used to produce large-scale digital maps is compared.

Table 2
$p$-values of Tukey's test

|  | method 1 | method 2 | method 3 | method 4 |
|---|---|---|---|---|
| method 1 | – | $<10^{-5}$ | $5 \cdot 10^{-5}$ | $<10^{-5}$ |
| method 2 | – | – | 0.76781 | $<10^{-5}$ |
| method 3 | – | – | – | $<10^{-5}$ |

One can notice that the accuracy of the first method is different from all other cases, since the corresponding $p$-values are close to zero. The similar situation occurs when the last method is considered. However, one cannot find the difference between the second and third method, because the $p$-value is 0.76781. Besides, regularity assumptions (given above) are not satisfied in this problem: samples do not have normal distributions and their variances differ. It has been already conjectured on the basis of the preliminary analysis that is given at the beginning of this section. These suspicions are confirmed by the Shapiro-Wilk test (all $p$-values less than $<10^{-5}$) and the Levene test ($p$-value also less than $<10^{-5}$). Thus, ANOVA is not the appropriate tool to investigate the considered problem, and conclusions formulated basing on it could be incorrect. It can be seen while comparing Table 2 and Table 3 (Section 4). Thus, one should use statistical methods that do not need such strong assumptions to be satisfied. The natural candidates are procedures related to ranks.

## 4. Rank-based methods

The model that is considered in this section is similar to (2), but one does not need as strong assumptions as in the previous section. Namely, one only assumes that random errors $\varepsilon_{ij}$ in (2) are independent and come from the same continuous population. In fact, the assumption that considered populations should be the same can be weakened (see Chapter 6 in [29]). Parameters $\mu$ and $\mu_j$ play the same roles as in the previous section. One, again, compares the accuracy of methods by comparing their effects $\mu_j$'s Notice that the proceeding is similar to the one from Section 3. The main difference (and the reason for the "success") is replacing the values of errors with their ranks.

Table 3
$p$-values of test based on ranks

|  | method 1 | method 2 | method 3 | method 4 |
|---|---|---|---|---|
| method 1 | – | $<10^{-5}$ | $<10^{-5}$ | $<10^{-5}$ |
| method 2 | – | – | $3 \cdot 10^{-5}$ | $<10^{-5}$ |
| method 3 | – | – | – | $<10^{-5}$ |

Thus, one first calculates the ranks of objects, that is sorts all $n$ observations from least to greatest, and the rank $r_{ij}$ of an element $z_{ij}$ is the number of $z_{ij}$ in the sorted list of objects. Moreover, let

$$\overline{R}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} r_{ij}, \qquad j = 1, 2, 3, 4 \qquad (7)$$

be the mean of errors ranks of the $j$-th method. The starting point is the test that checks if all four methods have the same accuracy. Thus, one again tests

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_1 : \text{at least two of } \mu_j \text{ effects are different,}$$
$$(8)$$

but now, the Kruskal-Wallis test [29] is used. It can be viewed as a rank analog of $F$-test from the previous section. The $p$-value of this test is less than $10^{-5}$, which means that the four considered samples are significantly distinct. To find out which of them differ, multiple comparisons are used. Namely, the Bonferroni correction is applied, so six null hypotheses are tested jointly:

$$H_0^{jk} : \mu_j = \mu_k \qquad H_1^{jk} : \mu_j < \mu_k, \qquad (9)$$

where $j < k$, $j, k = 1, 2, 3, 4$. The method proposed in [38] is used. It rejects $H_0^{jk}$, if:

$$\overline{R}_j - \overline{R}_k < -u_{\alpha/6} \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_j} + \frac{1}{n_k} \right)}, \qquad (10)$$

where $u_{\alpha/6} = \Phi^{-1}(1 - \alpha/6)$ and $\Phi$ is the standard normal distribution function. Besides, the value $\alpha$ is the experiment-wise significant level. In fact, this method is an approximation of the pure procedure [29], but the sizes of the considered samples are varying, so this approximation works well, and at the same time, calculations are much easier than in the strict algorithm. The results ($p$-values of comparison tests) are given in Table 3.

Method 1 has the highest accuracy, because $p$-values of tests that compare $\mu_1$ with $\mu_2$, $\mu_3$ and $\mu_4$ are close to zero. Moreover, method 2 is better than method 3 and method 4. It is worth to stress that using tests based on ranks, one can find the better method among the second and the third one. It was not possible while using ANOVA in the previous section. Finally, method 4 is the least accurate.

It is important to emphasize the advantages of rank-based methods. Obviously, they need only weak assumptions to be satisfied, comparing to ANOVA. Moreover, using the very conservative Bonferroni correction, one can indicate differences between the accuracy of methods. In the previous section, applied was the Tukey's test which is much less conservative, but it did not manage to estimate the methods' accuracy. Furthermore, improvements of the Bonferroni correction for ranking the methods can be also found in the literature, for instance the Holm's procedure [37].

## 5. Conclusions

The positional accuracy of different methods of producing digital map data is investigated in this paper. Two approaches to this problem are described, both belonging to the multiple comparisons theory. The first of them is the well-known ANOVA that gives reliable results only if quite strong assumptions are satisfied. Unfortunately, they are often violated while working with surveying data sets. As a remedy, multiple comparisons based on ranking methods are used. Applying them, one can reach the aim of the paper that is comparing the positional accuracy of the considered methods of producing data sets. Indeed, the statistical analysis confirms the highest positional accuracy of the digital map produced on the basis of survey with a total station (method 1). The second place in the ranking is occupied by the digital map produced using past field surveys (method 2). Finally, the data stored in digital orthophoto map (method 3) is characterized by higher positional accuracy than the digital map data obtained by means of graphical-and-digital processing of analogue large-scale maps (method 4). However, the obtained results are only qualitative. They should be completed by a quantitative comparison of the accuracy of methods. This will be the aim of future investigations.

In this paper, it was demonstrated that the use of geo-databases or digital map data sets to support decision-making meets with difficulties, because decisions are based on combined data from different sources. Therefore, information (metadata) about the positional accuracy of data plays a key role in GIS (geographic information system) and other support systems. In authors' opinion, the evaluation of the positional accuracy of digital map data via multiple comparisons is promising in automation of assessment of the quality of databases, and can help with producing an expert system for management of geo-data and their rational use (e.g. [39]). This is particularly important in the case of basic maps, which due to their purpose, are the most accurate cartographic product [40].

## REFERENCES

[1] "Directive 2007/2/EC of European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)", *Official Journal of the European Union* L108 (50), (2007).

[2] "Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 9 listopada 2011 r. w sprawie standardów technicznych wykonywania geodezyjnych pomiarów sytuacyjnych i wysokościowych oraz opracowywania i przekazywania wyników tych pomiarów do państwowego zasobu geodezyjnego i kartograficznego", *Dz.U. 2011 nr 263 poz. 1572*, (2011), [in Polish].

[3] "Rozporządzenie Ministra Administracji i Cyfryzacji z dnia 21 października 2015 r. w sprawie powiatowej bazy GESUT i krajowej bazy GESUT", Dz.U. poz. 1938, (2015), [in Polish].

[4] „Rozporządzenie Ministra Administracji i Cyfryzacji z dnia 2 listopada 2015 r. w sprawie bazy danych obiektów topograficznych oraz mapy zasadniczej", *Dz.U. 2015 poz. 2028*, (2015), [in Polish].

[5] "Rozporządzenie Rady Ministrów z dnia 15 października 2012 r. w sprawie państwowego systemu odniesień przestrzennych", *Dz.U. poz. 1247*, (2012), [in Polish].

[6] C. Rönsdorf, "Positional integration of geodata. Positional accuracy improvement: Impacts of improving the positional accuracy of GI databases", *EuroSDR Publication* 48 – Related Papers, 2004.

[7] A. Doskocz, "Accuracy assessment of planimetric large-scale map data for decision-making", *Geodesy and Cartography* 65 (1), doi: 10.1515/geocart-2016–0006, (2016).

[8] "Elements of spatial data quality", *International Cartographic Associations, Commission on Spatial Data Quality*, eds. S.C. Guptill, and J.L. Morrison, Pergamon Press, Oxford, 1995.

[9] "INSPIRE geoportal of infrastructure for spatial information in the European Community", http://inspire-geoportal.ec.europa.eu, 2016.

[10] "European Location Framework (ELF) service – a single point of access for harmonised reference data from national mapping, Cadastre and Land Registry Authorities", http://www.location-framework.eu, (2016).

[11] A. Hopfstock, D. Laurent, and T. Ellett von Brasch, "ELF BaseMap – Offering a european reference map service", in *Proceedings Paper of 1st ICA European Symposium on Cartography*, eds. G. Gartner, and H. Huang, EuroCarto, Vienna, 2015.

[12] Complex project of reference geo-information source for Europe, http://www.elfproject.eu, 2016.

[13] A. Hopfstock and D. Laurent, "Using INSPIRE data for cartographic purpose: The BaseMap experience", in *Geospatial World Forum*, Lisbon, 2015.

[14] A. Jakobsson, "Using INSPIRE data for cartographic purpose: The BaseMap experience", in *Geospatial World Forum*, Lisbon, 2015.

[15] "Positional accuracy improvement programme companion", Ordnance Survey, Southampton, http://www.ordnancesurvey.co.uk/pai, (2004).

[16] W. Dąbrowski and A. Doskocz, "Estimation of accuracy of the large-scale digital topographic map data", in *Proceedings Paper of 7th International Conference on Environmental Engineering* 1–3, ed. D. Cygas, VGTU Press Technika, Vilnius, 1293–1299 (2008).

[17] K. Husar, "The accuracy of digital spatial data", *Kartografické Listy* 4, 69–78 (1996), [in Slovak].

[18] F.J.A. Lopez and A.D.A. Gordo, "Analysis of some positional accuracy assessment methodologies", *Journal of Surveying Engineering* 134, 45–54 (2008).

[19] P.V. Bolstad, P. Gessler, and T.M. Lillesand, "Positional uncertainty in manually digitized map data", *Int. J. Geogr. Inf. Systems* 4, 399–412 (1990).

[20] A. Zakarevicius and I. Jonauskiene, "Opportunity to use statistical control for quality assessment of land parcels cadastral data", *Geodesy and Cartography* 33, 115–119 (2007), [in Lithuanian].

[21] P.A. Zandbergen, "Positional accuracy of spatial data: Non-normal distributions and a critique of the national standard for spatial data accuracy", *Transactions in GIS* 12, 103–130 (2008).

[22] A. Doskocz, "The statistical approach for estimation of horizontal uncertainty of the points of digital large-scale maps", *Geodetski List* 67 (90) No. 3, 201–212 (2013).

[23] P. Bogaert, J. Delince and S. Kay, "Assessing the error of polygonal area measurements: A general formulation with applications to agriculture", *Measurement Science and Technology* 16, 1170–1178 (2005).

[24] A. Croitoru and Y. Doytsher, "Accounting for discontinuities in cadastral data accuracy: Toward a patch-based approach", in *Papers of FIG Working Week*, Paris, 2003.

[25] S. de Bruin, "Modelling positional uncertainty of line features by accounting for stochastic deviations from straight line segments", *Transactions in GIS* 12, 165–177 (2008).

[26] T. Podobnikar, "Modelling and visualisation of spatial data error. Monte Carlo simulations in Slovenia", *GIM International* 13, 47–49 (1999).

[27] A. Doskocz, "Methodology for assessing the accuracy of digital large-scale maps", *Dissertations and Monographs* 193, Publishing House of University of Warmia and Mazury in Olsztyn, 2013, [in Polish].

[28] A. Doskocz, "The use of statistical analysis for estimation of positional accuracy of large-scale digital maps", *Geodesy and Cartography* 54, 131–150 (2005).

[29] M. Hollander and D.A. Wolfe, "Nonparametric statistical methods", Wiley, New York, 1999.

[30] R. Duchnowski, "Robustness of strategy for testing leveling mark stability based on rank tests", *Survey Review* 43, 687–699 (2011).

[31] A. Doskocz and W. Rejchel, "Proposition of automatization of analysis of accuracy of large-scale digital maps databases", *Journal of Civil Engineering, Environment and Architecture* 59, 85–93, (2012), [in Polish].

[32] J.W. Tukey, "The philosophy of multiple comparisons", *Statistical Science* 6, 100–116 (1991).

[33] Y. Benjamini and H. Braun, "John W. Tukey's contributions to multiple comparisons", *The Annals of Statistics* 30, 1576–1594 (2002).

[34] R Development Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, 2015. http://www.R-project.org.

[35] J.W. Tukey, "The problem of multiple comparisons", *The Collected Works of John W. Tukey*, vol. VIII, ed. H.I. Braun, Chapman and Hall, New York, 1994.

[36] H. Scheffé, "A method for judging all contrasts in the analysis of variance", *Biometrika* 40, 87–104 (1953).

[37] S. Holm, "A simple sequentially rejective multiple test procedure", *Scandinavian Journal of Statistics* 6, 65–70 (1979).

[38] O.J. Dunn, "Multiple comparisons using rank sums", *Technometrics* 6, 241–252 (1964).

[39] A. Doskocz, "About accuracy of analytical determination of areas for cadastre and other purposes", in *Proceedings Paper of 9th International Conference on Environmental Engineering*, doi: 10.3846/enviro.2014.203, eds. D. Cygas and T. Tollazzi, VGTU Press Technika, Vilnius, 2014.

[40] "Ustawa z dnia 17 maja 1989 r. Prawo geodezyjne i kartograficzne", Dz.U. poz. 1433 (2015), [in Polish].