TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 13, Y 2010

DOI 10.2478/v10022-010-0013-8

INITIAL USE OF STATISTICAL ANALYSIS FOR ESTIMATION OF HORIZONTAL ACCURACY OF LARGE-SCALE DIGITAL MAPS

Adam Doskocz

Chair of Surveying
University of Warmia and Mazury in Olsztyn

K e y w o r d s: digital map, horizontal accuracy, statistical analysis.

A b s t r a c t

The paper presents estimation of horizontal accuracy of maps using of initial statistical analysis. Investigations have been performed for four large-scale digital maps made with different methods of producing digital map data: new total station survey (object *A*), re-calculation of previous direct measurements (orthogonal and polar surveys) (object *B*), manual vectorisation of a raster orthophotomap image (object *C*) and graphical-and-digital processing of analogue maps (object *D*).

Analysis has been performed for large statistical samples of sets of vectors of shift of control points ε_L and their components, i.e. true errors of increments of co-ordinates ε_X, ε_Y. In the case of a map produced by means of new survey with an electronic tacheometer, the true errors were represented by differences between co-ordinates of control points obtained from two separate set outs. In the case of other methods of data collection for digital map production true errors were represented by differences of co-ordinates acquired from an investigated map and co-ordinates calculated from new direct surveys.

WSTĘPNE WYKORZYSTANIE ANALIZY STATYSTYCZNEJ DO OCENY DOKŁADNOŚCI OPRACOWANIA SYTUACYJNEGO WIELKOSKALOWYCH MAP CYFROWYCH

Adam Doskocz

Katedra Geodezji Szczegółowej
Uniwersytet Warmińsko-Mazurski w Olsztynie

S ł o w a k l u c z o w e: mapa cyfrowa, dokładność opracowania sytuacyjnego, analiza statystyczna.

A b s t r a k t

W pracy przedstawiono ocenę dokładności opracowania sytuacyjnego map numerycznych z wykorzystaniem wstępnej analizy statystycznej. Badania zrealizowano na przykładzie czterech wielkoskalowych map numerycznych wykonanych różnymi metodami pozyskiwania danych numerycznych: nowy pomiar tachimetrem elektronicznym (obiekt *A*), przeliczenie wyników wcześniejszych pomiarów bezpośrednich (ortogonalnych i biegunowych) (obiekt *B*), manualną wektoryzację rastrowego obrazu ortofotomapy (obiekt *C*) oraz przetworzenie graficzno-numeryczne map analogowych (obiekt *D*).

Analizę wykonano na dużych próbach statystycznych zbiorów długości wektorów przesunięcia punktów kontrolnych ε_L oraz jego składowych, tj. błędach prawdziwych przyrostów współrzędnych ε_X, ε_Y. W przypadku mapy wykonanej z nowych pomiarów tachimetrem elektronicznym jako błędy prawdziwe przyjęto różnice dwukrotnie wyznaczonych współrzędnych punktów kontrolnych. W odniesieniu do pozostałych metod pozyskiwania danych do tworzenia map numerycznych błędy prawdziwe uzyskano z różnic współrzędnych pozyskanych z badanej mapy i współrzędnych wyznaczonych z nowego pomiaru bezpośredniego.

Introduction

The feature of horizontal accuracy of the digital map database is very important for state geodetic administration and surveying documentation centres, and for good relation between producers and users of digital map data (DĄBROWSKI et al. 2007).

The discussed issue is not a local problem, that occurs in Poland only. Accuracy of digital data has already been investigated by numerous researchers (e.g. HUSÁR 1996). Those investigations are also performed with the use of statistical analysis (BOLSTAD et al. 1990), as well as other, modern research methods (PODOBNIKAR 1999). They concern, however, maps produced at medium and small scales available in particular countries. In Poland, the large-scale maps exist for the entire country in the form of the base map; the investigations were thus focused on that map (in opinion author of this paper).

The aspect of direct relations between results of analysis and correctness (accuracy) of data (the GIGO principle, i.e. "Garbage In, Garbage Out") (URBAŃSKI 1997) has been often discussed in literature. Correlations between data quality and costs of data acquisition are also stressed. Costs of survey depend, in general, on the accuracy of survey (GAŹDZICKI 1995).

The paper presents the initial use of statistical analysis for estimation of horizontal accuracy of large-scale digital maps.

Methodology of research and characteristics of investigated objects

Statistical analysis for estimation of accuracy of digital maps was executed basing on true errors ε_X, ε_Y of increments of plain coordinates of geometric details of the 1st accuracy group (of selected details of the 1st group, being the so-called, control points). In the case of a map produced on the basis of measurements with the use of an electronic tacheometer (object *A*), the differences of coordinates of control points, surveyed two times (following the theory of measurements in pairs) were used as true errors. With respect to other analysed methods of acquisition of location details used for creation of large-scale digital maps (objects *B*, *C* and *D*), true errors of control points were calculated basing on differences of coordinates obtained from the investigated map and coordinates determined from a new field survey.

Coordinates of analysed details of locations of the objects *A*, *B* and *D* were acquired in the form of text listings or database reports; in the case of the object *C* coordinates were acquired by means of manual vectorisation of the raster image of the orthophotomap (DOSKOCZ 2002).

Four digital maps (objects) for the cities of Olsztyn and Zielona Góra were the subject of research. The object *A* is the digital map in the scale 1:500, produced on the basis of direct survey performed at the university campus in Olsztyn in 1995–1996 by student teams with use of an electronic tacheometer (total station survey). The survey was based on the restorable 3rd order geodetic control network. 481 control points covering the area of approximately 200 ha were used for estimation of accuracy.

The object *B* is the digital base map of the City of Zielona Góra. The digital map has been produced basing on existing results of surveys, performed in the period of 1974–1999 (basing on technical traversing of the 2nd order from 1973–1974, developed in accordance with the B–III Instruction), by means of the method of orthogonal measurements, and, in the recent period, by means of the polar method, using an electronic tacheometer (in relation to the restorable 3rd order network). The map accuracy was evaluated for the area of approximately 330 ha (out of 5800 ha of the total area of the city), using 1619 control points.

The object *C* is the digital orthophotomap of the City of Olsztyn. The digital orthophotomap was produced basing on 1:5000 aerial photographs taken within the Phare Programme in 1995. Aerial photographs were processed to the digital form using a matrix scanner with the resolution of 1000 dpi; then the orthophotomap was developed at the scale of 1:2000. The orthophotomap accuracy was evaluated basing on 311 control points from the area of approximately 115 ha.

The object *D* is the digital base map of the City of Olsztyn. The digital map was produced using the method of graphical-and-digital processing of the analogue base map at the scale of 1:500 with the layers of utilities, at the scales of 1:500 and 1:1000. The base map was produced basing on technical traversing of the 2nd order from 1974. Survey of details was performed by means of a photogrammetric method, for initial conditions of October 1977. The map was updated by means of direct surveys, connected to the control network of 1974, and, after 1986 – to the newly established restorable control network of the 3rd order. Accuracy of the digital map was estimated for the area of approximately 355 ha (out of 8800 ha, being the total area of the city), using 2282 control points.

Statistical analysis for estimation of horizontal accuracy of large-scale maps

In the course of statistical analysis for estimation of horizontal accuracy of large-scale digital maps based on sets of true errors, estimation of parameters of distribution of the random variable and verification of compliance of its distribution with theoretical models of errors were performed. It highly influenced the statistical estimation of accuracy of control objects, since – in the case when compliance between the distribution of both sets of true errors ε_X, ε_Y of the given control objects with the normal distribution is stated – the random variable ε_L^2 ($\varepsilon_L^2 = (\varepsilon_X^2 + \varepsilon_Y^2)$) – the square of the length of the shift vector of the control point) would have the χ^2 distribution with two degrees of freedom.

For the needs of surveying, the εL type variable is often used. That variable of two degrees of freedom, expresses the length of the vector, the components of which represent two independent random variables of normal distribution and of equal standard deviations. Some transformations of the χ^2 distribution of practical value are known from literature (CRAMÉR 1958). An example of such a variable is the linear error of the point location (NEY 1976).

The strategy of estimation of accuracy of large-scale digital maps with use of initial statistical analysis of true errors is shown in Table 1.

Determination of parameters of a random variable

Two groups of parameters, i.e. measures of location and measures of scattering are most frequently applied. The basic numerical characteristics of the distribution of the random variable are particular cases of moments of the random variable (BARAN 1983).

Some values of the basic numerical characteristics are helpful to specify (hypothetical) reasons of anomalies of distributions of empirical sets (SZACHERSKA, KURPIEWSKA 1976, WIŚNIEWSKI 1986):

– Existing skewness ($S \neq 0$) in an empirical set proves that the set was created from combination of subsets of various size and of various expected values. This also proves that a deterministic factor occurs (a factor of non-random characteristics) which is systematic with respect to the sign but variable with respect to the value.

Table 1

Stages of initial statistical analysis for estimation of horizontal accuracy of map

Statistical analysis of the sets of true errors ε_X, ε_Y	
I stage: Determination of parameters of the random variable	
Calculation of empirical parameters of the random variable "true error"	
II stage: Analysis of distribution of the sets of true errors Determination of the type of the random variable "true error" (continuous or discrete type)	
Formulation of initial assumptions on the distribution of true errors (basing on empirical parameters and determined type of the random variable)	
Verification of non-parametric hypotheses with respect to compliance of distribution of the sets of true errors ε_X, ε_Y with the normal distribution	
III stage: Estimation of accuracy of control objects (independence of ε_X, ε_Y errors assumed)	
In the case when compliance of distribution of the sets of true errors ε_X, ε_Y of the given control object with the normal distribution is stated	*In the case when no compliance of distribution of at least one of the sets of true errors ε_X or ε_Y of the given control object with the normal distribution is stated*
Estimation of accuracy of the control object using the χ^2 distribution (with two degrees of freedom) for the random variable $\varepsilon_L{}^2$ ("the square of length of the shift vector of the control point")	Estimation of accuracy of the control object with the use of statistical definition of probability, where the probability of occurrence is expressed as the relative frequency of occurrence – occurrence of a specified length of the shift vector (ε_L) in the set of control points of the given object (for the sufficient number of sample population)

- Theoretical justification of $e > 0$ coefficient (slender empirical curve) is the occurrence of elements of various accuracy in a set.
- Flatteness of an empirical curve ($e < 0$) proves that systematic errors occur.
- Non-zero expected value of the error ($\bar{x} \neq 0$) indicates the occurrence of permanent errors.

In the initial statistical analysis of true errors the following values have been empirically determined: the average values $\bar{x} = \sum_{i=1}^{N} \varepsilon_{X_i}/N$ (expressed by its estimator – the arithmetic mean) and standard deviation of the random variable $m = \sqrt{\sum_{i=1}^{N} \varepsilon_{X_i}{}^2/N}$ (estimated by the mean square error). Asymmetry of the distribution has been specified by asymmetry factor (skewness) $S = \mu_3/m^3$ and the level of flattening of curves of empirical distribution with respect to theoretical models has also been determined. The factor of flattening (excess) has been calculated as $e = (\mu_4/m^4)$-3. The following quantities have also been calculated: the mean square error of the average value of the random variable $m_{\bar{x}} = m/\sqrt{N}$, the average error of the random variable $d = \sum_{i=1}^{N} |\varepsilon_{X_i}|/N$, and the d/m ratio, where N is the size of the set. Parameters of random variables of control sets investigated are given in Table 2 and Table 3.

Table 2

Empirical parameters of the random variables from control objects of digital maps of the City of Olsztyn

Control object (size)	Set of errors	Size of the set	Empirical parameters of random variables						
			$\bar{x}$ [m]	m [m]	$\bar{x}$ [m]	d [m]	d/m	S	e
A (200 ha)	ε_X	478	0.00	0.05	0.002	0.04	0.80	-0.23	0.04
	ε_Y	478	0.00	0.04	0.002	0.03	0.75	0.01	2.58
C (115 ha)	ε_X	311	-0.01	0.15	0.008	0.11	0.73	0.12	0.64
	ε_Y	311	-0.02	0.15	0.008	0.11	0.73	0.51	1.35
D-1 (46 ha)	ε_X	1001	0.10	0.26	0.008	0.19	0.73	-0.16	2.25
	ε_Y	1001	-0.08	0.28	0.009	0.21	0.75	0.31	1.54
D-2 (154 ha)	ε_X	549	0.10	0.31	0.013	0.23	0.74	-0.32	2.27
	ε_Y	549	-0.09	0.32	0.014	0.24	0.75	0.22	1.34
D-3 (25 ha)	ε_X	240	-0.06	0.31	0.020	0.23	0.74	0.00	0.95
	ε_Y	240	0.06	0.34	0.022	0.27	0.79	-0.03	0.34
D-4 (20 ha)	ε_X	236	0.01	0.22	0.014	0.16	0.73	-0.79	3.25
	ε_Y	236	0.05	0.25	0.016	0.17	0.64	-0.61	4.24
D-5 (100 ha)	ε_X	134	0.11	0.24	0.018	0.18	0.75	-0.60	1.26
	ε_Y	134	-0.02	0.18	0.016	0.14	0.78	0.06	1.17
D-6 (10 ha)	ε_X	115	0.10	0.29	0.027	0.23	0.79	-0.11	-0.77
	ε_Y	115	-0.08	0.25	0.023	0.19	0.76	0.14	0.22

Table 3

Empirical parameters of the random variables from control objects of digital base map of the City of Zielona Góra

Control object (size)	Set of errors	Size of the set	Empirical parameters of random variables						
			$\bar{x}$ [m]	m [m]	$m_{\bar{x}}$ [m]	d [m]	d/m	S	e
B-1 (92 ha)	ε_X	257	-0.04	0.11	0.007	0.08	0.73	0.14	0.70
	ε_Y	257	-0.01	0.10	0.006	0.07	0.70	0.56	1.73
B-2 (13 ha)	ε_X	217	0.00	0.14	0.010	0.09	0.64	0.77	3.46
	ε_Y	217	0.00	0.16	0.011	0.10	0.62	-0.77	9.91
B-3 (46 ha)	ε_X	209	-0.02	0.20	0.013	0.14	0.70	-0.80	2.38
	ε_Y	209	0.06	0.27	0.019	0.17	0.63	1.47	4.61
B-4 (60 ha)	ε_X	241	0.07	0.15	0.010	0.12	0.80	0.08	-0.15
	ε_Y	241	-0.06	0.15	0.010	0.11	0.73	0.15	1.30
B-5 (90 ha)	ε_X	318	0.00	0.12	0.007	0.08	0.67	1.09	13.90
	ε_Y	318	0.00	0.15	0.008	0.07	0.47	1.11	39.34
B-6 (18 ha)	ε_X	264	-0.01	0.11	0.007	0.08	0.73	1.31	4.34
	ε_Y	264	-0.01	0.09	0.006	0.06	0.67	0.03	9.59
B-7 total (7 ha)	ε_X	96	0.05	0.26	0.026	0.16	0.62	1.58	2.57
	ε_Y	96	0.15	0.31	0.032	0.20	0.64	1.15	0.39
B-7 sheet 2(2) (5 ha)	ε_X	72	0.04	0.13	0.015	0.09	0.69	0.43	0.59
	ε_Y	72	-0.06	0.16	0.019	0.12	0.75	-0.34	-0.88

Analysis of distribution of true errors

The second stage of analysis was focused on the determination of the distribution of empirical sets. At the beginning the type of the random variables "true error of increment of coordinates" ε_X and ε_Y was specified. Initially the type of the random variable was visually recognised, since in empirical sets of discrete random variables the number of elements of various values would be small (less then twenty). However, in the case of a theoretically continuous variable each element of a set has different value but the same value of variable may occur several times (NEY 1976). Finally, the type of the random variable was specified with use of the hypothesis of wave structure of errors (ADAMCZEWSKI 1961). In the case of sets of true errors ε_X, ε_Y of all control objects, the continuous function of attenuation has been stated. For example, in the case of set with greatest number of elements (Fig. 1) and set with least number of elements (Fig. 2).

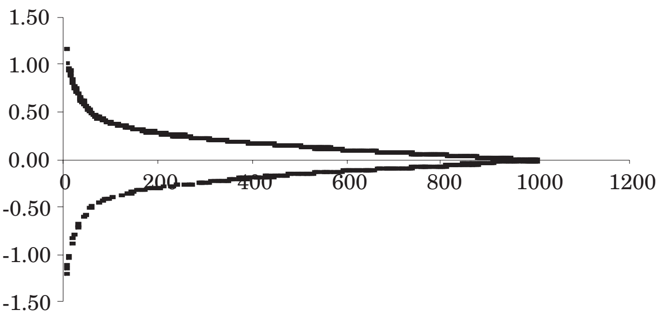


Fig. 1. Continuous function of attenuation of the empirical set ε_X of *D*-1 control object (set with greatest number of elements)

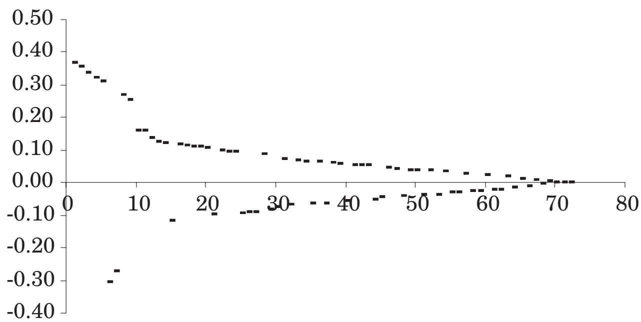


Fig. 2. Continuous function of attenuation of the empirical set ε_X of *B*-7 control object (set with least number of elements)

The specified type of random variable and calculated empirical parameters of the random variable (Tables 2 and 3) allowed to assume that distributions of investigated sets of true errors may belong tò a family' of normal distributions. Besides, basing on empirical parameters, initial specification of distribution of true errors was performed. It was noted that the majority of sets are asymmetric ($S \neq 0$) and are affected by a systematic factor ($\bar{x} \neq 0$). It was also found that, for the considerable part of sets, that obtained d/m ratio vary from its theoretical value in the normal distribution (for Gaussian distribution that ratio equals to $d/m = \sqrt{2/\pi} \cong 0{,}80$ (NEY 1976)). The excess values in empirical sets are almost all positive, what indicates more distinctive slenderness of empirical curves of probability density of errors than a normal curve of the variance determined in the given empirical set. Therefore, the use of modified normal distribution, in particular the Modulated Normal Distribution (MND) (ROMANOWSKI 1974, 1979) was considered. The analysis of relative frequency of true errors ε_X, ε_Y in the investigated sets indicates the similarity of their distribution with the MND distribution of the 1st-type.

The empirical sets of observation errors frequently substantially differ from the Gaussian model; it concerns asymmetry and excess (SZACHERSKA 1974, ŁYSZKOWICZ 1975b, WIŚNIEWSKI 1984, 1986). Such differences were also exhibited by the empirical sets investigated. In order to determine the empirical distribution density curves, Pearson method was applied, that requires calculating numerical coefficients β_1, β_2, k (basing on values of central moments of empirical sets, calculated earlier): β_1, μ_3^2/μ_3^2, β_2, μ_4/μ_2^2, $k = \beta_1 \cdot (\beta_2 + 3)^2 / 4 \cdot (2 \cdot \beta_2 - 3 \cdot \beta_1 - 6) \cdot (4 \cdot \beta_2 - 3 \cdot \beta_1)$. Those coefficients become the basis for determination the curve, out of 12 Pearson curves, that might be used for approximation of the given empirical set. Pearson distributions of I, II, IV and VII types were identified as suitable for typical surveying data (ŁYSZKOWICZ 1975a, 1975b).

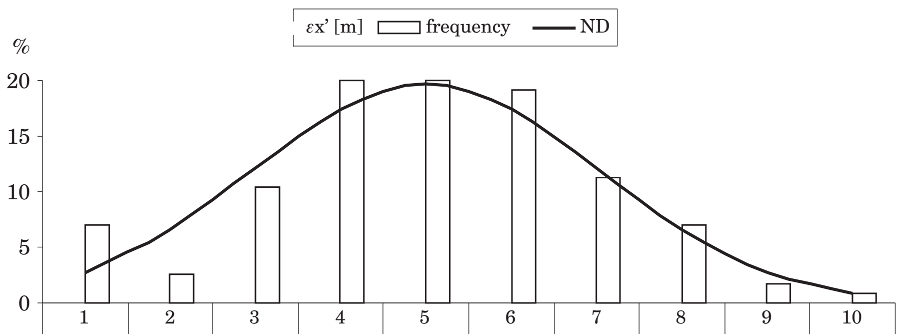
The search for Pearson distributions that might be used for approximation of empirical sets was undertaken. Besides, two statistical tests, i.e. Pearson χ^2 test and Kołmogorow λ test were used for verification of non-parametric hypotheses concerning the compliance of distributions of empirical sets with theoretical distributions (Gaussian and MND 1st-type distribution characterized by the leptokurtosis index ω_t). Investigations of compliance of distribution of true errors of ε_X, ε_Y with theoretical models was performed for the standardized random variable (e.g. $\varepsilon_X' = (\varepsilon_X - \bar{x})/m$).

The χ^2 test is of higher sensitivity than λ test, but, at the same time, it requires larger set (at least above one hundred elements) (NEY 1970). The χ^2 test better characterizes the difference between the empirical distribution and the theoretical model, since it uses all class intervals of a distributive series. In the λ test the level of compliance between the empirical and theoretical distributions is expressed by the point value D_i. The argument that supports the use of the Kołmogorow λ method is that the λ test is not sensitive on discrepancies which occur for final small class intervals of the distributive series (which often prevail in empirical sets) (HELLWIG 1998). The Kołmogorow λ method was applied when it was necessary as an auxiliary test.

Analysed the sets ε_X, ε_Y are slender (leptokurtic) in majority of cases, with respect to Gaussian distribution (except of ε_X and ε_Y sets of the object A, the ε_Y set of the object D-5 and the ε_Y set of the object B-7); prevailing discrepancies of their distribution occur in central class intervals of the distributive series. Detailed results of investigations of compliance between distributions of sets of true errors with theoretical distributions are presented in the paper (DOSKOCZ 2005).

Summary of results of analysis of distribution of true errors

In the first two stages of statistical analysis (Tab. 1) it was found that the distributions of ε_X, ε_Y sets "are governed by their own law" which makes it difficult to present them by means of known, theoretical models of errors. Only in a few cases the compliance between empirical distribution with the theoretical distribution was stated. The only set which proved the compliance between the distribution of the empirical set and the normal distribution (ND) was the empirical set ε_X of D-6 control object (Fig. 3).



εx' [m]	-0.580	-0.435	-0.290	-0.145	0	0.145	0.290	0.435	0.580	0.725
Frequency	0.070	0.026	0.104	0.200	0.200	0.191	0.113	0.070	0.017	0.009
ND	0.0278	0.0656	0.1210	0.1747	0.1974	0.1747	0.1210	0.0656	0.0278	0.0092

Fig 3. Compliance between the distribution of the empirical set ε_X of D-6 control object with the normal distribution

Distributions of investigated sets of true errors are, in general, not compliant with MND Romanowski distributions (DOSKOCZ 2005), what results from their skewness as it is known from literaturè utilisation of modulated normal distributions is not fully consistent in conditions of existing asymmetry' (WIŚNIEWSKI 1986).

Also Pearson curves, which accept the presence of asymmetry and excess in an empirical empirical set, could, to a very limited extend (ε_Y set of A, ε_Y set of *B*-7, ε_X and ε_Y sets of *D*-3, ε_Y set of *D*-5) approximate investigated sets of true errors (DOSKOCZ 2005).

Estimation of accuracy of investigated digital maps

Because of lack of compliance between the random variable "the square of the length of the shift vector" and the theoretical distribution χ^2, the probability of occurrence of specified values of the point location error m_p (expressed by the length of the vector of shift of the control point ε_L) has been estimated on the basis of the relative frequency W_N of occurrence of particular values ε_L within the control objects. The required probability was determined by means of the method presented in (SMIRNOW, DUNIN-BARKOWSKI 1969). Probabilities have been estimated for the confidence level $\gamma = 0.997$ for $t_\gamma = 3$.

The confidence interval for the unknown probability may be determined by that method when the condition $N \cdot P \cdot Q > 9$ is met (N is the size of the set; P is the event probability to be determined; Q is the probability of the complementary event). This condition define sufficient size of the samples who is needed for estimating probability of an analysed event.

The limits of confidence interval for the estimated probability P are expressed as $P_1(W, N) < P < P_2(W, N)$ (SMIRNOW, DUNIN-BARKOWSKI 1969), where

$$P_1(W, N) = \frac{2 \cdot N \cdot W_N + t_\gamma \cdot \sqrt{D}}{2 \cdot (N + t_\gamma^2)}$$

$$P_2(W, N) = \frac{2 \cdot N \cdot W_N + t_\gamma \cdot \sqrt{D}}{2 \cdot (N + t_\gamma^2)}$$

$$\sqrt{D} = \sqrt{4 \cdot N \cdot W_N \cdot (1 - W_N) + t_\gamma^2}$$

Statistical estimation of the horizontal accuracy of large-scale digital maps in particular control objects is given in Tables 4 and 5. Columns 6 and 8, present probabilities that the map error does not exceed stated point location error (m_P) multiplied by factor 3 and 2, respectively, at the given control object. Column 10 presents the probability that the theoretical error will be exceeded

(m_t – determined in accordance to the accuracy standard for mapping). The symbol (*) by the confidence interval means that the probability is determined in situation when the relative frequency of the considered event in the empirical set had not ensured that the condition of the sufficient sample size was fulfilled.

Table 4

Statistical estimation of horizontal accuracy of control objects on digital maps of the City of Olsztyn

Control object	Analysed value	Size of the set	m_P [m]	Statistical characteristics of accuracy of control objects determined for the confidence level $\gamma = 0.997$					
				Estimated probabilities of occurrence of the (Δ) error in general population					
				W_N	$\Delta \leq 3m_p$	W_N	$\Delta \leq 2m_p$	W_N	$\Delta > m_t$
1	2	3	4	5	6	7	8	9	10
A	ε_L	478	0.04	0.95	$0.91<P<0.97$	–	–	0.02	$0.01<P<0.05$*
C	ε_L	311	0.21	–	–	0.97	$0.92<P<0.99$	0	$0<P<0.03$*
D-1	ε_L	1001	0.38	0.99	$0.98<P<1$*	0.95	$0.93<P<0.97$	0.19	$0.16<P<0.24$
D-2	ε_L	549	0.45	0.99	$0.97<P<1$*	0.95	$0.92<P<0.97$	0.27	$0.22<P<0.33$
D-3	ε_L	240	0.46	1	$0.96<P<1$*	0.95	$0.90<P<0.98$	0.32	$0.24<P<0.42$
D-4	ε_L	236	0.33	0.98	$0.93<P<0.99$*	0.96	$0.90<P<0.98$*	0.15	$0.09<P<0.23$
D-5	ε_L	134	0.30	–	–	0.96	$0.88<P<0.99$*	0.14	$0.07<P<0.26$
D-6	ε_L	115	0.39	–	–	0.97	$0.88<P<0.99$*	0.29	$0.18<P<0.42$

Table 5

Statistical estimation of horizontal accuracy of control objects on the digital base map of the City of Zielona Góra

Control object	Analysed value	Size of the set	m_P [m]	Statistical characteristics of accuracy of control objects determined for the confidence level $\gamma = 0.997$					
				Estimated probabilities of occurrence of the (Δ) error in general population					
				W_N	$\Delta \leq 3m_p$	W_N	$\Delta \leq 2m_p$	W_N	$\Delta > m_t$
1	2	3	4	5	6	7	8	9	10
B-1	ε_L	257	0.15	–	–	0.97	$0.92<P<0.99$*	0.004	$0<P<0.04$*
B-2	ε_L	217	0.21	0.99	$0.94<P<1$*	0.96	$0.90<P<0.98$*	0.03	$0.01<P<0.09$*
B-3	ε_L	209	0.33	0.98	$0.93<P<1$*	0.92	$0.85<P<0.96$	0.13	$0.08<P<0.22$
B-4	ε_L	241	0.22	–	–	0.94	$0.87<P<0.97$	0.05	$0.02<P<0.12$
B-5	ε_L	318	0.20	0.98	$0.94<P<0.99$*	0.97	$0.93<P<0.99$*	0.03	$0.01<P<0.07$*
B-6	ε_L	264	0.14	0.98	$0.94<P<0.99$*	0.96	$0.90<P<0.98$	0.02	$0<P<0.06$
B-7 total	ε_L	96	0.41	0.98	$0.88<P<1$*	0.91	$0.78<P<0.96$*	0.15	$0.07<P<0.28$
B-7 sheet 2(2)	ε_L	72	0.21	–	–	0.97	$0.84<P<1$*	0.03	$0<P<0.16$*

Conclusions

Results of statistical analysis proved the lack of compliance of considered sets of true errors with the normal distribution (except ε_X set of the control object *D*-6). That is why estimation of horizontal accuracy of investigated digital maps was based on statistical definition of the probability, where the probability of an event is identified with the relative frequency of this event. Obtained results of accuracy estimation using statistical analysis are coherent with conclusions developed on the basis of classical estimation of accuracy (DĄBROWSKI, DOSKOCZ 2008).

Initial statistical analysis confirmed high accuracy of the digital map produced on the basis of survey with an electronic tacheometer (object *A*). Probability that the theoretical error (0.3 mm at the map scale) will be exceeded in the object *A* does not exceed 5% (Table 4, col. 10).

The probability of occurrence of the point location error which does not meet the accuracy standards in the case of a digital map produced on the bases of the past field surveys is relatively low (object *B*); in general it does not exceed 10% (except control objects *B*-3 and *B*-7) (Table 5, col. 10).

Statistical analysis confirmed the high accuracy (in relation to well identified details of the 1st group, i.e. characteristic points of technical utilities) of a digital orthophotomap (object *C*). In the case of that object the probability of occurrence errors larger than the theoretical error does not exceed 3% (Table 4, col. 10).

The lowest accuracy was confirmed for a digital map produced by means of graphical-and-digital processing of analogue maps (object *D*). The estimated probability of occurrence of the point location error that exceeds the theoretical error value (0.3 mm at the map scale) within the object *D* equals to 20 ÷ 30% (Table 4, col. 10).

Results of performed investigations allow to formulated the general conclusions:

1) Large-scale digital maps based on data acquired from various methods do not always meet the requirements of accuracy standards specified in technical standards.

2) There is a need for estimation accuracy of large-scale digital maps to ensure an appropriate quality of the state surveying resources.

3) In author opinion of this paper, future investigations should be focused on automation process of estimation of digital map data bases with application mathematical and statistical methods.

Acknowledgments

This work is the result of own research within the frames of the university internal grant (theme) No. 522-0309-0213 "Creation and updating of databases of digital maps for the needs of surveying administration and for educational purposes".

Accepted for print 10.01.2010

References

ADAMCZEWSKI Z. 1961. *Hypothesis of wave structure of the errors* (in Polish). Geodezja i Kartografia, X (2): 151–159.

BARAN L.W. 1983. *Theoretical bases of elaborate of results of the geodetic surveys*. PWN, Warsaw.

BOLSTAD P.V., GESSLER P., LILLESAND T.M. 1990. *Positional uncertainty in manually digitized map data*, Int. J. Geogr. Inf. Systems, 4(4): 399–412.

CRAMÉR H. 1958. *Mathematical methods in statistic*. PWN, Warsaw.

DĄBROWSKI W., DOSKOCZ A., MRÓWCZYŃSKI T. 2007. *Utility accuracy of large-scale digital maps of the City of Zielona Góra – The city appreciated orthophoto*. GEODETA, 1(140): 26–28, http://www.geoforum.pl/pages/index.php?page=edition&subpage=artykuly&id=30

DĄBROWSKI W., DOSKOCZ A. 2008. *Estimation of accuracy of the large-scale digital topographic map data*. Proceeding of the 7th International Conference "Environmental Engineering" Volume III, pp. 1293–1299. https://www.vgtu.lt/upload/leid_konf/dabrowski_et_al_estimation.pdf

DOSKOCZ A. 2002. *Examining of accuracy of large-scale digital maps made with different methods.* Ph.D. Dissertation, University of Warmia and Mazury in Olsztyn, p. 174.

DOSKOCZ A. 2005. *The use of statistical analysis for estimation of positional accuracy of large-scale digital maps*. Geodezja i Kartografia, 54(3): 131–150.

GAŹDZICKI J. 1995. *Cadastral systems.* PPWK, Warsaw–Wroclaw.

HELLWIG Z. 1998. *Elements of probability and mathematical statistic.* PWN, Warsaw.

HUSÁR K. 1996. *Presnost' digitálnych priestorových údajov.* Kartografickiε Listy, 4: 69–78.

ŁYSZKOWICZ A. 1975a. *Define of Pearson curves description of sets of errors of the geodetic surveys.* Geodezja i Kartografia, XXIV, 4: 275–282.

ŁYSZKOWICZ A. 1975b. *Choice of mathematical models for empirical distribution of results of the geodetic surveys.* Ph.D. Dissertation, University of Warmia and Mazury in Olsztyn.

NEY B. 1970. *Criteria of compliance of empirical distribution with models*. Polska Akademia Nauk – Oddział w Krakowie, Prace Komisji Górniczo-Geodezyjnej, Seria Geodezja, 7: 31–46.

NEY B. 1976. *Statistical methods in geodesy*. AGH Uczelniane Wydawnictwa Naukowo-Dydaktyczne, 497, Kraków.

PODOBNIKAR T. 1999. *Modelling and Visualisation of Spatial Data Error. Monte Carlo Simulations in Slovenia*, GIM International, 13(7): 47–49.

ROMANOWSKI M. 1974. *Theory of modulation of the random errors in observations*. Polska Akademia Nauk – Oddział w Krakowie, Prace Komisji Górniczo-Geodezyjnej, Seria Geodezja, 18: 19–46.

ROMANOWSKI M. 1979. *Random Errors in Observations and the Influence of Modulation on their Distribution.* Verlag Konrad Wittwer, Stuttgart.

SMIRNOW N.W., DUNIN-BARKOWSKI I.W. 1969. *Course of probability and mathematical statistic in technical applications*. PWN, Warsaw.

SZACHERSKA M.K. 1974. *Model of composition of errors of the geodetic surveys*. Geodezja i Kartografia, XXIII (1): 21–51.

SZACHERSKA M.K., KURPIEWSKA A. 1976. *Accuracy of precision leveling in light of statistical investigations*. Geodezja i Kartografia, t. XXV, 1: 17–33.

URBAŃSKI J. 1997. *Understand GIS – analysis of spatial information*. PWN, Warsaw.
WIŚNIEWSKI Z. 1984. *Use of Pearson distributions II and VII types in adjustment of the geodetic networks*. Geodezja i Kartografia, XXXIII (3): 85–104.
WIŚNIEWSKI Z. 1986. *Adjustment of geodetic networks with use of probabilistic models of errors of the surveys*. Acta Acad. Agricult. Techn. Olst. Geodaesia et Ruris Regulatio, 15, Supplementum C., University of Warmia and Mazury in Olsztyn.