Local height transformation through polynomial regression

Marcin Ligas, Piotr Banasik

AGH University of Science and Technology,
Faculty of Mining Surveying and Environmental Engineering
Department of Geomatics, 30 Mickiewicza Al., 30-059 Krakow, Poland
e-mail: ligas@agh.edu.pl, pbanasik@agh.edu.pl

Abstract: The paper presents results of the transformation between two height systems Kronstadt'60 and Kronstadt'86 within the area of Krakow's district, the latter system being nowadays a part of National Spatial Reference System in Poland. The transformation between the two height systems was carried out based on the well known and frequently applied in geodesy polynomial regression. Despite the fact it is well known and frequently applied it is rather seldom broader tested against the optimal degree of a polynomial function, goodness of fit and its predictive capabilities. In this study some statistical tests, measures and techniques helpful in analyzing a polynomial transformation function (and not only) have been used.

Keywords: height transformation, polynomial regression, confidence intervals, crossvalidation

1. Introduction

A change of a reference system and what goes further horizontal and vertical datums on the area of a given country is usually a long term process. In Poland new geodetic datums were introduced in the year 2000 within the framework of the National Spatial Reference System (GUGiK, 2000). Since that time till now there have been works carried out in order to complete resolutions included in this document. They concern all existing geodetic resources, including geodetic networks. The introduction of the new vertical datum for the minor vertical networks (III – V class) is carried out by local governmental units such as districts. One of the methods used for computing heights in the compulsory system is a height transformation between "old" and "new" system preformed on the basis of benchmarks of the basic vertical control network, having heights in both systems. The essence of such transformation is an algorithm assuring the homogeneity and appropriate accuracy of computed heights.

In this study, in order to satisfy these two conditions a polynomial transformation model has been used. Its simplicity and well established theory in the framework of linear regression models make it very useful and common in solving (approximating) different problems within the field of geodesy. Height transformations are not

an exception in this respect (Hadaś and Bosy, 2009; Gucek and Basic, 2009; Osada, 2008). Although the use of polynomials in height transformations is recently mainly concentrated on the transformation between geometric heights (derived from GPS) and those referenced to the Earth's gravity field (orthometric, normal orthometric heights) the issue raised in this study (transformation between height systems Kronstadt'60 and Kronstadt'86) certainly is not of less practical importance. In this study the main focus is centered on testing polynomial transformation models in order to emerge the best one assuring acceptable accuracy with high level of certainty as to the heights of transformed points, thus some level of redundancy will be present in the following parts of this work.

2. Study area and data

The Krakow's district does not constitute a compact area. In the center there is the city of Krakow which is a separate territorial unit (Fig. 1). There are several leveling lines of the Ith class and tens of leveling lines of the IInd class of the basic (primary) vertical network crossing the area (Fig. 1). Information on this vertical control network in Kronstadt'60 system is contained in *Catalogues of Leveling Lines* (Katalog, 1960) whereas *Sketches of Leveling Lines* (Pulkowo'42 reference system, scale 1:100000) provide one with the localization of benchmarks necessary in the process of transforming between the two height systems. Determination of horizontal coordinates of benchmarks was possible by an appropriate calibration and digitization of *Sketches of Leveling Lines*.

Data on the basic vertical control network in Kronstadt'86 system are contained in digital registers provided by Central Geodetic and Cartographic Documentation Center. The horizontal coordinates of benchmarks contained in digital registers are determined in the (former) national horizontal system 1965. Confronting both data sources concerning the basic vertical network within the study area enabled a preliminary selection of benchmarks common for both height systems (176 benchmarks) and 150 benchmarks have left after a more detailed analysis (Fig. 1).

Minor vertical network (III-V class) within the area consists of about 600 benchmarks concentrated mainly near big cities. Heights of benchmarks in Kronstadt'60 system are contained in a digital database and localization may be acquired from analogue height overlays in 1:10000 scale. Digitization of benchmarks on overlays made the determination of their horizontal coordinates possible.

Uncertainty as to the identification of horizontal position of benchmarks of primary and minor vertical networks may vary considerably. An analysis revealed that the accuracy of horizontal coordinates may be on the level of several or tens of meters and in individual cases for the minor vertical network may exceed even 100 meters. Hence, and additional study of the impact of the horizontal position of control points on the transformation function was conducted.

Table 1 summarizes the dataset of 150 common points of the Ith class of the primary vertical network as to the height difference between systems Kronstadt'60 and
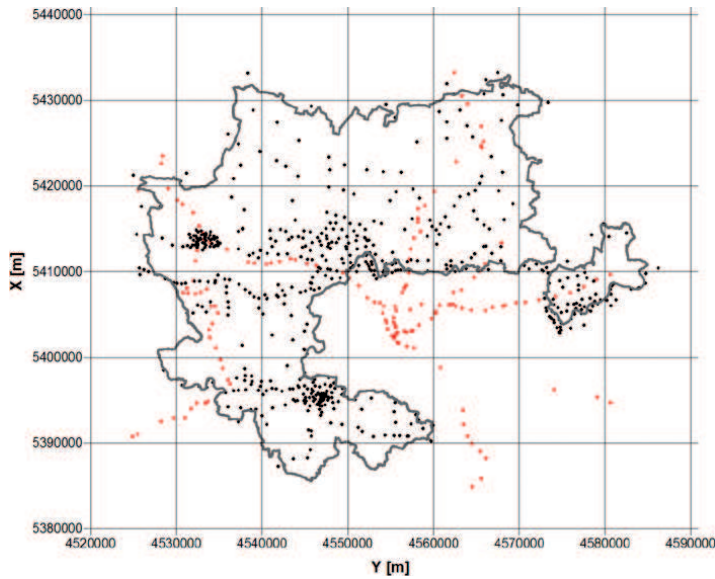
Fig. 1. The boundaries of Krakow's district (study area), red dots – the basic vertical control network
(I class), black dots – minor vertical network (III-V class)

Kronstadt'86 (dH$_{86-60}$). These points established the base for determining a transformation model.

Table 1. Summary of height differences (dH$_{86-60}$) on common benchmarks, in meters

Number of points	min	first quartile	median	third quartile	max	range	average	std. dev.
150	-0.0514	-0.0364	-0.0344	-0.0327	-0.0281	0.0233	-0.0351	0.0037

3. Transformation function

In order to find a suitable transformation function tying both height systems (Kronstadt'60 and Kronstadt'86) a polynomial model has been used. This kind of model has widespread applications in different fields of the Earth sciences i.e. geology, geophysics, geography, soil science and many others. It is very often associated with the name Trend Surface Analysis (TSA) which is the particular case of the well known in statistics multiple regression analysis (Davies, 2003; Schabenberger and Gotway, 2005). Particularity of the TSA relies on the fact that the independent variables of the model are spatial coordinates of the observed phenomena. Thus, in the paper the names: polynomial transformation function, polynomial regression and trend surface will be used exchangeable. As stated above, the polynomial transformation model is nothing but a polynomial regression thus the observed variable (or pseudo – observation, here the difference between the heights of the same point in the two height systems) is

decomposed into two components: a trend and a disturbance. The trend represents the global effect (average behavior of the phenomena under study) and the disturbance reveals the local effects (local fluctuations). This may be written as:

$$l_i = f(x_i, y_i) + \varepsilon_i \qquad (1)$$

where:

l_i – observed value (here the difference between the heights of the same point in the two height systems dH_{86-60})

x_i, y_i – spatial coordinates

ε_i – disturbance (error term)

General p^{th} degree trend surface is given as:

$$l_i = \sum_{k=0}^{p} \sum_{m=0}^{p} a_{km} x_i^k y_i^m + \varepsilon_i \quad k + m \le p \qquad (2)$$

or in compact matrix notation as:

$$\mathbf{L} = \mathbf{Xa} + \boldsymbol{\varepsilon} \qquad (3)$$

It is easy to notice that the number of coefficients of a polynomial increases rapidly with the degree of the polynomial, this may be expressed as $N = (p+1)(p+2)/2$, where N is the number of coefficients and p is the degree of a polynomial. Hence, the number of observations is strictly connected with the degree of the polynomial. A Polynomial approach being quite simple in use it also carries some load of difficulty. *It has an undesirable feature to wave the edges to fit points in the center. This effect is well known in polynomial regression, but is more severe in two or more dimensions where there are more boundaries to be affected* (Ripley, 2004). This drawback may be fixed to some extent by taking some buffer outside the study area. Also, from the numerical viewpoint it is known to be an ill – conditioned least squares problem, due to the fact that with higher and higher powers of spatial coordinates it leads to extreme values in the design matrix. From this reason a suitable transformation of spatial coordinates is necessary. The literature concerning the subject offers some transformation (centering/shifting/scaling) techniques (Shacham and Brauner, 1997), e.g.:

$$\begin{cases} x_i' = \dfrac{x_i}{x_{\max}} \\ y_i' = \dfrac{y_i}{y_{\max}} \end{cases} \qquad (4)$$

$$\begin{cases} x_i' = \dfrac{x_i - x_{\min}}{x_{\max} - x_{\min}} \\ y_i' = \dfrac{y_i - y_{\min}}{y_{\max} - y_{\min}} \end{cases} \qquad (5)$$

$$\begin{cases} x_i' = \dfrac{2x_i - x_{\max} - x_{\min}}{x_{\max} - x_{\min}} \\[2ex] y_i' = \dfrac{2y_i - y_{\max} - y_{\min}}{y_{\max} - y_{\min}} \end{cases} \tag{6}$$

$$\begin{cases} x_i' = \dfrac{x_i - \hat{x}}{s_x} \\[2ex] y_i' = \dfrac{y_i - \hat{y}}{s_y} \end{cases} \tag{7}$$

where:

x_i, y_i – original spatial coordinates

x_{min}, y_{min}, x_{max}, y_{max} – maximum and minimum values of original spatial coordinates

$\hat{x}, \hat{y}$, s_x, s_y – averages and standard deviations of original spatial coordinates

x_i', y_i' – spatial coordinates after transformation

Thus, applying one from the above transformation techniques one obtains a new form of (3) i.e.:

$$\mathbf{L} = \mathbf{X'b} + \boldsymbol{\varepsilon} \tag{8}$$

Extraction of coefficients **b** (**a**) from (8) or (3) is performed via ordinary least squares. The Table 2 reports estimates of parameters, standard errors as well as t – statistics and corresponding p – values for bilinear and quadratic trend surfaces fitted to data. All the parameters for both trend surfaces are statistically significant and this constitutes the base for further analysis. In addition, a graphical analysis of residuals for the quadratic transformation function as the further model of choice is presented in Table 2 and Figure 2. The graphical analysis supports the Jarque – Bera test verifying the normality of residuals.

Table 2 Estimates of parameters, standard errors, t – statistics for bilinear and quadratic trend surfaces

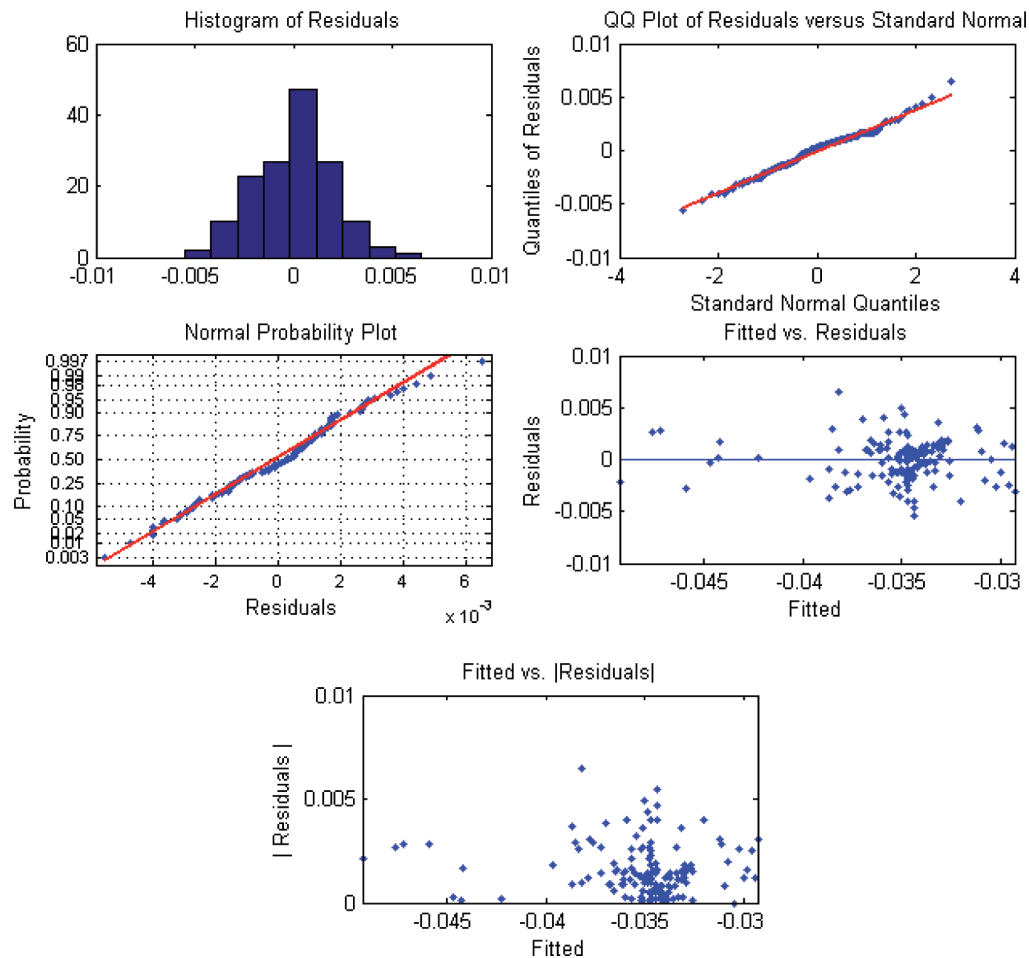| CASE		coeff.	std. error	t – stat	p – value
BILINEAR	a_{00}	-2.96949248E+02	3.26282517E+01	-9.10	0.0000
	a_{10}	5.49240737E-05	6.03532140E-06	9.10	0.0000
	a_{01}	6.55357889E-05	7.16638275E-06	9.14	0.0000
	a_{11}	-1.21229979E-11	1.32558219E-12	-9.15	0.0000
	On the basis of Jarque – Bera test there is no ground to reject the hypothesis of normality of residuals.				
QUADRATIC	a_{00}	-5.53054544E+02	4.65889392E+01	-11.87	0.0000
	a_{10}	1.29474514E-04	1.49210160E-05	8.68	0.0000
	a_{01}	8.95083119E-05	1.06008273E-05	8.44	0.0000
	a_{11}	-1.15434100E-11	1.16846584E-12	-9.88	0.0000
	a_{20}	-7.13728380E-12	1.36857683E-12	-5.22	0.0000
	a_{02}	-2.97831441E-12	8.67257557E-13	-3.43	0.0008
	On the basis of Jarque – Bera test there is no ground to reject the hypothesis of normality of residuals.				

Fig. 2. Graphical analysis of residuals for the quadratic transformation function

4. Significance of the polynomial transformation function and polynomial's degree

Analysis of variance for significance of bilinear and quadratic transformation functions has been used. The analysis of variance relies on dividing the total variation in the response variable, here **L** (height difference between the two height systems) into two components: SS_R – sum of squares of regression which measures the amount of variation explained by the model and SS_D (SS_E) – sum of squares of deviations (or errors) which on the other hand measures the residual variation that is not explained by the adopted model. The sum of these two quantities equals to the corrected total sum of squares SS_T (more on ANOVA may be found in e.g. Rao and Toutenburg, 1999;

Davies, 2003; Rawlings et al., 1998). The method of ANOVA is helpful in determining whether there is a relationship between response variable and the set of explanatory variables, in order to do this one tests the hypothesis:

$$H_0 : a_1 = a_2 = \ldots = a_k = 0$$
$$H_1 : \text{ not all } a_i = 0$$

(9)

Conclusions on the above hypothesis are drawn on the basis of the test statistics:

$$F = \frac{SS_R / (k - 1)}{SS_D / (n - k)}$$

(10)

which is compared to the critical value $F(\alpha, k\text{-}1, n\text{-}k)$ of the F distribution on the adopted significance level α with $(k\text{-}1, n\text{-}k)$ degrees of freedom. If $F > F(\alpha, k\text{-}1, n\text{-}k)$ hypothesis H_0 may be rejected and one may conclude that the response variable is related to one or more explanatory variables. Table 3 presents the results of analysis of variance for significance of analyzed models as well as the residual standard deviation m_0, and coefficient of determination R^2 and corrected coefficient of determination R^2_{corr} as measures of goodness of fit.

Table 3. Analysis of variance (ANOVA) for significance of polynomial regression (height transformation function) for bilinear and quadratic cases

CASE	Source of variation	Degrees of freedom		Sum of squares		Mean squares	F – test
BILINEAR	Total (uncorrected)	n	150	SS_U	0.186446		
	Total (corrected)	n − 1	149	SS_T	0.002072	0.000014	F = 80.52
	Polynomial regression	k − 1	3	SS_R	0.001291	0.000430	
	Deviations from polynomial	n − k	146	SS_D	0.000781	0.000005	
	$m_0 = \pm2.3mm$, $R^2 = 0.62$, $R^2_{corr} = 0.62$						
QUADRATIC	Total (uncorrected)	n	150	SS_U	0.186446		
	Total (corrected)	n − 1	149	SS_T	0.002072	0.000014	F = 73.98
	Polynomial regression	k − 1	5	SS_R	0.001491	0.000298	
	Deviations from polynomial	n − k	144	SS_D	0.000581	0.000004	
	$m_0 = \pm2mm$, $R^2 = 0.72$, $R^2_{corr} = 0.71$						

n – number of observations, k – total number of coefficients

Analyzing entries of Table 3 one notices that the hypothesis H_0 from (9) may be rejected in favor of H_1 what implies that there is some trend in the data (for both bilinear and quadratic trend surfaces). For the quadratic case the corrected coefficient of

determination reaches the value of 0.71 what may be interpreted as 71% of variability in the response variable (height differences) is explained by this model. This implies that the quadratic model should be better then the bilinear one.

As far as authors' knowledge goes there is not any a priori method answering the question what degree of the polynomial should be used to data at hand. This question is answered a posteriori usually by successive fitting higher degree polynomials to the data and each time inspecting the residual variance and the range of residuals. This is usually the base on which conclusions as to the polynomial's degree are drawn. Here, a more formal way of testing whether the gain associated with the increase of polynomial's degree is statistically significant is presented. This test relies also on some form of the analysis of variance but this time the difference between the sums of squares of regression of ($p+1$) and (p) degree is confronted with the sum of squares of deviations from the ($p+1$) degree model. The scheme of computation may be found in e.g. Davies (2003), Krumbein and Graybill, (1965). Here, only the results will be reported.

Table 4 presents the abovementioned analysis of variance for significance of increase from bilinear to quadratic trend surface, in fact statistically significant. This may constitute the formal confirmation of the choice of the quadratic polynomial transformation function based on analysis of variance.

Table 4. Analysis of variance for significance of increasing the degree of transformation function, bilinear vs. quadratic

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F – test
Quadratic regression	5	0.00149141	0.00029828	
Deviations from quadratic regression	144	0.00058057	0.00000403	
Bilinear regression	3	0.00129144	0.00043048	F = 24.80
Deviations from bilinear regression	146	0.00078054	0.00000535	
Increase form bilinear to quadratic regression	2	0.00019997	0.00009999	
Total variation	149	0.00207198	0.00001391	

5. Akaike Information Criterion

Another statistical tool useful in identification the best model among all nested models is Akaike Information Criterion (AIC). Among the models (estimated on the basis of the same dataset) one chooses the one with the lowest AIC value (single AIC value has no meaning) (Schabenberger and Gotway, 2005). AIC is defined by the following formula:

$$AIC = -2 \cdot \ln\left(L\left(\hat{\mathbf{a}}|\mathbf{L}\right)\right) + 2 \cdot u \tag{11}$$

In case of small samples ($n/k < \sim 40$) it is recommended to use corrected AIC expressed as:

$$AIC_c = -2 \cdot \ln(L(\hat{\mathbf{a}}|\mathbf{L})) + 2 \cdot u + \frac{2u(u+1)}{n-u-1} = AIC + \frac{2u(u+1)}{n-u-1} \qquad (12)$$

where:

$\ln(L(\hat{\mathbf{a}}|\mathbf{L}))$ – maximized value of likelihood function for estimated model

u – the number of estimated parameters including the residual variance

n – the sample size

When models are estimated with least squares and there is no ground to reject the hypothesis of normality of errors, AIC is given as:

$$AIC = n \ln\left(\frac{RSS}{n}\right) + 2 \cdot u \qquad (13)$$

where: RSS is residual sum of squares (sum of squares of deviations) and $\dfrac{RSS}{n}$ is a biased estimator of the residual variance

Table 5. Residual sum of squares, AIC and AIC_c for bilinear and quadratic trend surfaces

Model	Number of parameters	RSS	AIC	AIC_c
Bilinear	5	0.00078054	-1814.92	-1814.51
Quadratic	7	0.00058057	-1855.32	-1854.53

6. Crossvalidation

In the ideal case, one would like to have an extensive data set with as much additional data as one may imagine. Thus, every model constructed on such dataset could be tested under different condition and probably yielding satisfactory results each time. Unfortunately, additional data are seldom available due to additional cost, extra work done and so on. In order to omit the problem of lack of additional data one may use the crossvalidation technique.

The idea of crossvalidation is pretty simple, and rests on the division of the original dataset on two subsets; the first one is called a training set and the other is called a validation set. On the basis of the training set a model is built and one tries to forecast the data from the validation set. Among different types of crossvalidation models the Leave One Out CrossValidation (LOOCV) is probably the most often used. It relies on dropping out only one observation at a time from the entire dataset (n – times) and making a prediction of its value from the remaining n-1 observations (Efron and Tibshirani, 1993; Wackernagel, 2003). Thus, for every single observation from the data set one may determine the difference between the real observed value l_i and that provided by the model fitted to n-1 observations. In this way one obtains the crossvalidation error in the form:

$$e_{CV_i} = l_i - \hat{l}_i^{(n-1)} \qquad (14)$$

Marcin Ligas, Piotr Banasik

Averaging these differences over all observations one obtains the average crossvalidation error, given as:

$$\hat{e}_{CV} = \frac{\sum\limits_{i=1}^{n}\left(l_i - \hat{l}_i^{(n-1)}\right)}{n} \approx 0 \tag{15}$$

If its value is close to zero one may infer that prediction was made without an apparent bias. On the other hand, if there is a significant deviation from zero one may expect a systematic overestimation (or underestimation) introduced by the model (Wackernagel, 2003).

In order to compare the real prediction capabilities of different models the mean squared crossvalidation error (16) is constructed for every individual model.

$$\sigma_{CV} = \sqrt{\frac{\sum\limits_{i=1}^{n}\left(l_i - \hat{l}_i^{(n-1)}\right)^2}{n}} \tag{16}$$

The model with the least value of the mean squared crossvalidation error (16) will claim to be the best model among all tested but, of course, it cannot be considered in isolation from other quantities. The Table 6 presents the results of crossvalidation and least squares model fitting for bilinear and quadratic transformation functions.

Table 6. Results of crossvalidation and least squares fit for bilinear and quadratic surfaces

Case	Crossvalidation [m]		Fitted model [m]	
Bilinear	Minimum error	-0.020$_0$	Minimum residual	-0.006$_0$
	Maximum error	0.006$_8$	Maximum residual	0.006$_6$
	Averaged crossvalidation error	-0.000$_6$	Averaged residual	0 (from theory)
	Mean squared crossvalidation error	0.004$_2$	Residual standard deviation	0.002$_3$

Summary of crossvalidation errors [mm]							Summary of residuals [mm]						
[0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,8]	>8	[0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,8]	>8
35	36	25	21	12	13	8	38	52	39	10	7	4	0

Case	Crossvalidation [m]		Fitted model [m]	
Quadratic	Minimum error	-0.012$_2$	Minimum residual	-0.005$_5$
	Maximum error	0.005$_4$	Maximum residual	0.006$_5$
	Averaged crossvalidation error	-0.000$_6$	Averaged residual	0 (from theory)
	Mean squared crossvalidation error	0.003$_8$	Residual standard deviation	0.002$_0$

Summary of crossvalidation errors [mm]							Summary of residuals [mm]						
[0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,8]	>8	[0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,8]	>8
29	39	31	20	13	11	7	56	54	23	10	5	2	0

In every case (bilinear and quadratic) the crossvalidation statistics are slightly higher comparing to statistics derived from the fitted model. There is no apparent bias visible. The residual standard deviation from the fitted models for each case is almost half the mean squared crossvalidation error what may indicate that the overall fit described by the residual standard deviation may be somewhat too optimistic.

7. Confidence intervals on transformation function and on a single predicted value

In order to verify how reliable (at least from statistical standpoint) for prediction purposes the estimated quadratic trend surface is, a confidence interval was spread around it. It was to show places within the study area where we can expect the best and the worst accuracy. To do so, the method of construction of confidence intervals from (Krumbein, 1963; Krumbein and Graybill, 1965) is adopted. Thus, denoting the true trend surface as $L(x,y)$ and its estimated counterpart as $\hat{L}(x, y)$ we may symbolically write:

$$\hat{L}(x, y) - m_0 \sqrt{kF_\alpha \mathbf{x}^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{x}} \leq L(x, y) \leq \hat{L}(x, y) + m_0 \sqrt{kF_\alpha \mathbf{x}^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{x}} \quad (17)$$

where: k – total number of coefficients of the model, F_α - quantile of the F distribution with $(k, n\text{-}k)$ degrees of freedom and a significance level α, $\mathbf{x}$ – vector of coordinates (and their powers depending on the polynomial's degree) for a point for which the upper and lower bound is estimated, $(\mathbf{X}^T\mathbf{X})^{-1}$ – inverse of the normal equation matrix, m_0 – residual standard deviation.

This general formula is applied to as many points (a regular grid is preferred) as desired in order to produce a reasonable map of upper and lower bounds of confidence interval. But here, in order to save some space and to be more informative a contour map of half of the 0.95 confidence interval is presented (Fig. 3). As could be expected the most accurate fit of the polynomial surface is around the center of the study area providing the half-width of the 0.95 confidence interval on the level of 1 mm and decreasing out-of-center down to 3 mm.

As the main goal of the transformation function is height prediction for new points rather then just description of the phenomena itself it is reasonable to use mean square error of prediction (18) and confidence intervals for new predicted values (19) as statistical measures of prediction accuracy.

$$m_p = m_0 \sqrt{1 + \breve{\mathbf{x}}_i^T (\mathbf{X}^T\mathbf{X})^{-1} \breve{\mathbf{x}}_i} \quad (18)$$

$$\breve{l}_i \pm t_{\frac{\alpha}{2}, n-k} m_0 \sqrt{1 + \breve{\mathbf{x}}_i^T (\mathbf{X}^T\mathbf{X})^{-1} \breve{\mathbf{x}}_i} = \breve{l}_i \pm t_{\frac{\alpha}{2}, n-k} m_p \quad (19)$$

where: k – total number of coefficients of the model, $t_{\alpha/2}$ - quantile of the t distribution with $(n\text{-}k)$ degrees of freedom and a significance level α, $\breve{\mathbf{x}}$ – vector of coordinates (and their powers depending on the polynomial's degree) for a point for which the upper and lower bound is estimated, $(\mathbf{X}^T\mathbf{X})^{-1}$ – inverse of the normal equation matrix, m_0 – residual standard deviation.

These measures of prediction accuracy provided by the model could be confronted with the true errors for 8 points (little but always something) of the IInd class of the basic vertical network having their heights in both systems and clearly identified within the study area (Fig. 4, blue dots).
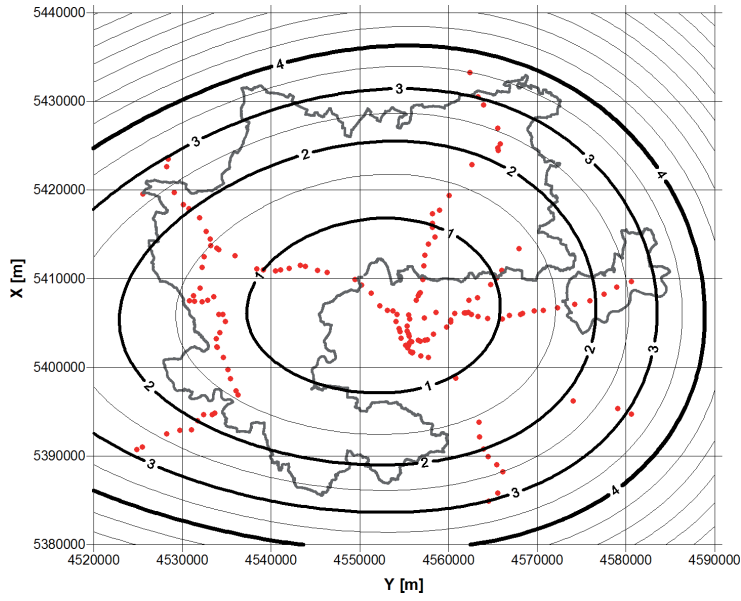
Fig. 3. Half of the confidence interval width for the quadratic transformation function (confidence level
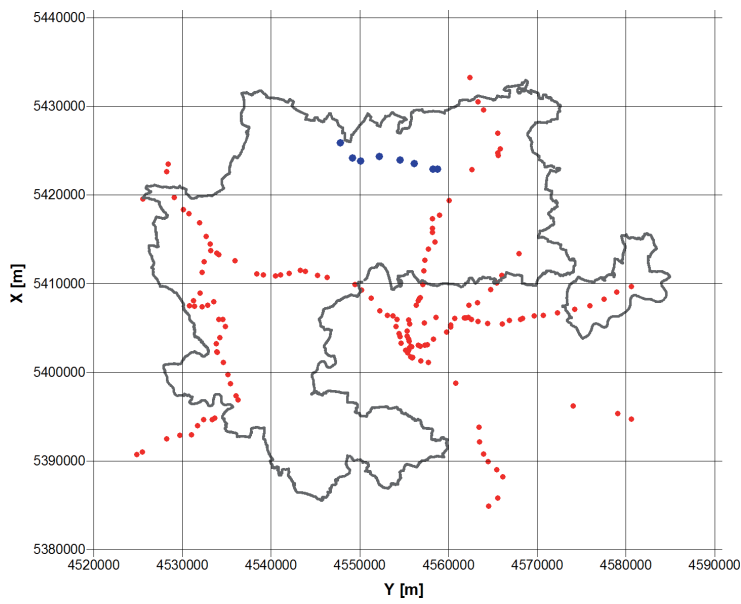1-α = 0.95, half confidence interval width given in [mm])

Fig. 4. Localization of 8 points of the IInd class of the basic vertical network used for the "true"
validation purposes (blue dots)

Table 7 presents the result of the true validation – comparison between the "true" prediction errors defined as "True dH_{86-60}" minus "Estimated dH_{86-60}" and those provided by the transformation model. As can be seen from the Table only one "true prediction error" is significantly larger then the prediction error from the model (point no 1 in the Table). Thus in some cases prediction errors provided by the model may be too optimistic. On the other hand, taking the 0.95 confidence interval around each predicted point a more realistic level of uncertainty is achieved ($\approx \pm 5$ mm) and this should be taken as a measure of accuracy of predicted (transformed) points.

Table 7. Comparison of the "true" prediction errors and those provided by the quadratic polynomial model on 8 test points of the IInd class of the primary vertical network

Point	True dH [m]	Estimated dH [m]	True – Estimated [m]	Prediction error [m]	$1-\alpha = 0.95$ half of the confidence interval width [m]
1	-0.0369	-0.0404	0.0035		
2	-0.0393	-0.0398	0.0005		
3	-0.0427	-0.0399	-0.0028		
4	-0.0381	-0.0406	0.0025	$\approx \pm 0.002$	$\approx \pm 0.005$
5	-0.0397	-0.0409	0.0012		
6	-0.0407	-0.0411	0.0004		
7	-0.0421	-0.0412	-0.0009		
8	-0.0427	-0.0414	-0.0013		

8. Impact of the horizontal position of control points on transformation function

In addition to the analysis of the polynomial transformation function presented above another very important issue had to be resolved. Namely, as stated in the "study area" section, horizontal position of control and transformed points was identified with the accuracy varying from several to tens of meters and in some individual cases reaching even 100 m. In order to verify whether this kind of identification "shift" has any impact on the results a simple simulation study has been performed. Having the "original" positions of control points derived through digitization process, new sets of control points have been obtained by random shifting the "original" position by 10, 50, 100, 500 m. in eight directions i.e.: N, NE, E, SE etc. preserving, of course, the height difference between two systems constant and assigned to proper point. After each "shifting" a new transformation function was obtained. This random shifting was performed 1000 times and every time the results from shifted data were confronted with the "original" transformation function. This was to give the answer to what extent the identification of horizontal position affects the final result. The simulation study showed that the dataset under study is not sensitive to the abovementioned shifts. For the largest shift of 500 m the maximum difference between height differences derived form the polynomial transformation function for the "original" and "shifted" datasets was on the level of

0.04 mm. This value implies that even the exaggerated shift of 500 m has no practical meaning in the analysis and its influence on the results may be completely ignored. But this cannot be said in general, this value is certainly dataset – dependent thus this kind of simulation study (whenever this type of uncertainty is present) should be performed for each dataset separately.

9. Conclusions

In the paper, a fairly detailed analysis of the local transformation between two height systems – Kronstadt'60 and Kronstadt'86 – has been presented. In order to achieve some level of certainty as to the correctness (accuracy, usability) of the local transformation performed some redundancy was introduced to fulfill the saying "always make sure and double check". On the basis of some statistical tests (analysis of variance), techniques (crossvalidation) and measures (Akaike Information Criterion) the quadratic polynomial transformation function was chosen as well describing the dataset of control points of the Ith class of the major vertical network as well as giving satisfactory results for transforming points of the minor vertical network. In order to assess the level of uncertainty associated with height prediction for points of the minor vertical network the prediction error offered by the polynomial regression was used. The magnitude of the prediction error on the level of ± 2 mm turned out to be slightly too optimistic in comparison to the "true prediction errors" obtained on 8 points of the IInd class of the basic vertical network. Hence, as a more realistic measure of prediction accuracy the 0.95 confidence interval for predicted height differences was constructed giving the value of prediction uncertainty on the level of ± 5 mm. Even the latter value could slightly be increased by using the mean squared crossvalidated error instead of the residual standard deviation, then giving the prediction uncertainty on transformed points on the level of ± 8 mm, and this value is so careful that probably only blunders will exceed this limit.

Acknowledgments

References

Davies, J.C. (2003). *Statistics and data analysis in geology*. New York: John Wiley & Sons.
Efron, B. & Tibshirani, R.J. (1993). *An introduction to the Bootstrap*. New York: Chapman&Hall/CRC.
Gucek, M. & Basic, T. (2009). *Height transformation models from ellipsoidal into the normal orthometric height system for the territory of the city of Zagreb*. Stud. Geophys. Geod., 53, 17-38.

GUGiK. (2000). *Regulation of Council of Ministers of 8 August 2000 on the State Spatial Reference System.* (In Polish). Dz.U. Nr 70, poz.821.

Hadaś, T. & Bosy, J. (2009). *Niwelacja satelitarna GNSS z wykorzystaniem serwisu NAWGEO systemu ASG – EUPOS.* Acta Sci. Pol. Geod. Descr. Terr., 8(2), 53-66.

Katalog. (1960). *Catalogues of Leveling Lines M-34-64 (Kraków), M-34-65 (Nowa Huta), M-34-76 (Myślenice) i M-34-77 (Bochnia).* (In Polish). Warszawa: GUGiK MSW.

Krumbein, W.C. (1963). Confidence intervals on low – order polynomial trend surfaces. *Journal of Geophysical Research*, 68(20), 5869-5878.

Krumbein, W.C. & Graybill, F.A. (1965). *An introduction to statistical models in geology*. New York: McGraw–Hill, Inc.

Osada, E. (2008). *Analysis of base map transformation methods to "2000" and Kransztag 1986 coordinate systems, the Wroclaw example- the best fit.* (In Polish). Magazyn Geoinformacyjny „Geodeta" 10(161) 44-46.

Rao, C.R. & Toutenburg, H. (1999). *Linear models: least squares and alternatives*. New York: Springer.

Rawlings, J.O. Pantula, S.G. & Dickey, D.A. (1998). *Applied regression analysis: a research tool*. New York: Springer.

Ripley, D.B. (1981). *Spatial Statistics*. New Jersey: John Wiley & Sons.

Schabenberger, O. & Gotway, C.A. (2005). *Statistical methods for spatial data analysis*. Boca Raton: Chapman&Hall/CRC.

Shacham, M. & Brauner, N. (1997). Minimizing the effect of collinearity in polynomial regression. *Ind. Eng. Chem.* Res., 36, 4405-4412.

Wackernagel, H. (2003). *Multivariate Geostatistics*, Berlin: Springer.

Lokalna transformacja wysokości przy uyciu regresji wielomianowej

Marcin Ligas, Piotr Banasik

AGH Akademia Górniczo-Hutnicza,
Wydział Geodezji Górniczej i Inżynierii Środowiska
Katedra Geomatyki, al. A. Mickiewicza 30, 30-059 Kraków
e-mail: ligas@agh.edu.pl, pbanasik@agh.edu.pl

Streszczenie

W artykule przedstawiono wyniki transformacji wysokości między układami Kronsztadt'60 i Kronsztadt'86 na obszarze powiatu krakowskiego. Ostatni z wymienionych układów jest obecnie częścią obowiązującego w Polsce Państwowego Systemu Odniesień Przestrzennych. Transformacja między wymienionymi układami wysokości została wykonana w oparciu o dobrze znaną i często stosowaną w geodezji regresję wielomianową. Mimo jej powszechności w zastosowaniach rzadziej można spotkać w literaturze jej szerszą analizę pod względem optymalnego stopnia wielomianu, jakości dopasowania oraz zdolności predykcyjnych. W niniejszym opracowaniu wykorzystano różne metody w celu uzyskania statystycznej pewności co do poprawności i praktycznej użyteczności opisywanego modelu.