Two covariance models in Least Squares Collocation (LSC) tested in interpolation of local topography

Wojciech JARMOŁOWSKI[1], Mieczysław BAKUŁA[1]

[1] Faculty of Geodesy and Land Management
University of Warmia and Mazury in Olsztyn, ul. Heweliusza 5, 10-724 Olsztyn, Poland
e-mail: wojciech.jarmolowski@uwm.edu.pl, mbakula@uwm.edu.pl

Abstract: Advantages and disadvantages of least squares collocation (LSC) and kriging have recently been discussed, especially as interdisciplinary research becomes popular. These statistical methods, based on a least squares rule, have infinite number of applications, also in the domains different than Earth sciences. The paper investigates covariance parameters estimation for spatial LSC interpolation, via a kind of cross-validation, called hold-out (HO) validation. Two covariance models are applied in order to reveal also those differences that come solely from the covariance model.

Typical covariance models have a few variable parameters, the selection of which requires analysis of the actual data distribution. Properly chosen covariance parameters result in accurate and reliable predictions. The correlation length (CL), also known as the correlation distance in the Gauss-Markov covariance functions, the variance (C_0) and a priori noise parameter (N) are analyzed in this paper, using local terrain elevations. The covariance matrix is used in LSC, as analogy to the correlation matrix often present in the kriging-related investigations. Therefore the covariance parameter N has the same scale as the data and can be analyzed in relation to the data errors, spatial data resolution and prediction errors.

The vector of the optimal three covariance parameters is sometimes determined approximately for the purposes of modeling with limited accuracy requirements. This is done e.g. by the fitting of analytical model to the empirical covariance values. The more demanding predictions need precise estimation of the covariance parameters vector and the researchers solve this problem via least squares methods or maximum likelihood (ML) inference. Nevertheless, both least squares and ML produce an error of the parameters and it is often large. The reliability of LSC or kriging using parameters with an error of e.g. a quarter of the parameter value is usually not discussed. This paper involves a kind of cross-validation, performed to observe possible influence of the parameters error on the prediction accuracy. This kind of validation serves for a basis of considerations on the accuracy of covariance parameters estimation with other different techniques.

Key words: covariance function, correlation length, hold-out validation, Least Squares Collocation, noise

doi: 10.2478/congeo-2013-0001

1. Introduction

Many engineering and scientific tasks require precise estimates of spatial correlation and accurate interpolation, especially when high precision and reliability are desired in the case of sparse or noisy data. To ensure accuracy, it is necessary to determine the optimal parameters for the modeling process and to incorporate the most precise data, free of systematic errors. Digital elevation models (DEMs) have many details of limited correlation, e.g. drainage structure or the structures created by human activity, which are hard to process via least squares collocation (LSC). Uncorrelated details are not investigated in this paper. The correlated part of the terrain provides the data for the current research and its covariance parameters are analyzed. This correlated part, predicted with optimal parameters may submit the base for mentioned details, especially if it's unbiased.

The simplest accuracy measure of any spatial or space-time prediction is usually the RMS of difference between the predicted value and the true data. The RMS may be obtained by comparing digital elevation model (DEM) surface of any origin to the corresponding elevation values obtained from high-accuracy survey techniques such as total station or real-time kinematic (RTK) survey. A second important estimate is bias, i.e., the systematic error usually obtained as the mean of the differences between the tested data and the control points. This is the vertical shift between the true surface and the actual model, and it is usually a result of errors in data processing or imperfect reference level implementation. The bias often occurs in DEM models created with different techniques (*Gonçalves-Seco et al., 2006; Veneziano et al., 2004*). Today, various methods of creating DEMs can result in an accuracy of a few decimeters, but they can suffer from complicating factors, like those caused by forests or large terrain slopes (*Veneziano et al., 2004*). The random error rapidly increases in difficult observational conditions, making estimation without control points less reliable.

A wide spectrum of techniques is available in geostatistics, even in the form of the commercial software. The properties, advantages and disadvantages of some tools are described in the literature and in the actual context of terrain modeling (*Bater and Coops, 2009; Erdogan, 2010; Wise, 2007*). Different modeling techniques are useful for individual purposes because of their specific mathematical behavior, but least squares collocation (LSC) and kriging methods, judging from the number of their applications, appear to be commonly applied (*Darbeheshti and Featherstone, 2010; Hengl et al., 2008; Kotsakis, 2007; Pringle et al., 2009*). The article investigates the spatial prediction using LSC with the covariance models based on an exponential function and called Gauss-Markov models or sometimes Markov models.

The exponential models of the covariance have nowadays a large number of applications in the spatial data modeling e.g. atmospheric density distribution (*Eshagh, 2009*), altimetry tracks processing (*Andersen and Knudsen, 1998*) or geoid modeling (*Kavzoglu and Saka, 2005*). Moreover, exponential models are present not only in the spatial modeling context, but time interpolation examples may be also found (*Revallo et al., 2010*). Therefore LSC and kriging of the various data types with various covariance models may be recognized as a wide, interdisciplinary problem. The special research problem related closely to these methods is the estimation of covariance function parameters that represent the best local spatial covariance of the investigated field.

The authors do not attempt to compare terms or to assess differences between LSC, known from geodetic investigations (*Kotsakis, 2007; Moritz, 1980*) and kriging method known from different domains applying geostatistics (*Kitanidis, 1983; Pringle et al., 2009*). A specific comparison of kriging and LSC within the context of the local geoid determination may be found in *Reguzzoni et al. (2005)*. Some statistical assumptions on the covariance models differ among these methods, but the general rule of data weighting is similar. It should also be mentioned that similar methods are simply called linear predictors in some papers dealing with terrain modeling problems (*Briese et al., 2002; Lohmann et al., 2000*). We use here the LSC terminology, as the authors are more familiar with the geodetic literature.

Least squares techniques for interpolation or adjustment processes are widely used in many geodetic applications and in other fields of Earth sciences. The aim of this paper is not to present a detailed DEM creation, but analyze the process of local terrain modeling using LSC.

A set of 2458 points from a GNSS/RTK survey is used to best fit the parameters of Gauss-Markov second-order (GM2) and third-order (GM3) covariance models to empirical covariance function. The dataset used in this study was assumed to have no specific long-wave trend to be subtracted for the LSC process because no global model has been used and we avoided use of polynomials. After a general look at residuals variations around the mean we decided to treat them as an approximately residual field. The aim of this research is not to test the precision of LSC modeling, because the dataset used is of limited density and is irregular. The tested factor is the influence of varying parameters in a covariance model on the efficiency of the prediction process. The question of the relation of the empirical covariance function with the analytical model is partially answered here. In particular, the efficiency of two covariance models is thoroughly investigated here with some insights regarding optimum correlation length (CL) and noise (N) values. The estimation of optimal CL and N parameters is an essential problem in LSC. Assuming the spatial correlation of the field, it is necessary to assess these parameters to retain the optimum accuracy.

The results of the fitting of covariance parameters may, of course, be applied to all DEM data acquisition techniques, other than RTK. However, it should be kept in mind that differences in terrain roughness and differences in the type of data require specific covariance function parameters for modeling, or even specific modeling functions every time. Therefore every dataset may need different covariance parameters and their precise estimation has to be made. Determination of the variance (C_0) and CL parameters by practitioners is often performed in an approximate way, by fitting the functional model to the empirical covariance function or empirical semivariogram. Determination of the N value is especially difficult using this technique. Moreover, inaccurate determination of N strongly affects other parameters, i.e. C_0 and CL, as well as the prediction accuracy.

Numerous tests on the covariance parameters estimation may be found in the literature, involving e.g. various applications of Least Squares (LS) rule or Maximum Likelihood (ML) estimation (*Kitanidis, 1983; Pardo-Igúzquiza et al., 2009*). The estimates of the covariance function parameters are always obtained with an estimation error. This error is often significantly large in comparison to the computed parameter and its influence on LSC or kriging is not presented in many cases. Therefore this paper focuses on the general behavior of parameters and may be an attractive background for comparisons with different techniques of the covariance parameters estimation. The results of this study, involving a kind of cross-validation, shows what ranges of the parameters choice are especially erroneous in the case of approximate parameters determination or significant errors of the parameters estimation. The differences in the functioning of two covariance models are also shown with some conclusions.

2. Datasets description

A fully operational system of reference GNSS stations in Poland, which covers the entire country, was put into operation in the middle of 2008 (*Bosy et al., 2007*), although the testing network has been operational since 2004. Commonly, the ASG-EUPOS system, shown partially in Fig. 1, offers several real-time services for RTK and for differential positioning. There are also two services for the post-processing of collected data. The development of ASG-EUPOS services was a milestone in national surveying, and made GNSS a widely popular technique with the potential to replace classical ones to some extent. There is a strong factor limiting the use of real-time, relative GNSS positioning in general, i.e., sky obstructions (*Bakuła et al., 2009*). It can even make a survey impossible. The correlation-based modeling with weighting considered in this paper may help in minimizing the effects of a

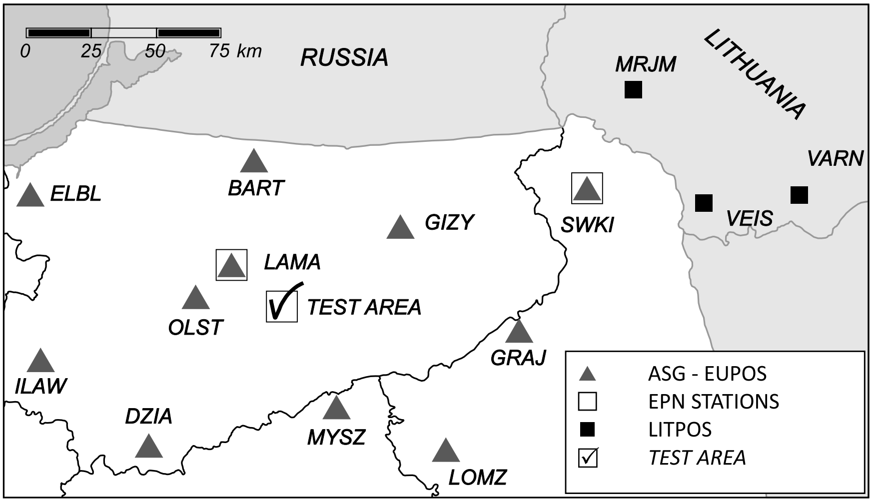

Fig. 1. Deployment of reference stations in northern Poland near test area (ASG-EUPOS).

noisy data. A specific test with a large set of data is performed to find the most accurate 3D model of the surface, assuming its spatial correlation.

Two sets of data were collected: 2077 points of the sparse field survey and 472 points surveyed at parcels' boundary marks. Both datasets cover approximately the same area. First set is the prediction set and the second is the control dataset. First dataset is used for the prediction on the positions of the control one. Only the data with vertical accuracy estimates smaller than 0.08 m were used for the interpolation. The number of the observations used in the prediction set was therefore reduced to 1986 by removing points with poor vertical accuracy estimates. There were no points with significant measurement errors in the control dataset. Figure 2 shows the prediction (crosses) and the control (circles) datasets, both coming from the independent surveys.

Two datasets of a few centimeters accuracy were used in LSC, but only prediction dataset provides surveyed heights for an interpolation. The point validation was performed for the optimal CL and N estimation. This kind of validation is often called hold-out (HO) validation (*Arlot and Celisse, 2010; Kohavi, 1995*), as far as the data are split only once into separate prediction and control subsets. The prediction set of data was a sparse RTK survey, and it was assumed to be a database for the interpolation process. Figure 3 presents the histograms of the observed RMS for the position and height, excluding outliers.

The histograms show the observational noise. The use of an a priori noise variance is a key factor in LSC prediction. However, some data acquisition techniques do not have an internal method for a priori error estimation, and they need certain control values from the other techniques, as mentioned in *Hengl et al. (2008).* It is usually assumed that the error is not correlated, so only diagonal error variances are present. This affects the spatial correlation in a positive sense, i.e., the smooth surface is not unnecessarily fitted to the noisy data.

The control points come from the survey of existing parcels' boundary marks over the same area. The points were sufficiently dense to perform HO validation, i.e., to interpolate height values based on one dataset at the positions of the second one. Therefore, two independent datasets were used to examine modeling process in detail.

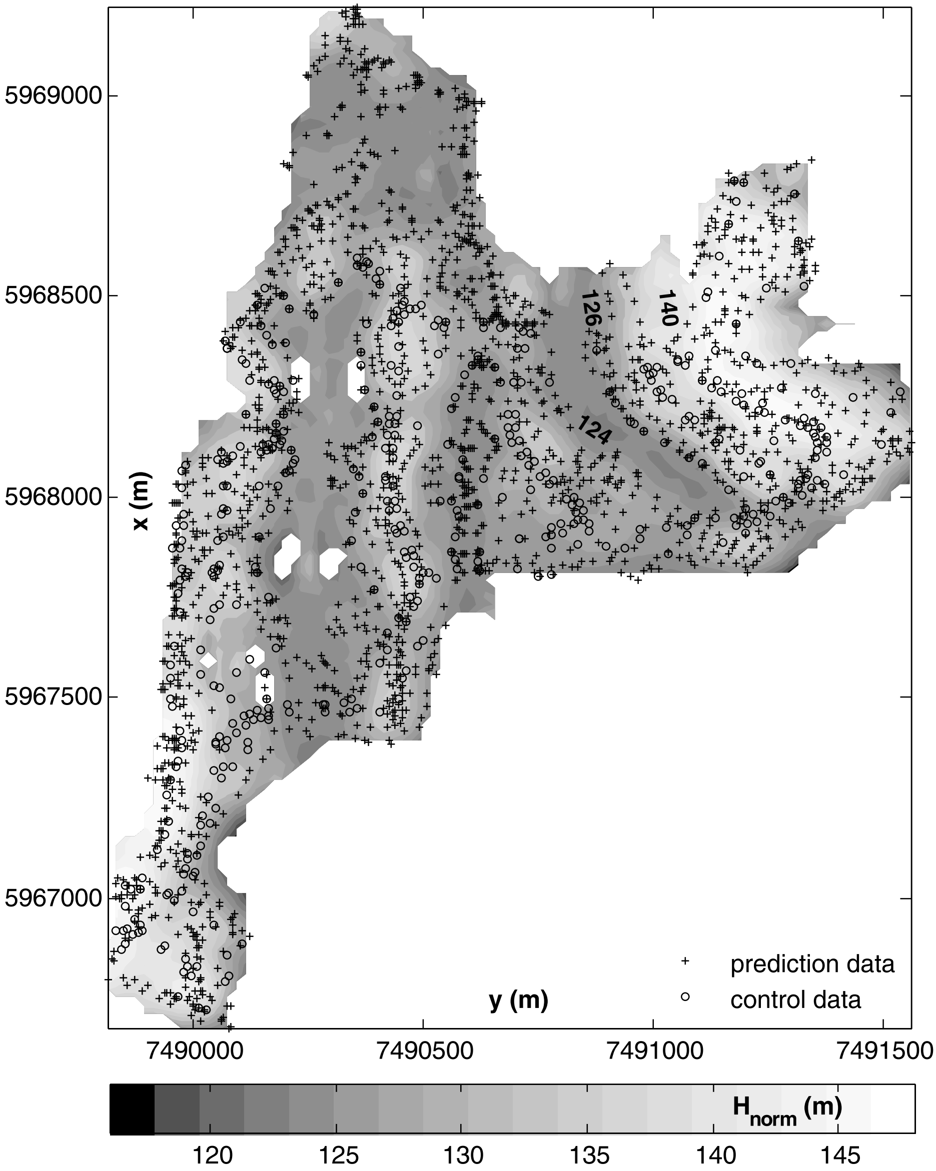

Fig. 2. Data plot and contour map. Crosses are prediction data and circles – control points.

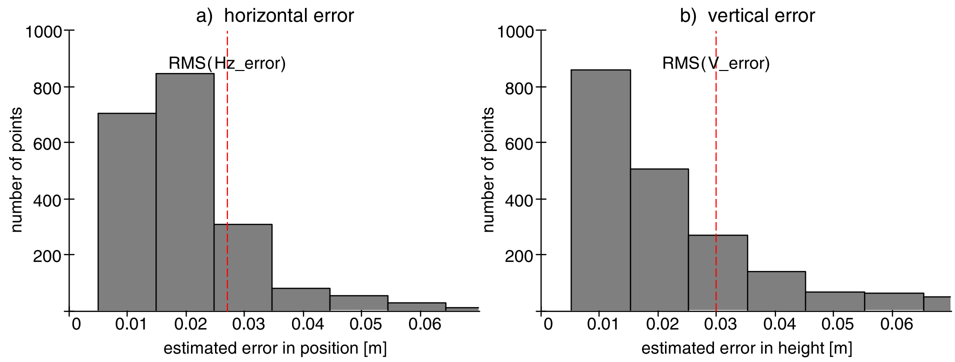

Fig. 3. Histograms of: (a) horizontal and (b) vertical observational errors (prediction dataset).

3. LSC prediction model

The GM2 and GM3 models were used as the covariance function for the interpolation using LSC method. The second order model is (*Andersen and Knudsen, 1998, p. 8130*):

$$\mathrm{C2}(\mathrm{C}_0, \mathrm{CL}, \mathrm{s}) = \mathrm{C}_0 \left(1 + \frac{\mathrm{s}}{\mathrm{CL}}\right) \cdot \mathrm{e}^{-\left(\frac{\mathrm{s}}{\mathrm{CL}}\right)}, \tag{1}$$

where s is the distance in meters between the stochastic field points, C_0 is the field variance in square meters and CL is a term in meters that determines the most appropriate shape of covariance function of the local field and that approximates the maximum distance for the data point values being correlated with each other. The same terms define the third-order model (*Kavzoglu and Saka, 2005, p. 523*), as follows:

$$\mathrm{C3}(\mathrm{C}_0, \mathrm{CL}, \mathrm{s}) = \mathrm{C}_0 \left(1 + \frac{\mathrm{s}}{\mathrm{CL}} + \frac{\mathrm{s}^2}{3 \cdot \mathrm{CL}^2}\right) \cdot \mathrm{e}^{-\left(\frac{\mathrm{s}}{\mathrm{CL}}\right)}. \tag{2}$$

These exponential covariance models are here adopted from geophysical and geodetic literature and are widely known in the interpolation of various functionals of the Earth's disturbing potential, e.g. geoid heights or gravity anomalies. Moreover, we found some similar models applied to the terrain modeling (*Briese et al., 2002; Lohmann et al., 2000*) and decided to test also GM2 and GM3 models with terrain data.

The analytical model should be well fitted to the empirical covariance function. The local variability of the actual topography may be observed in Fig. 4. However, the topography in different places and also different physical fields may represent different variance or smoothness. Only the proper choice of parameters produces the most accurate results when using LSC as an interpolation tool. The covariance model of the spatial field may be used in the following HO validation based on LSC formula (*Moritz, 1980; Hofmann-Wellenhof and Moritz, 2005*):

$$\mathrm{H}_{\mathrm{P}}^{\mathrm{res}} = \mathbf{C}_{\mathrm{P}}^{\mathrm{T}} \cdot (\mathbf{C} + \mathbf{N})^{-1} \cdot \mathbf{H}^{\mathrm{res}}. \tag{3}$$

$\mathrm{H}_{\mathrm{P}}^{\mathrm{res}}$ is the predicted value at the position of the control point from the control dataset. The matrix $\mathbf{C}$ is the covariance matrix of the data, $\mathbf{C}_{\mathrm{P}}$ is the covariance vector between the predicted point and the data, $\mathbf{N}$ is the noise covariance matrix and $\mathbf{H}^{\mathrm{res}}$ is the data residuals vector. The vector of residuals should have its expected value zero and to achieve this, the long term trend may be subtracted. We use simple data mean, but most advanced techniques may be applied to fulfill the condition more precisely,

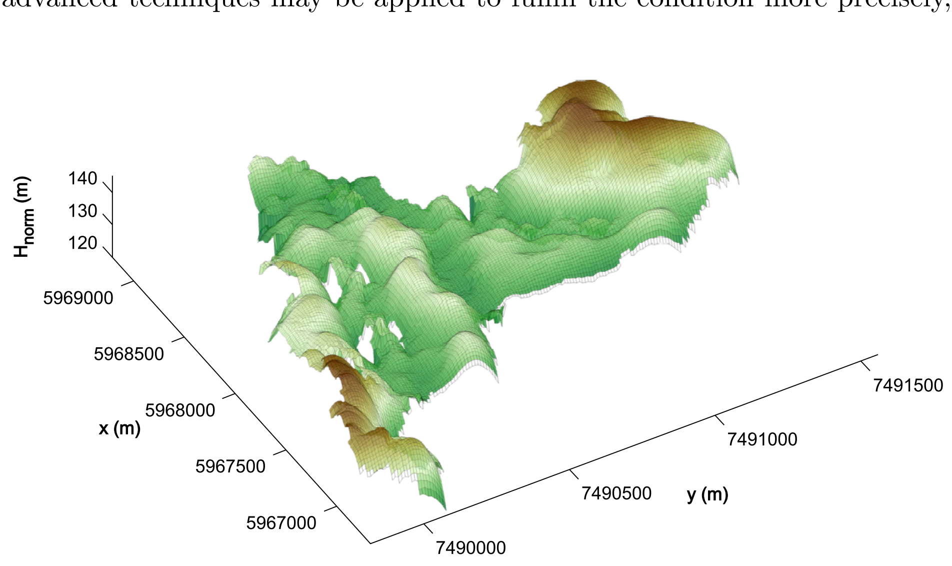

Fig. 4. Sample elevation model from LSC prediction of RTK data.

e.g. polynomial trend. After the prediction process (Eqs. 3 and 6), the trend part of the surface should be restored in order to obtain the final height. In some cases, where the noise is white, the matrix **N** is diagonal and it is often called diagonal loading, also known as regularization (*Lim and Mulgrew, 2007*). In such cases, we have the following relation:

$$\mathbf{N}_{\mathrm{ij}} = \begin{cases} \mathrm{N}_{\mathrm{i}}^2 & \mathrm{i} = \mathrm{j} \\ 0 & \mathrm{i} \neq \mathrm{j} \end{cases} \tag{4}$$

The indices in Eqs. 4 and 7 relate to sparse points used in the interpolation process. In the case of an RTK survey, the estimated vertical RMS coming from GNSS data adjustment may not be the only factor determining N_{ij}. The limited resolution may significantly affect the error covariance, when the spatial data distribution is sparse and the resolution of predicted field doesn't reflect data accuracy.

The notation in Eqs. 1, 2 and 3 is conventional in geodesy, but if we include N in the vector of the covariance function parameters $\boldsymbol{\theta}$, the kernel from Eq. 1 will be

$$\mathrm{C2}(\boldsymbol{\theta}, \mathrm{s}) = \mathrm{N}^2 + \mathrm{C}_0 \left(1 + \frac{\mathrm{s}}{\mathrm{CL}}\right) \cdot \mathrm{e}^{-\left(\frac{\mathrm{s}}{\mathrm{CL}}\right)} . \tag{5}$$

Emphasizing that N parameter is zero for non-diagonal elements of the data covariance matrix and for the whole covariance vector between prediction and data, we may rewrite the LSC formula to

$$\mathrm{H}_{\mathrm{P}}^{\mathrm{res}} = \mathbf{C}\left(\boldsymbol{\theta}\right)_{\mathrm{P}}^{\mathrm{T}} \cdot \mathbf{C}\left(\boldsymbol{\theta}\right)^{-1} \cdot \mathbf{H}^{\mathrm{res}} . \tag{6}$$

The error covariance represented by nugget inside the interpolation kernel is more frequent in case of kriging variograms or covariance functions, but it is also reasonable here in LSC, because we include N in the vector of parameters $\boldsymbol{\theta}$, which will be estimated.

4. Empirical covariance function

To determine a CL, which is equivalent to range in kriging, the empirical covariance function is analyzed (*Crombaghs et al., 2002; Hofmann-Wellenhof and Moritz, 2005*). The empirical covariance function comes from a specific statistical evaluation of the data following the rule:

$$\forall(i,j) \ \ | \ \ \mathrm{s} = \left[(x_i - x_j)^2 + (y_i - y_j)^2\right]^{\frac{1}{2}} \ : \ \mathrm{EC(s)} = \frac{\sum_{i,j}^{\mathrm{n}} \mathrm{H}_i^{\mathrm{res}}\,\mathrm{H}_j^{\mathrm{res}}}{\mathrm{n}} \,. \qquad (7)$$

The data $\mathrm{H}^{\mathrm{res}}$ has to be residual after the trend surface or mean value subtracted, i.e., when the expected value is close to zero. The obtained estimate of covariance is the mean product of n data values when the distance between points is equal to s. In our case, s is an interval that comes from the intervals in distance repeated several times to cover the entire area of the data. The values of EC computed for increasing distances results in the so-called empirical covariance function, which shows the behavior of local fields and may be fundamental in the selection of the best covariance model for the interpolation process. Prediction dataset was used to estimate the empirical covariance parameters for the distances increasing up to 400 m. The empirical covariance function estimates began to oscillate from 400 m, where the function decreased to zero. This result indicates that no correlation should be expected between data points separated by distances greater than 400 m (Fig. 5).

The empirical covariance function has a specific value at distance zero, i.e., when the point heights are multiplied by themselves. A mean product

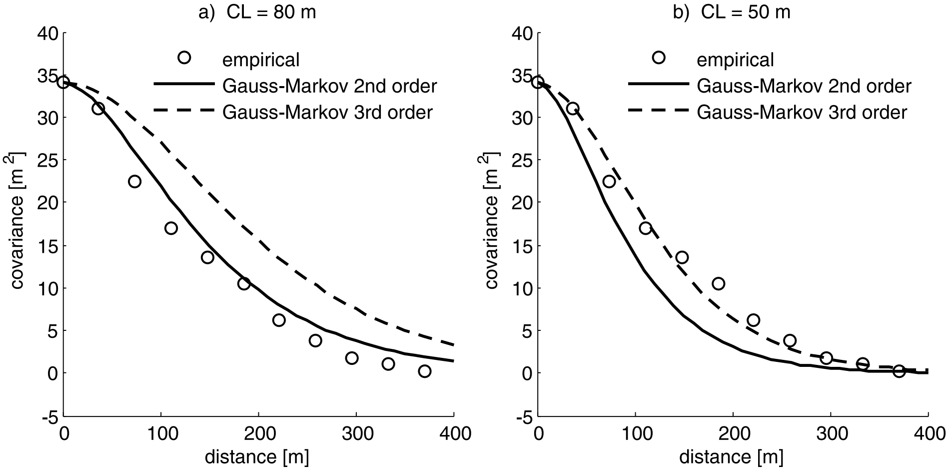

Fig. 5. Empirical covariance function estimated from data (circles) with GM2 (solid) and GM3 (dashed) models when CL is equal: (a) 80 m and (b) 50 m.

of about 35 m^2 represents the general variance of the data and is an estimation of the scale factor for the analytical model (Fig. 5), i.e. C_0 parameter in Eqs. 1–2 and 5. The assumption that half of the variance indicates CL on the distance axis may be an example of CL approximation for some covariance models (*Moritz, 1980*). This assumes a CL of approximately 120 m in the case of a second-order function. However, fitting of the shape of the analytical function requires the CL parameter to be equal to about 70–80 m for the GM2 model (Fig. 5).

The practical choice of an optimal CL parameter has been discussed numerous times and has been estimated in relation to the empirical covariance parameters. The optimum is sometimes assumed to be somewhere between zero value of the empirical covariance function and the distance indicating half of the data variance at the covariance axis (*Crombaghs et al., 2002; Lohmann et al., 2000*). This paper presents an investigation of CL and N with a detailed look at the loss of accuracy when LSC prediction is performed with approximate covariance parameters or the parameters' estimates have significant errors.

5. Results of prediction with varying parameters

As mentioned in sections 2 and 3, the data comprised of two sets. The prediction dataset was the basis for the interpolation process with different covariance model's parameters. The control set provided control values for the HO test of the surface fit after the interpolation involving 30 closest points from prediction set. No covariance function fit was performed in the experiment. The solution was checked for a wide spectrum of probable CL and N parameters to search for the optimum parameters of the covariance model empirically. An increasing CL was used in consecutive predictions, starting at 40 m and ending at 160 m for GM2 and from 20 m to 140 m for GM3. The N parameter was set as uniform for all points in the prediction set and no correlation was assumed between the data points. The approximate C_0 value observed from an empirical covariance function is 35 m^2. One smaller and two larger values are also tested in order to assess general influence of this parameter on the prediction and the other parameters (Figs. 6 and 7).

The statistical quantities computed for comparisons show the level of fitness between the original values and the interpolated ones. The smallest RMS available from the HO validation is around 0.58 m. This value represents accumulated error of RTK accuracy, data distribution and the precision of the interpolation technique. The prediction accuracy is here strongly affected by spatial data distribution and it is normal that the accuracy decreases when data is distributed too sparse in relation to the spectrum represented by its measurement error. The estimated error of LSC depends strongly on the N parameters chosen for the LSC process. There was also no maximum threshold for the closest point taken for the prediction, therefore some predictions may find closest data in a few tens of meters distance. The standard measure of the prediction accuracy for the whole dataset may be calculated as $\varepsilon^2 = (\mathbf{V}^{\mathrm{T}}\,\mathbf{V})/\mathrm{n}$, where vector $\mathbf{V}$ consists of point prediction errors computed via known formula from *Hofmann-Wellenhof and Moritz (2005, p. 362)*. Applying such assumptions, for nearly optimal parameters, the mean square error for GM2 is 0.69 m ($\mathrm{C_0} = 35\ \mathrm{m}^2$, CL = 80 m, N = 0.3 m) and for GM3 is 0.50 m ($\mathrm{C_0} = 35\ \mathrm{m}^2$, CL = 50 m, N = 0.3 m). Biases between predicted values (residuals) and the control dataset are equal respectively to 0.16 m and 0.12 m and may indicate that the mean is not the best trend here. Nevertheless, the observed bias is of much smaller order than the prediction errors, therefore we recognize the covariance parameters estimation as valuable.

Two models of the analytical covariance function were tested to compare the behavior of their results when changing the CL parameter. First, the GM2 model was used as the covariance model. Figure 6 presents RMS of the difference between predicted values of residuals and the control residuals, where the CL is variable from 40 to 160 m. The diagonal loading in terms of an a priori noise N is also variable from 0 m to 0.9 m.

The RMS in Fig. 6 shows that the accuracy of LSC rapidly decreases when using a CL shorter than about 60 m. The differences start to increase also when the CL is longer than 120 m, depending on the parameter N. Decreasing N leads also to decrease of the accuracy if CL and $\mathrm{C_0}$ are fixed. Therefore, optimal results require different $\mathrm{C_0}$ parameters for different diagonal input, i.e. the level of the field smoothness. On the other hand, if $\mathrm{C_0}$ is fixed, one pair of CL and N arguments gives optimum interpolation results. Both covariance models provide similar minimum RMS at the level

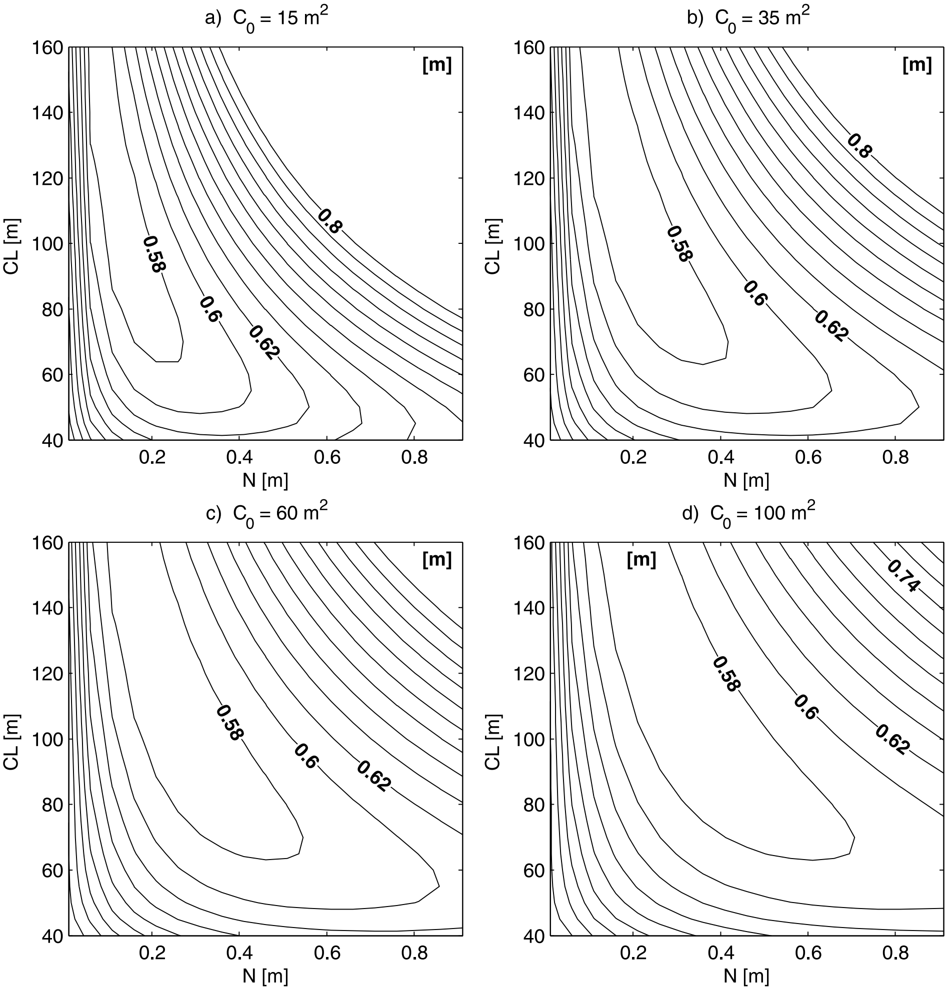

Fig. 6. RMS of differences between control values and LSC prediction based on the GM2 function for varying CL and N.

of 0.58 m, but in case of GM3 it achieves the value 0.56 m. Moreover, different CL and N parameters achieve the same level of modeling precision when properly combined. These parameters play a key role if we want to fit the analytical model to the empirical covariance values. It is also evident, that different covariance functions may require different values of CL (Figs. 6

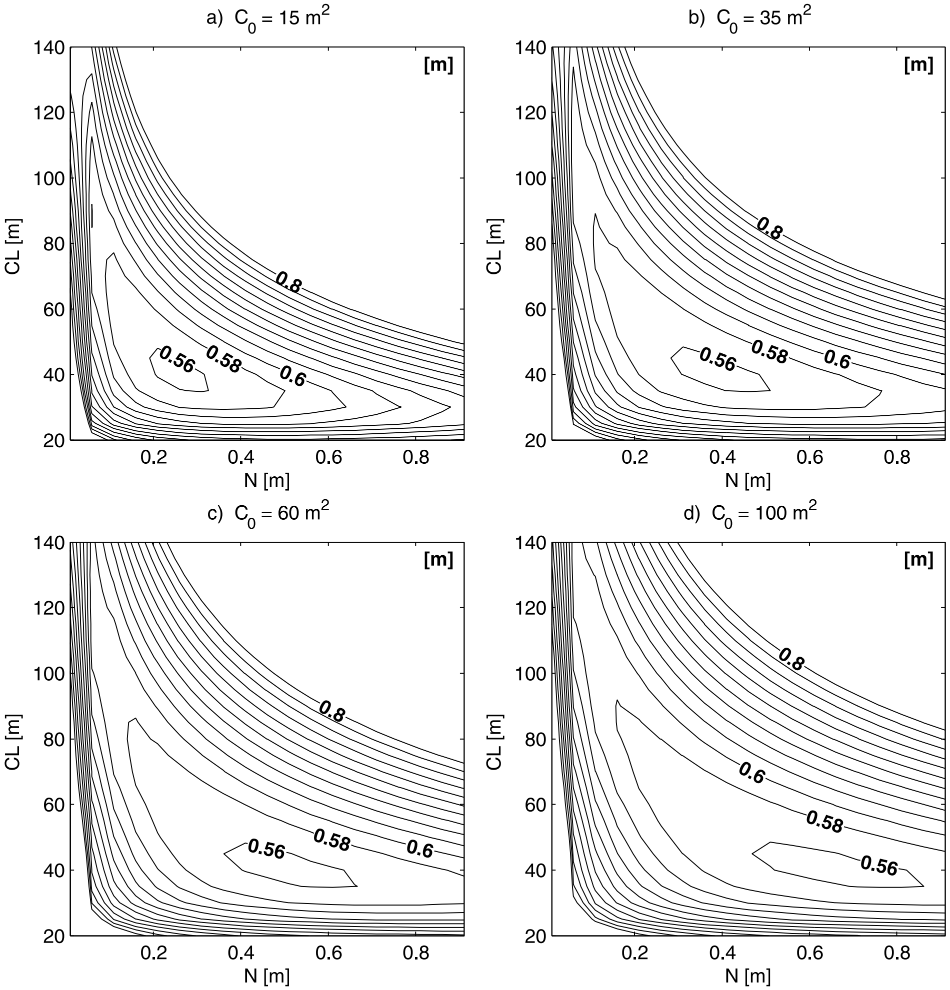

Fig. 7. RMS of differences between control values and LSC prediction based on the GM3 function for varying CL and N.

and 7).

Figures 6 and 7 present the statistics of the comparisons between the original and interpolated heights when the a priori noise N is varying. The estimate of N for an optimum RMS of differences between predicted and measured values, when C_0 value is similar to the residuals variance (Figs. 6b and 7b), is larger than RMS of GNSS observations. Figures 6 and 7 present two cases of N applied in the modeling with the two employed functions. It is evident here that all RMS minima move along the N axis due to the change of the C_0 parameter and are bigger than the surveying error (0.08 m is VRMS threshold in investigated datasets). This may indicate certain constancy of the CL parameter, especially if C_0 becomes larger than 35 m^2 (Fig. 8).

Furthermore, the optimal values of CL move away from those estimated with the empirical covariance function in the case of rough predictions, where small N is assumed. The results in Fig. 7 indicate that smooth prediction (e.g. for N = 0.3 m) with the GM3 model requires particularly precise estimation of the CL to avoid large modeling errors (Fig. 7b).

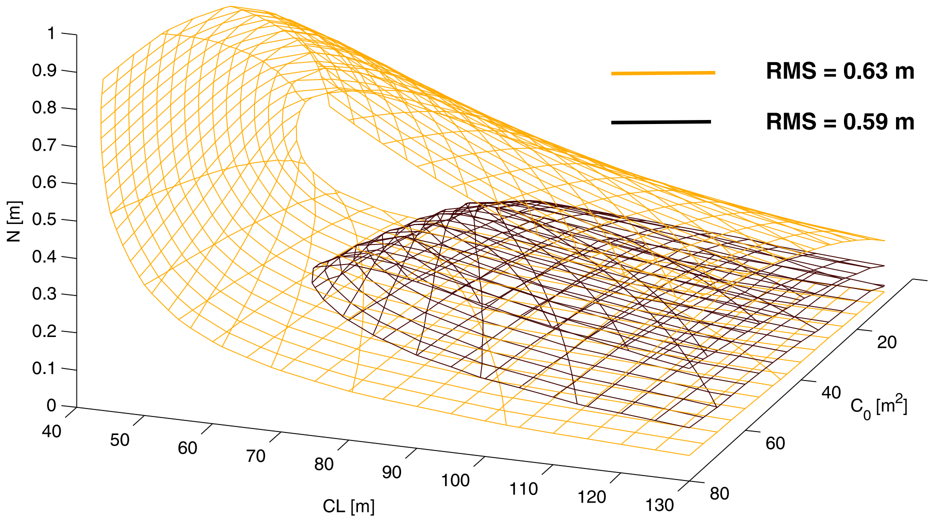

Fig. 8. RMS of differences between control values and LSC prediction based on the GM2 function drawn in space of three parameters of the covariance function.

6. Discussion and conclusions

Terrain data are significantly correlated what may be presumed from the empirical covariance function (Fig. 5). There was no long-wave trend information used besides the data mean and the mean was only subtracted. Therefore C_0 parameter based on the empirical covariance function is approximately equal 35 m^2 and it's hard to improve this estimate, because C_0 is correlated with N (Figs. 6 and 7). The results show strong dependency between C_0 and N parameters in case of both functional models. Therefore the most general conclusion is that terrain may require more advanced covariance parameters estimation than fitting of the covariance function only.

The optimum CL varies depending on the analytical model chosen and on the a priori noise parameter N responsible for the surface smoothness. It is observable for the GM2 model (Fig. 6) that the prediction tailored extremely to the used data, when N value is too small, may require a larger CL, to supply additional smoothness. In case of the GM3 (Fig. 7), the minimum is more precisely situated, what may need more precise estimation of CL and N. Moreover, we find the RMS below 0.56 m for GM3, which is not achievable in case of GM2. From Fig. 6 we may observe that deficiency of the third order term in GM2 makes the range of useful parameters wider, decreasing also slightly the accuracy of the prediction. The GM3 assures slightly more precise results, but the covariance parameters have to be more accurate. Furthermore, the influence of third order term suggests generalized functional models or series expansion of the covariance model to be worth investigation.

It is evident that the optimal RMS moves along the CL and N axes, i.e. the best CL decreases with an increase of a priori noise. The range of possible CL becomes narrow with a large N argument, especially in case of the GM3 function, which may need particularly precise CL determination. It is necessary, then, to be careful with increasing CL, especially in conditions of significant noise assumed in the dataset, i.e., when the smoothness factor has to be applied. The CL estimation is also strongly dependent on the function chosen what is suspected from Fig. 5 and confirmed in Figs. 6 and 7. Although CL may be approximately assessed from functional model fitting (Fig. 5), the graphical estimation of the N parameter is difficult. The error covariance is hard to observe amongst the residuals variance, because we don't know how the functional model represents the data distribution.

The actual accuracy resulting from the precise modeling of RTK data may be significantly better with the use of denser and spatially regular data. The data used in this experiment was rather sparse and randomly distributed because it came from the practical survey, not from designed experiment. Despite the sparse data, the RMS at the 0.56 m level was achieved, without any assumption on the closest point for the interpolation. These RMS values, as well as the N parameter, are strongly affected by the data spatial distribution and the obtainable resolution of the prediction.

Future work will be focused on different methods that investigate covariance parameters estimation, because a lack of the practical applications is found amongst considerable theoretical background in this domain. The parameters are often estimated with significant errors and not tested in the modeling process. The current research was necessary to start with the assessment of the parameters influence on LSC prediction. Many spatial modeling processes in Earth sciences and engineering require precise estimates of the covariance, therefore practical testing is desired in this area.

References

Andersen O. B., Knudsen P., 1998: Global marine gravity field from the ERS-1 and Geosat geodetic mission altimetry. Journal of Geophysical Research, **103**, (C4), 8129–8137.

Arlot S., Celisse A., 2010: A survey of cross-validation procedures for model selection. Statistics Surveys, **4**, 40–79 (doi:10.1214/09-SS054)

Bakuła M., Oszczak S., Pelc-Mieczkowska R., 2009: Performance of RTK Positioning in Forest Conditions: Case Study. Journal of Surveying Engineering-ASCE, **135**, 3, 125–130 (doi:10.1061/(ASCE)0733-9453(2009)135:3(125))

Bater C. W., Coops N. C., 2009: Evaluating error associated with lidar-derived DEM interpolation. Computers & Geosciences, **35**, 2, 289–300.

Bosy J., Graszka W., Leończyk M., 2007: A Multifunctional Precise Satellite Positioning System in Poland. European Journal of Navigation, **5**, 4, 2–6.

Briese C., Pfeifer N., Dorninger P., 2002: Applications of the Robust Interpolation for DTM determination. In: R. Kalliany, F. Leberl and F. Fraundorfer (Editors), Photogrammetric Computer Vision. International Archives of Photogrammetry and Remote Sensing, **34**, 3A, 55–61.

Crombaghs M., Oude Elberink S., Brügelmann R., de Min E., 2002: Assessing Height Precision of Laser Altimetry DEMs. International Archives of Photogrammetry and Remote Sensing, **34**, 3A, 85–90.

Darbeheshti N., Featherstone W. E., 2010: Tuning a gravimetric quasigeoid to GPS-levelling by non-stationary least-squares collocation. Journal of Geodesy, **84**, 7, 419–431.

Erdogan S., 2010: Modeling the spatial distribution of DEM error with geographically weighted regression: An experimental study. Computers & Geosciences, **36**, 34–43.

Eshagh M., 2009: On the convergence of spherical harmonic expansion of topographic and atmospheric biases in gradiometry. Contrib. Geophys. Geod., **39**, 4, 273–299.

Gonçalves-Seco L., Miranda D., Crecente R., Farto J., 2006: Digital terrain model generation using airborne LiDAR in a forested area Galicia, Spain. Proc. of 7th International Symposium on Spatial Accuracy Assessment in Natural Resource and Environmental Sciences, 169–180.

Hengl T., Bajat B., Blagojević D., Reuter H. I., 2008: Geostatistical modeling of topography using auxiliary maps. Computers & Geosciences, **34**, 12, 1886–1899.

Hofmann-Wellenhof B., Moritz H., 2005: Physical Geodesy. Springer, New York.

Kavzoglu T., Saka M. H., 2005: Modeling local GPS/levelling geoid undulations using artificial neural networks. Journal of Geodesy, **78**, 520–527.

Kitanidis P. K., 1983: Statistical estimation of polynomial generalized covariance functions and hydrologic applications. Water Resources Research, **19**, 4, 909–921.

Kohavi R., 1995: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada, 1137–1143.

Kotsakis C., 2007: Least-squares collocation with covariance-matching constraints. Journal of Geodesy, **81**, 10, 661–677.

Lim C-H., Mulgrew B., 2007: Linear prediction of range-dependent inverse covariance matrix (PICM) sequences. Signal Processing, **87**, 1412–1420.

Lohmann P., Koch A., Schaeffer M., 2000: Approaches to the filtering of laser scanner data. International Archives of Photogrammetry and Remote Sensing, **33**, (B3/1), 534–541.

Moritz H., 1980: Advanced Physical Geodesy. Herbert Wichmann Verlag, Karlsruhe.

Pardo-Igúzquiza E., Mardia K. V., Chica-Olmo M., 2009: MLMATERN: A computer program for maximum likelihood inference with the spatial Matérn covariance model. Computers & Geosciences, **35**, 1139–1150.

Pringle M. J., Schmidt M., Muir J. S., 2009. Geostatistical interpolation of SLC-off Landsat ETM+ images. ISPRS Journal of Photogrammetry and Remote Sensing, **64**, 6, 654–664.

Reguzzoni M., Sanso F., Venuti G., 2005: The theory of general kriging, with applications to the determination of a local geoid. Geophysical Journal International, **162**, 4, 303–314.

Revallo M., Valach F., Váczyová M., 2010: The geomagnetic storm of August 2010 as a lesson of the space weather modeling. Contrib. Geophys. Geod., **40**, 4, 313–322.

Veneziano D., Hallmark S., Souleyrette R., 2004: Accuracy of Light Detection and Ranging Derived Terrain Data for Highway Location. Computer-Aided Civil and Infrastructure Engineering, **19**, 2, 130–143.

Wise S. M., 2007: Effect of differing DEM creation methods on the results from a hydrological model. Computers & Geosciences, **33**, 10, 1351-1365.