

# A priori noise and regularization in least squares collocation of gravity anomalies

Wojciech Jarmolowski

University of Warmia and Mazury  
Faculty of Geodesy and Land Management  
Department of Satellite Geodesy and Navigation  
ul. Heweliusza 5, 10-724 Olsztyn, Poland  
e-mail: wojciech.jarmolowski@uwm.edu.pl

Received: 12 September 2013 / Accepted: 21 November 2013

**Abstract:** The paper describes the estimation of covariance parameters in least squares collocation (LSC) by the cross-validation (CV) technique called leave-one-out (LOO). Two parameters of Gauss-Markov third order model (GM3) are estimated together with a priori noise standard deviation, which contributes significantly to the covariance matrix composed of the signal and noise. Numerical tests are performed using large set of Bouguer gravity anomalies located in the central part of the U.S. Around 103 000 gravity stations are available in the selected area. This dataset, together with regular grids generated from EGM2008 geopotential model, give an opportunity to work with various spatial resolutions of the data and heterogeneous variances of the signal and noise. This plays a crucial role in the numerical investigations, because the spatial resolution of the gravity data determines the number of gravity details that we may observe and model. This establishes a relation between the spatial resolution of the data and the resolution of the gravity field model. This relation is inspected in the article and compared to the regularization problem occurring frequently in data modeling.

**Keywords:** gravity anomaly, least squares collocation, leave-one-out, covariance, noise

---

## 1. Introduction

Least squares collocation (LSC) is a spatial technique used in geodesy and related fields for estimating values of physical field at the positions, where it is unknown. The estimates are based on the measured values in some other individual points of the analyzed, spatially correlated field. LSC is used in geodesy for the interpolation or the interpolation combined with the transformation between different quantities, e.g. different functionals of anomalous gravity potential. This paper investigates the interpolation of gravity anomalies by planar LSC, which is a frequent technique used with gravity and other geodetic data. A special attention in this paper is put on the covariance parameters estimation for the planar covariance model, with particular focus on a priori noise standard deviation.

LSC is applied with the use of the local planar covariance function. Different spherical or planar covariance models are often investigated in geoid or gravity field modeling (Arabelos and Tscherning, 2003; Forsberg, 1987), however quite simple one is selected here. It has two parameters and is easy in the investigation of the problem. This study investigates a local covariance model with typical, local parameters, so that the model is in some sense intuitively chosen, as an example. Although the model is planar, the planar distance is replaced with spherical. The description and practical sense of this replacement is explained in the next section. The estimation of two covariance parameters and a priori noise standard deviation is performed by the leave-one-out (LOO) validation (Arlot and Celisse, 2010; Kohavi, 1995; Kusche and Klees, 2002). This technique is a cross-validation (CV) technique and the description of its current application is provided in the next section. LOO validation is proposed in the article despite its time-consuming nature, because this method is very straightforward and accurate. The aim of current paper is to find a priori noise standard deviation ( $\delta n$ ), signal variance ( $C_0$ ) and correlation length ( $CL$ ) using the smallest RMS of misfit between data and the prediction.

A special attention is focused on a priori noise variance estimation and all considerations are compared to the regularization problem, which may be found frequently in least squares applications. The regularization method in LSC prediction equations is equivalent to Tikhonov regularization (Eshagh and Sjöberg, 2011; Koch and Kusche, 2002; Moritz, 1980). Sometimes this technique is also associated with the name of D. L. Philips (Rummel et al., 1979; Kotsakis, 2007). The presented investigations follow the same mathematical rule and provide some observations related to LSC of the correlated, scalar data in 2D space. The problem of the regularization is known in geodetic literature for years, usually in the context of improperly posed problems. Rummel et al. (1979) apply the regularization to LSC and find its significant relation with the number and geographical distribution of the data. They also inspect data spacing in the downward continuation problem in the context of the regularization. The downward continuation and its regularization are frequently investigated together in the geodetic literature (Eshagh and Sjöberg, 2011; Kusche and Klees, 2002; Kotsakis, 2007; Jekeli and Garcia, 2002; Xu and Rummel, 1994). The applications of Tikhonov regularization in least squares techniques different than LSC, refer often to the spatial distribution of the data, which is main subject of this work. For example, (Eshagh and Sjöberg, 2011) apply various spatial resolutions and noises in the downward continuation of the satellite gravity gradiometry data and show their influence on the estimation results. Some authors use regularized approach to the data combination, which assumes the regularization together with variance component estimation (Koch and Kusche, 2002; Xu, 2009). Trojanowicz (2012) applies weighting in the quasigeoid modelling using gravity data inversion in a way that is close to above mentioned works.

Returning to implementation of the regularization in the LSC technique, an interesting example is provided by Pail et al. (2010), who find Tikhonov regularization as suitable for covariance model correction in the combination of the global and local

data used in collocation. Arabelos and Tscherning (1998) have applied an increasing error variance in the LSC of altimeter data and found that optimum prediction is attained with the error variance larger a few times than the measurement error. They have also compared finite and infinite functions in LSC and found an advantage of the correlated errors in some applications (Arabelos and Tscherning, 1999). This article, however, investigates the local function and uncorrelated a priori noise only, because these are convenient to handle and present the problem with ease.

The paper focuses on relations between the spatial resolution, actual data noise and the regularization factor. The data for the numerical experiment are selected to be approximately regular in the horizontal direction and therefore the gravity field resolution is individually limited in each of four sets. Strykowski (2000), in turn, applies the circular cutoff filter to remove some high frequency gravity signal from the gravity field model. Such filter is approximately equivalent to the assumption of some minimal distance limit to the closest point used in LSC. Both operations exclude the nearest gravity information from the interpolated model. This rule is applied in selection of data for the current LSC study. Some basic knowledge in the field of spectral data analysis may be helpful for the understanding of the stated problem, because although there is no spectral analysis in the numerical part, the discussion on the LSC problems is related with the spectrum of gravity data (Forsberg, 1984; Schwarz, 1984).

A very closely related investigation of the simple covariance models may be found in Smith and Milbert (1999). They, however, do not use regularization parameter, but match assigned noise and misfit of the prediction in the iterative process. The results of the current study may be also inspected this way, because some similar coherency will be also visible in the later sections. Marchenko et al. (2003) apply Tikhonov regularization parameter to collocation in the same way as it is presented here, but avoid time-consuming iterations using its simplified estimation. The current study, however, applies iterative LOO to show some details occurring in the process. The noise standard deviation is regarded as entire parameter, rather, than the product of the noise and regularization factor and this choice will be illustrated in the numerical part.

## 2. Trend removal, LSC and LOO validation

The scalar field of gravity anomalies in 2D plane may be represented by the addition of the trend and residuals of the signal (Moritz, 1980; Rao and Toutenburg, 1995):

$$\Delta \mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \Delta \mathbf{g}^r \quad (1)$$

where  $\Delta \mathbf{g}$  is the observed gravity,  $\boldsymbol{\beta}$  includes the vector of unknown trend parameters and  $\mathbf{X}$  is design matrix of the trend.  $\Delta \mathbf{g}^r$  represents residual gravity field after the subtraction of the long-wavelength signal. Various theoretical assumptions and

numerical techniques are used for its derivation. The spherical or ellipsoidal harmonic expansion is always regarded as an efficient method for signal analyses and may provide an excellent trend model. Such model constitutes an accurate mathematical approximation of the long wavelength signal part, however, simpler models of the trend are often efficient in the local investigations. Spatially limited areas may have similar representations of the long-wavelength part from harmonic expansion, as well as from the simple polynomial trend. Moreover, polynomial trend or even mean subtraction may be sufficient for some purposes. Various trend orders are often used in practice (Kryński and Łyszkowicz, 2006; Osada et al., 2005). The second order polynomial trend is applied in this paper, because this is found as sufficient to attain graphically acceptable normal distribution of the residuals and no evident bias. The matrix  $\mathbf{X}$  used for detrending of the data reads:

$$\mathbf{X} = \begin{bmatrix} 1 & \varphi_1 & \lambda_1 & \varphi_1^2 & \lambda_1^2 & \varphi_1 \lambda_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \varphi_n & \lambda_n & \varphi_n^2 & \lambda_n^2 & \varphi_n \lambda_n \end{bmatrix}, \quad (2)$$

where  $n$  is the number of the observations. This form of the trend depends on the spatial data distribution and is in many cases sufficient to obtain residuals representing approximately normal distribution. The trend removal is performed using orthogonal projection:

$$\Delta \mathbf{g}^T = \Lambda \Delta \mathbf{g}. \quad (3)$$

The creation of the  $\Lambda$  is based on singular value decomposition (SVD) of some combination of  $\mathbf{X}$ , i.e.  $\mathbf{U}\Sigma\mathbf{V}^T$  of the square block matrix  $[\mathbf{X};\mathbf{0}]$  is realized where  $\mathbf{0}$  is a matrix of zeros with dimensions  $n \times (n-p)$  (Golub and Van Loan, 1989; Rao and Toutenburg, 1995). Assuming  $\mathbf{U}$  to be composed of  $[\mathbf{U}_1;\mathbf{U}_2]$ ,  $\mathbf{U}_2$  with dimensions  $n \times (n-p)$  is used to construct  $\Lambda$  as follows:

$$\Lambda = \mathbf{U}_2^T. \quad (4)$$

$\mathbf{V}$  is the unitary matrix, as well as  $\mathbf{U}$  and both are orthogonal for real numbers in  $\mathbf{X}$ .  $\Sigma$  is the diagonal matrix with singular values in main diagonal. The obtained matrix  $\Lambda$  meets the conditions:

$$\Lambda \mathbf{X} \beta = 0. \quad (5)$$

and

$$\Lambda \Lambda^T = \mathbf{I}. \quad (6)$$

We know from Eq. (4) that the number of detrended data is  $n-p$  now and we lose first  $p$  points from the dataset. This number has no significance in terms of accuracy since all used sets have almost 300 points. Then the residuals can be computed using Eq. (3). The LSC equation for detrended gravity data reads:

$$\tilde{\Delta \mathbf{g}}^r = \mathbf{C}_p^T \cdot (\mathbf{C} + \mathbf{D})^{-1} \cdot \Delta \mathbf{g}^r, \quad (7)$$

where  $\tilde{\Delta \mathbf{g}}^r$  is predicted vector of residuals.  $\mathbf{C}$  is the covariance matrix of the residuals,  $\mathbf{C}_p$  is the covariance matrix between the predicted points and the data and  $\mathbf{D}$  represents noise covariance matrix. We investigate a case, when average a priori noise is homogeneous and uncorrelated, i.e.:

$$\mathbf{D} = \delta n^2 \cdot \mathbf{I}_n, \quad (8)$$

where  $\delta n$  represents a priori noise standard deviation and is equal for all points. Regularized LSC equation in the case, when error covariance matrix needs a factor  $\alpha$  to attain CV minimum will be:

$$\tilde{\Delta \mathbf{g}}^r = \mathbf{C}_p^T \cdot (\mathbf{C} + \alpha \mathbf{D})^{-1} \cdot \Delta \mathbf{g}^r. \quad (9)$$

The factor  $\alpha$  is usually selected empirically, because it is dependent on the local data sample. Some methods of its derivation are described in the literature given in the previous section. This work, however, does not attempt to estimate  $\alpha$ , but  $\mathbf{D}$  matrix with a priori noise.

The choice of the covariance model is arbitrary in some sense, but it is based on the frequent applications of Gauss-Markov third order model (GM3) in the literature in geoid (Kavzoglu and Saka, 2005) or gravity interpolation (Moreaux, 2008). GM3 model is

$$GM3(C_0, CL, s) = C_0 \left( 1 + \frac{s}{CL} + \frac{s^2}{3 \cdot CL^2} \right) \cdot \exp\left(\frac{-s}{CL}\right), \quad (10)$$

where  $s$  is replaced by the spherical distance between data points. Although spherical distance is a typical variable in the spherical covariance models it is adopted here to work as a variable in the planar model (Eq. 10). This combination is applied, because no advantage is expected from the cartographic projection, since the area of the regional data extends to several degrees. The maximum distance between the prediction point and the data that is used in the LSC is  $1^\circ$ . This implies  $CL$  values (Figs 3 and 4) that are often used with the planar models (Andersen and Knudsen, 1998; Smith and Milbert, 1999). Sometimes, a better modeling of the long-wavelength signal is considered as an advantage of the spherical models (Arabelos and Tscherning 1998). However, the long-wavelength signal is correlated at particularly long distances

and therefore is not expected to find it in the matrix  $\mathbf{D}$ , which cannot store correlated data parts. On the other hand, the second order trend (Eq. 2) removes much from the long-wavelength part. Summarizing, the signal covariance matrix  $\mathbf{C}$  may be assumed as more affected by long-wavelength errors, than  $\mathbf{D}$ , which holds the uncorrelated part that comes from the survey. The variable distance  $s$  is therefore calculated using the spherical distance formula in Eq. (10), as well as in the empirical covariance function (ECF) and  $s = \psi$  in the article. GM3 results in the following covariance values for the signal:

$$C(\psi) = \begin{cases} GM3(C_0, CL, \psi) & \text{for } \psi > 0 \\ C_0 & \text{for } \psi = 0 \end{cases} \quad (11)$$

The parameters of this or similar covariance models may be estimated by fitting the analytical model into the empirical covariance values calculated in the following way:

$$\begin{aligned} \forall(i,j) \quad | \quad \cos \psi &= \cos \theta_i \cos \theta_j + \sin \theta_i \sin \theta_j \cos(\lambda_i - \lambda_j) \\ &: \quad EC(\psi | \Delta \mathbf{g}^r) = \frac{\sum_{i,j}^k \Delta \mathbf{g}_i^r \Delta \mathbf{g}_j^r}{k} \end{aligned} \quad (12)$$

where  $\theta_i = \pi/2 - \varphi_i$  and  $\psi$  is the spherical distance. Calculated values are usually grouped using intervals of uniformly increasing spherical distance. It should be noted here, that the length of the interval, which may be also treated as the sampling rate, plays a significant role in the final shape of ECF and should be consistent with the actual resolution of the data.

The collocation formula by Eq. (7) is used in the validation test by LOO. The noise standard deviation  $\delta n$  is treated as third parameter of the data covariance matrix ( $\mathbf{C}+\mathbf{D}$ ). LOO validation rule is quite frequently applied in the literature (Darbeheshti and Featherstone, 2009; Kusche and Klages, 2002), but other, similar CV technique may be also efficient. In LOO, the vector  $\Delta \mathbf{g}^r$  estimated by Eq. (7) is compared to  $\Delta \mathbf{g}^r$  via computation of the RMS and this is repeated for each set of the covariance parameters. The original, residual data vector in the Eq.(7) is always replaced by the vector  $\Delta \mathbf{g}_{(n-1) \times 1}^r$ , i.e. the analyzed point  $i$  is omitted in the vector  $\Delta \mathbf{g}^r$  as well as in the matrices  $\mathbf{C}_p$ ,  $\mathbf{C}$  and  $\mathbf{D}$ . The formula of root mean square (RMS) in LOO (RMSL) may be written as:

$$\begin{aligned}
 &RMSL\left((C_0, CL, \delta n) \mid \left(\tilde{\Delta \mathbf{g}}_{n \times 1}^r, \Delta \mathbf{g}_{n \times 1}^r\right)\right) = \\
 &\left\{ \begin{array}{l} \sqrt{n^{-1} \sum_{i=1}^n \left(\Delta \mathbf{g}_i^r - \tilde{\Delta \mathbf{g}}_i^r\right)^2} \\ \tilde{\Delta \mathbf{g}}_i^r = \mathbf{C}_{P(n-1) \times 1}^T \cdot \left(\mathbf{C}_{(n-1) \times (n-1)} + \mathbf{D}_{(n-1) \times (n-1)}\right)^{-1} \cdot \Delta \mathbf{g}_{(n-1) \times 1}^r \wedge \\ \Delta \mathbf{g}_i^r \notin \Delta \mathbf{g}_{(n-1) \times 1}^r \end{array} \right. \quad (13)
 \end{aligned}$$

RMSL is a measure of prediction precision with the chosen parameter set, therefore the smallest RMSL is assumed as indicating the optimum parameter set out of all combinations that are tested. This kind of estimation provides optimal  $\delta n$ , which is hard to determine by the covariance function fitting. A posteriori error of the prediction, which is strongly related to  $\delta n$  is (Hofmann-Wellenhof and Moritz, 2005)

$$m_{P_i}^2 = C_0 - \mathbf{C}_{P_i}^T \cdot (\mathbf{C} + \mathbf{D})^{-1} \cdot \mathbf{C}_{P_i}, \quad (14)$$

assuming, that  $\mathbf{C}_{P_i}$  is a vector representing covariance for one point only.

### 3. Gravity samples and test assumptions

Gravity data for numerical test originate from the U.S. gravity database, which is available at the website of University of Texas at El Paso (Hildenbrand et al., 2002). Large, regional area of Bouguer gravity data (around  $6^\circ \times 9^\circ$ ) is useful for work with various resolutions of the gravity signal. This is an important matter of the current LSC test. Although the numerical studies of the covariance parameters by LOO are performed solely in the space domain, the horizontal resolution of the data has some relation with with the sampling in the frequency domain. The relation of the LSC with the actual spectrum of gravity anomalies leads to the discussion on the spectral properties of used data. All local data are characterized by specific variances of the signal at individual resolutions, which may be expressed by their own power spectrum (Forsberg 1984). Particular frequencies of sampling that correspond to spatial resolutions compose the gravity signal. The global gravity signal, which is

also used in this study is nowadays expanded in spherical or ellipsoidal harmonics (Hofmann-Wellenhof and Moritz, 2005; Pavlis et al., 2012).

Gravity anomalies have significant signal variance at lower and higher degrees of harmonic expansion, oppositely to geoid, where larger part of the signal is cumulated at lower degrees. We may expect some gravity variance at harmonic degrees equivalent to the highest possible resolution of the analyzed data and significantly larger variances at lower degrees related with the trend in case of set 1, set 2 and set 3. The trend is here a very rough approximation of the lower frequency signal. The rule of its removal is the same for all four sets and therefore particular sets obtain different lower limit of the spectrum, due to different sizes. The sampling density is completely different for every dataset, which results in the various upper limits of the gravity spectrum. The lower limit of the data spectrum is then roughly defined by the trend surface. The upper limit is related with the spatial resolution and therefore sampling rate plays a crucial role in the experiment. Four datasets are sampled from the full dataset of gravity anomalies, presented in Fig. 1a. Figure 1b describes the sampling scheme and shows four subsets: set 1 - largest with approximate resolution  $0.5^\circ$ , set 2 –  $0.25^\circ$ , set 3 –  $0.1^\circ$  and set 4, which covers the smallest area and has original resolution of about  $0.03^\circ$  (Fig. 1b). The intervals of one degree in north and east have different lengths at the current latitude. However, although angularly equal sampling may affect accuracy of the final results, it is assumed to be negligible for principal findings in the paper. Consequently, the intervals are angularly equal and this is convenient in further subsets description.

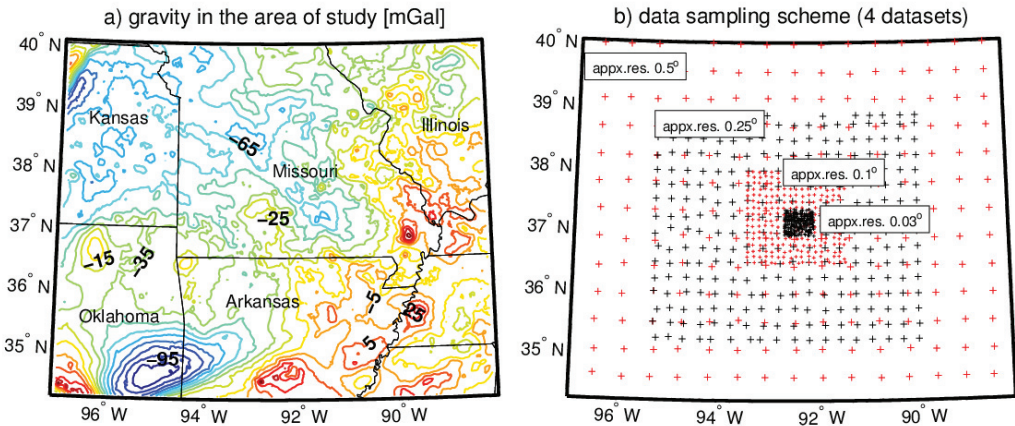


Fig. 1. Gravity data and scheme of sampling (four datasets)

Simply polynomial trend of the second order (Eq. 2) has been removed separately for each set to produce residuals, which represent approximately normal distribution. This condition is often used in the assessment of the data samples. The residuals



should sufficiently well represent the assumed distribution if the proper variance and covariance values are desired from e.g. ECF estimation. This is checked and shown in the Fig. 2. All four sets of the residuals represent graphically acceptable distribution, as no bias or significant discontinuity may be found. However, the influence of the incorrectness in respect to the normal distribution is not known here in numbers.

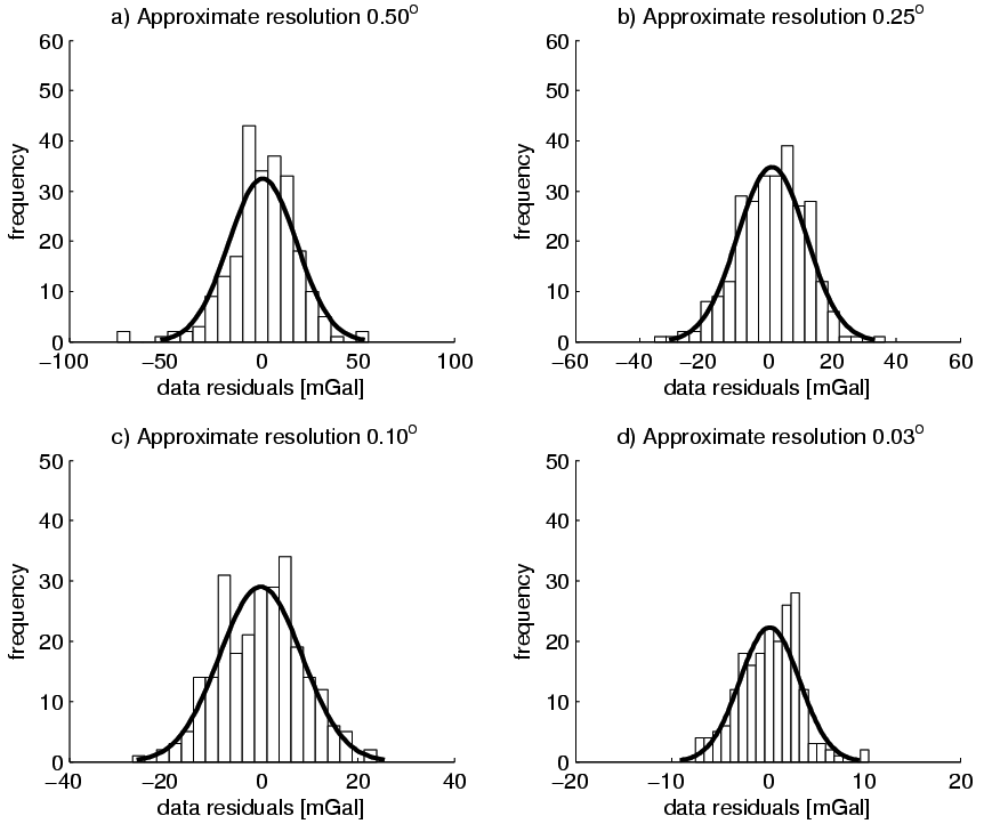


Fig. 2. Histograms of data residuals and normal distribution curves fitted

ECFs of the residuals are calculated to provide rough estimates of  $C_0$  and  $CL$  parameters. The GM3 model is graphically fitted to the estimated empirical covariance values neglecting  $\delta n$ .  $C_0$  of the residuals may be approximately estimated as the covariance at the distance zero. The individual  $CL$  parameters are obtained from the fitting of analytical model (Fig. 3). Fitting is a widely used technique in the estimation of the covariance parameters (Arabelos and Tscherning, 2003; Darbeheshti and Featherstone, 2009; Smith and Milbert, 1999). Some authors use advanced analytical fitting methods, however an important factor is the sampling rate, because it may affect the shape of the ECF, especially at shorter distances.

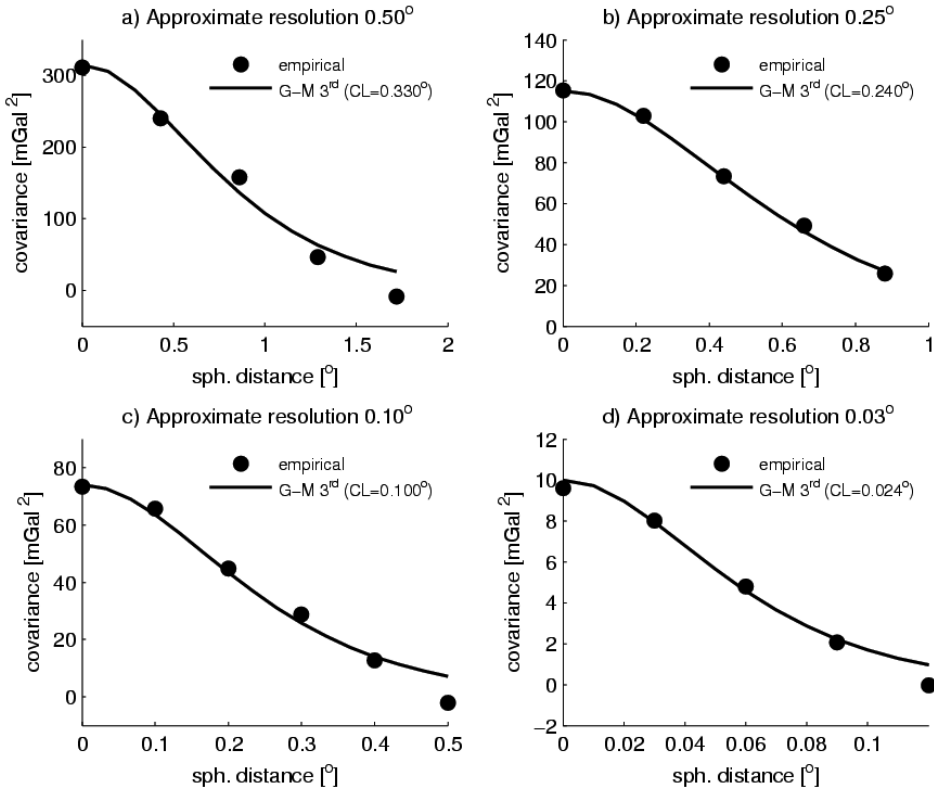


Fig. 3. ECFs of residuals and approximate fit of GM3 model

Too dense sampling results in the variations of ECF values at short distances. The sampling rate has also a general influence on the empirical covariance values. The sampling of ECFs for respective sets is different here and dependent on approximate data resolution. It should be stressed, that  $\delta n$  is not estimated from Fig. 3. The noise is treated as uncorrelated part of the data and it is analyzed only by LOO.

#### 4. Numerical test and discussion

LOO validation of the LSC results is performed with four datasets using different ranges of  $CL$  and  $\delta n$  parameters, because different ones are expected to be estimated. The minimum RMSL indicates the optimal values of covariance parameters  $CL$  and  $\delta n$ .  $C_0$  is constant in this test for one dataset and is based on the residuals variance. The LSC process is repeated with different pairs of  $CL$  and  $\delta n$  to fill assumed grid of RMSL values. RMSL is computed by Eq. 13, i.e. by subtracting estimates from respective original values. The variable, average a priori error and  $CL$  result in RMSL

values for particular pairs of parameters. The one of those combinations provides the minimum RMSL.

$CL$  optimums based on minimum RMSL in Fig. 4 are close to those from fitting (Fig. 3), although  $\delta n$  was neglected in the fitting. Some factors affect  $CL$  approximation by LOO, however a general conclusion may be drawn that LOO search of  $CL$  values gives close results to those from the GM3 model fitting. The horizontal resolution of the data increases starting from set 1 and ending at set 4. The parameter  $\delta n$ , which represents noise standard deviation decreases respectively from about 7.5 mGal (Fig. 4a) to 1.0 mGal (Fig. 4d). Set 2 needs intermediate value of around 5.5 mGal (Fig. 4b) and set 3 requires 3.0 mGal for minimum RMSL. The  $\delta n$  step size in LOO is 0.5 mGal for set 1 and 2, 0.25 for set 3 and 0.05 for set 4. The

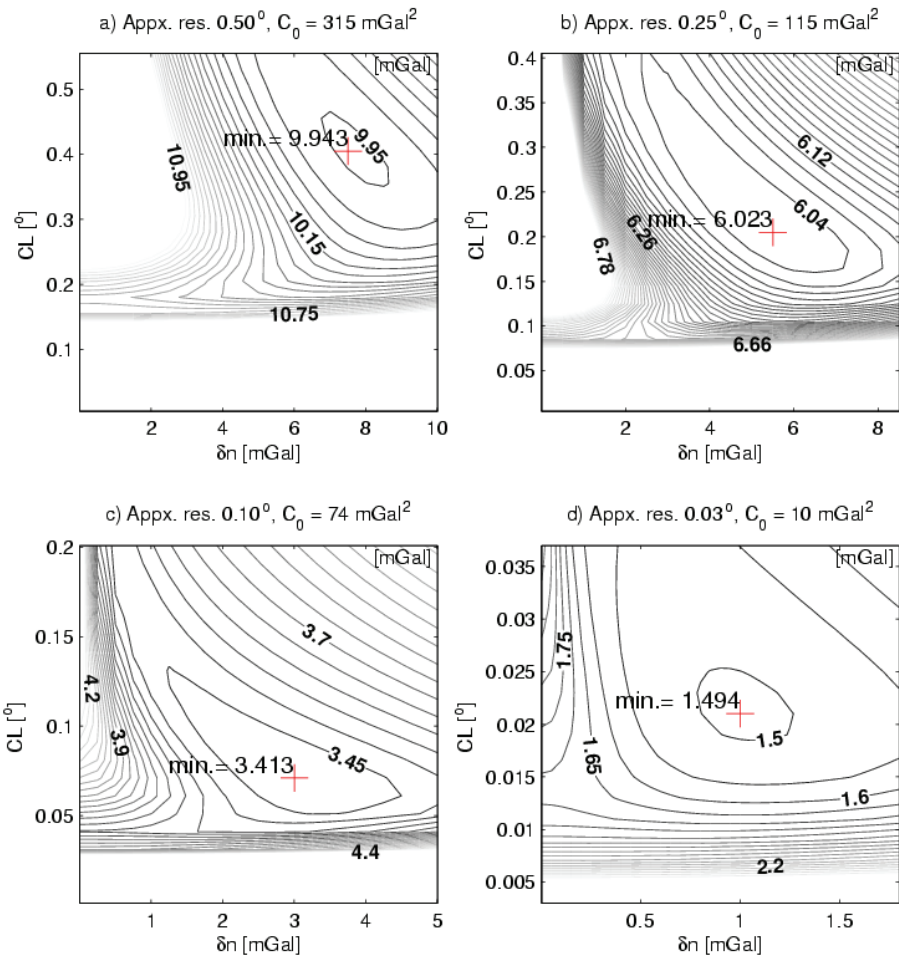


Fig. 4. LOO estimation of  $CL$  and  $\delta n$  based on optimal RMSL ( $C_0$  is approximated by the residuals variance)

parameters from the regions of small  $CL$  or  $\delta n$  are especially inapplicable, because RMSL rises rapidly in these places, especially for the lower resolution data (Figs. 4a-c). This indicates that the spatial resolution or some factors closely related may substantially contribute to a priori noise of the data. The fact that the observational error is not sufficient to represent a priori noise was previously observed and reported in the literature (Arabelos and Tscherning 1998, Sadiq et al. 2010).

The examples of LSC with optimal  $CL$  and  $\delta n$  used are given in the Table 1 in terms of RMSL and the statistics related to a posteriori error estimate  $m_p$ . The computation with the same  $C_0$  and  $CL$ , but decreased  $\delta n$  is also provided to show an influence of too small  $\delta n$  on RMSL and  $m_p$ . A posteriori error is in some cases more sensitive to covariance parameters change, than the result itself. This effect was reported before e.g. by Sansó et al. (1999), however RMSL also varies noticeably.

Table 1. RMSL and a posteriori standard deviation ( $m_p$ ) of the predictions with optimal parameters and with decreased  $\delta n$  (Eq. 14)

[mGal]	$\delta n$	RMSL	Min. $m_p$	Mean $m_p$	Max. $m_p$
<b>SET 1 (232 points)</b>					
$C_0 = 315 \text{ mGal}^2$ $CL = 0.330^\circ$	7.50	9.94	5.53	6.09	9.34
	1.00	12.17	2.95	3.44	6.23
<b>SET 2 (273 points)</b>					
$C_0 = 115 \text{ mGal}^2$ $CL = 0.240^\circ$	5.50	6.02	3.55	4.23	6.18
	1.00	7.06	1.88	2.48	4.19
<b>SET 3 (259 points)</b>					
$C_0 = 74 \text{ mGal}^2$ $CL = 0.100^\circ$	3.00	3.41	2.75	3.28	4.91
	0.50	3.93	1.80	2.33	4.04
<b>SET 4 (202 points)</b>					
$C_0 = 10 \text{ mGal}^2$ $CL = 0.024^\circ$	1.00	1.49	0.76	1.23	2.43
	0.10	1.74	0.39	0.91	2.36

The estimated  $\delta n$  value is subsequently compared with some spectrum of gravity anomaly under some assumption. More precisely, it is necessary to find this part of the gravity signal, which is present in the data, however the limited horizontal resolution excludes it from the correlated field. The global geopotential model EGM2008 is used for this purpose. The maximum spatial resolution in each dataset limits the upper signal frequency to the degree, which is approximately equivalent to average minimum distance between the neighboring points. Of course, we need quite homogeneous horizontal resolution of the data to discuss the problem. The maximum harmonic expansion of EGM2008 corresponds to  $5'$  of the horizontal resolution (Pavlis et al., 2012).  $5'$  is around  $0.08333^\circ$ , so it is decided to use EGM2008 with the smallest grid spacing equal  $0.05^\circ$ , which does not exceed much the maximum

resolution. Three grids of gravity anomalies are calculated online using ICGEM website to limited harmonic degree expansion (set 1 – 360, set 2 – 720, set 3 - 1800). The limited degrees correspond to the approximate maximum resolutions of the sets 1, 2 and 3. The gravity residuals are next computed for all three areas, by subtracting

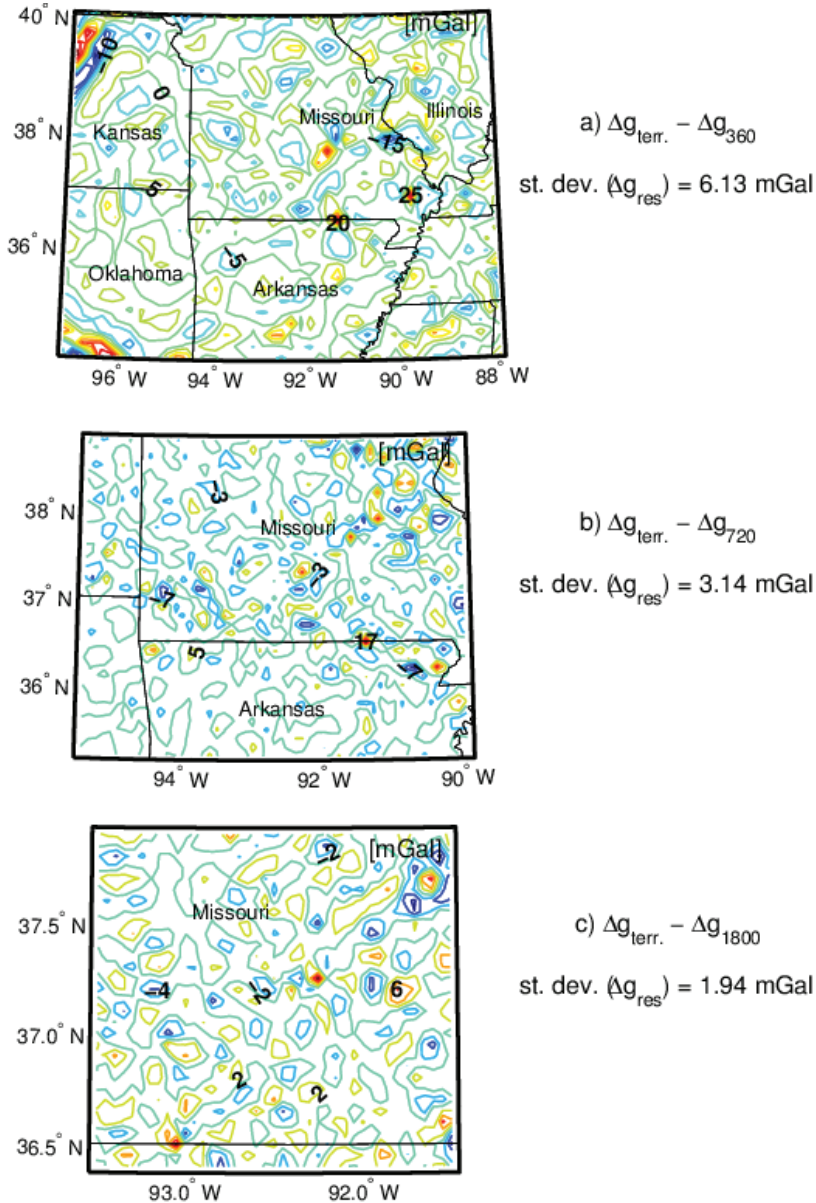


Fig. 5. Residual gravity from differences between roughly gridded point data and EGM2008 limited to degree that corresponds to datasets resolution (samples 1, 2, 3 – horizontal scales are different)

the degree-limited gravity signal from the grids calculated from all available gravity information in the involved point data. The grids of maximum frequency have been made by the simple triangulation to keep maximum available gravity information and avoid smoothing. EGM2008 grids and residual gravity anomalies are computed using different grid spacing for particular sets in order to be graphically representative and computationally efficient i.e.: set 1 –  $0.2^\circ$ , set 2 –  $0.1^\circ$  and set 3 –  $0.05^\circ$ . The residuals for set 4 have been not calculated, since the resolution of set 4 ( $0.03^\circ$ ) is maximum resolution available in the frame of this work and EGM2008 has also no degrees corresponding to  $0.03^\circ$ . Fig. 5 shows residual gravity signal for the areas of sets 1, 2 and 3.

The standard deviation of the residuals in Fig. 5 can be compared to the noise standard deviation  $\delta n$ , since the correlation of these signal frequencies cannot be recognized by the ECF. The sampling of the empirical covariance cannot be much smaller than the data resolution, because improper values may occur at small distances, due to the insufficient number or even lack of data pairs representing small distances. LSC cannot find some parts of the signal (Fig. 5) as correlated, since the smallest distance used to compute  $C$  matrix corresponds to minimum distance between the neighboring points. To resume, the signal can be interpolated also between the data points, but will be smoothed to the frequency dependent on the data resolution. In other words, it's hard to interpolate the signal at frequency  $0.03^\circ$ , when the prediction point is situated between two data points and  $0.25^\circ$  from each one. The optimum  $\delta n$  parameter based on RMSL in Fig. 4 is always slightly larger than the standard deviations of gravity residuals in Fig. 5. This should be explained in the further studies, as the limited resolution may be not only factor that influences  $\delta n$ .

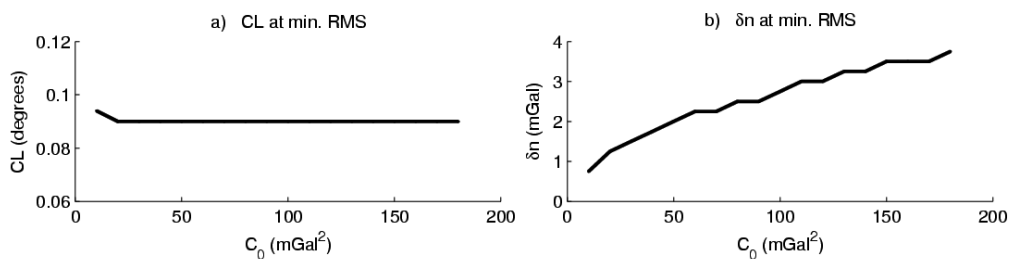


Fig. 6. CL and  $\delta n$  for minimum RMSL computed with variable  $C_0$  parameter applied to set 3.

The test described above is based on the assumption that  $C_0$  in the matrix  $C$  has to be equal or approximately equal to the residuals variance. This is reasonable, since the covariance function should represent actual variance and covariance of the residuals. However, in the case of inaccurate removal of the long-wavelength part of the signal, some bias from the lower harmonics may affect actual variance. An additional test is shown in Fig. 6, where variable  $C_0$  is applied to gravity residuals

in set 3.  $CL$  and  $\delta n$  values are derived in 2D plane in the same way as in Fig. 4, for minimum RMSL. This iteration is performed for the set 3 only. Fig. 6 presents  $CL$  and  $\delta n$  found as optimal at local minimums of RMSL. Some correlation is observed between  $C_0$  and  $\delta n$ , but  $CL$  remains apparently constant in relation to variable  $C_0$ .

## 5. Conclusions

Numerical test with large dataset allows for different sampling of surveyed data, which may have an unknown, heterogeneous measurement error. LSC of variously spaced data reveals large influence of the factors related with the data spatial resolution. The measurement error is presumably smaller than the estimated  $\delta n$  values, at least in sets 1, 2 and 3. The values of  $\delta n$ , which enable optimum RMSL, are different for different data resolutions. These  $\delta n$  are also very similar to minimum RMSL obtained, which additionally emphasizes their advance amongst different  $\delta n$  used. It can be suspected, that sparse data (e.g. GNSS/leveling, satellite data from short missions) may require taking into account their spectrum and the horizontal or spatial resolution when estimating a priori noise variance. These observations lead to the conclusion that if the data sampling is corresponding to accuracy or denser,  $\delta n$  may represent indeed the survey error. The loss of the accuracy related with the change of  $\delta n$  is smallest in Fig. 4d, which may be caused by small data variance at frequency corresponding to  $0.03^\circ$ .

Supplementary test with variable  $C_0$  shows scaling effect that exists between  $C_0$  and  $\delta n$ . This suggests fixing variance of the functional covariance model to have  $\delta n$  in its original spatial scale. Some least squares applications however, use normalized covariance matrices or normalized semi-variograms. It is obvious then, that the parameter representing noise variance will be also scaled and non-comparable with the actual noise.

In the future work, the spectral analysis of the local signals should be considered. It may be a helpful tool in the assessment of the variance of the higher frequency signal, which has no correlation due to the limited horizontal or spatial resolution.

## Acknowledgements

This work is prepared in the frame of statutory research. Gravity data originate from Gravity Database of the U.S. and were downloaded from the website of the University of Texas at El Paso (<http://research.utep.edu>). The service of International Centre for Global Earth Models (ICGEM) was the source of global gravity anomalies (<http://icgem.gfz-potsdam.de/ICGEM/>).

## References

- Andersen, O.B. & Knudsen, P. (1998). Global marine gravity field from the ERS-1 and Geosat geodetic mission altimetry. *J. Geophys. Res.*, 103 (C4), 8129-8137.
- Arabelos, D. & Tscherning C. C. (1998). The Use of Least Squares Collocation Method in Global Gravity Field Modeling, *Phys. Chem. Earth*, 23 (1), 1-12.
- Arabelos, D., & Tscherning C.C. (1999). Gravity field recovery from airborne gravity gradiometer data using collocation and taking into account correlated errors, *Phys. Chem. Earth*, (A) 24 (1), 19–25.
- Arabelos, D., & Tscherning C. C. (2003). Globally covering a-priori regional gravity covariance models, *Adv. Geosci.*, 1, 143-147, DOI:10.5194/adgeo-1-143-2003.
- Arlot, S. & Celisse A., 2010. A survey of cross-validation procedures for model selection. *Stat. Surv.*, 4, 40–79, DOI: 10.1214/09-SS054.
- Darbeheshti, N., & Featherstone W.E. (2009). Non-stationary covariance function modelling in 2D least-squares collocation. *J. Geod.*, 83(6), 495–508. DOI:10.1007/s00190-008-0267-0.
- Eshagh, M. & Sjöberg L E. (2011). Determination of gravity anomaly at sea level from inversion of satellite gravity gradiometric data. *Journal of Geodynamics*, 51, 366-377.
- Forsberg, R. (1987). A new covariance model for inertial gravimetry and gradiometry, *J. Geophys. Res.*, 92 (B2), 1305-1310.
- Forsberg, R. (1984). Local covariance functions and density distributions. Report No. 356, Department of Geodetic Science and Surveying, The Ohio State University, Columbus.
- Golub, G. A., & Van Loan C. F., 1989, *Matrix Computations: 2nd Ed.*, John Hopkins University Press, Baltimore.
- Hildenbrand, T. G., Briesacher A., Flanagan G., Hinze W. J., Hittelman A. M., Keller G. R., Kucks R. P., Plouff D., Roest W., Seeley J., Smith D. A., & Webring, M. (2002). Rationale and Operational Plan to Upgrade the U.S. Gravity Database: U.S. Geological Survey Open-File Report 02–463, 12 p.
- Hofmann-Wellenhof, B. & Moritz H. (2005). *Physical Geodesy*, Springer, New York.
- Jekeli, C. & Garcia R. (2002). Local geoid determination with in situ geopotential data obtained from satellite-to-satellite tracking data, in: Sideris, M.G. (ed.), *Gravity, Geoid and Geodynamics*, 2000, Springer, Berlin, 123-128.
- Kavzoglu, T. & Saka M. H. (2005). Modeling local GPS/levelling geoid undulations using artificial neural networks. *J. Geod.*, 78, 520-527. DOI 10.1007/s00190-004-0420-3.
- Koch, K. R., & Kusche J. (2002). Regularization of geopotential determination from satellite data by variance components. *J. Geod.*, 76, 259-268.
- 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, Canada, 2, 1137-1143.
- Kotsakis, C. (2007). Acovariance-adaptive approach for regularized inversion in linear models. *Geophys. J. Int.*, 171, 509–522.
- Kryński, J., & Lyszkowicz A. (2006). Fitting gravimetric quasigeoid model to GPS/levelling data in Poland, IAG Proceedings from 1st International Symposium of the International Gravity Field Service (IGFS), 28 August - 1 September 2006, Istanbul, Turkey.
- Kusche, J., & Klees R. (2002). Regularization of gravity field estimation from satellite gravity gradients. *J. Geod.*, 76, 359–368.
- Marchenko, A., Tartachynska Z., Yakimovich A., & Zablotskyj F. (2003). Gravity anomalies and geoid heights derived from ERS-1,ERS-2, and Topex/Poseidon altimetry in the Antarctic peninsula area, Proceedings of the 5th International Antarctic Geodesy Symposium AGS'03, September 15-17, Lviv, Ukraine, SCAR Report No. 23, <http://www.scar.org/publications/reports/23/>
- Moreaux, G. (2008). Compactly supported radial covariance functions, *J. Geod.*, 82, 431-443, DOI 10.1007/s00190-007-0195-4.
- Moritz, H., 1980. *Advanced Physical Geodesy*, Herbert Wichmann Verlag, Karlsruhe.



- Osada, E., Kryński J. & Owczarek M. (2005). A robust method of quasigeoid modelling in Poland based on GPS/levelling data with support of gravity data, *Geodesy and Cartography* 54 (3), 99–117.
- Pail, R., Reguzzoni M., Sansò F., Kühtreiber N. (2010). On the combination of global and local data in collocation theory. *Studia Geophys Geod.*, 54 (2), 195-218.
- Pavlis, N. K., Holmes S. A., Kenyon S. C., & Factor J. F. (2012). [The development and evaluation of Earth Gravitational Model \(EGM2008\)](#), *J. Geophys. Res.*, 117, B04406, DOI:10.1029/2011JB008916.
- Rao, C. R., & Toutenburg H. (1995). *Linear Models: Least Squares and Alternatives*. New York: Springer-Verlag, pp. 352.
- Rummel, R., Schwarz K. P., & Gerstl M. (1979). Least squares collocation and regularization. *Bull. Geod.*, 53, 343–361.
- Sadiq, M., Tscherning C.C., & Ahmad Z. (2010). Regional gravity field model in Pakistan area from the combination of CHAMP, GRACE and ground data Using least squares collocation: A case study. *Adv. Space Res.*, 46, 1466-1476.
- Sansò, F., Venuti G. & Tscherning C.C. (1999). A theorem of insensitivity of the collocation solution to variations of the metric of the interpolation space. In Schwarz, K.P. (Ed.), *Geodesy Beyond 2000. The Challenges of the First Decade, International Association of Geodesy Symposia* 121, 233–240, Springer, Berlin.
- Schwarz, K. P. (1984). Data types and their spectral properties, in: Schwarz (ed.) *Local gravity field approximation*, Beijing International Summer School.
- Smith, D. A., & Milbert D. G. (1999). The GEOID96 high-resolution geoid height model for the United States. *J Geod.*, 73 (5), 219–236.
- [Strykowski, G. \(2000\). Silkeborg Gravity High Revisited: Horizontal Extension of the Source and its Uniqueness. \*Phys. Chem. Earth \(A\)\*, 25 \(4\), 375-380.](#)
- Trojanowicz, M. (2012). Local quasigeoid modelling using gravity data inversion technique - analysis of fixed coefficients of density model weighting matrix, *Acta Geodyn. Geomater.*, 9 (3) (167), 269–281.
- Xu, P. (2009). Iterative generalized cross-validation for fusing heteroscedastic data of inverse ill-posed problems. *Geophys. J. Int.*, 179, 182–200.
- Xu, P., & Rummel R. (1994). A simulation study of smoothness methods in recovery of regional gravity fields. *Geophys. J. Int.*, 117, 472–486.

## **Szum a priori i regularyzacja w kolokacji najmniejszych kwadratów anomalii grawimetrycznych**

**Wojciech Jarmolowski**

Uniwersytet Warmiński Mazurski  
Wydział Geodezji i Gospodarki Przestrzennej  
Zakład Geodezji Satalitarnej i Nawigacji  
ul. Heweliusza 5, 10-724 Olsztyn, Polska  
e-mail: wojciech.jarmolowski@uwm.edu.pl

### **Streszczenie**

Artykuł opisuje estymację parametrów kowariancji w kolokacji najmniejszych kwadratów (LSC) przy pomocy techniki krosvalidacji nazywanej leave-one-out (LOO). Wyznaczane są dwa parametry modelu Gaussa-Markova trzeciego rzędu (GM3) wraz z odchyleniem standardowym szumu a priori, które ma znaczny wpływ na macierz kowariancji złożoną z sygnału i szumu. Testy numeryczne przeprowadzono na dużym zbiorze anomalii grawimetrycznych Bouguera z obszaru centralnej części USA. Obszar ten mieści około 103000 pomiarów grawimetrycznych. Dane te wraz z regularnymi siatkami wygenerowanymi z modelu geopotencjalnego EGM2008 pozwalają na pracę z różną rozdzielczością przestrzenną i różnymi wariancjami sygnału i szumu. Odgrywa to kluczową rolę w badaniach numerycznych, ponieważ rozdzielczość przestrzenna danych grawimetrycznych wyznacza liczbę szczegółów pola siły ciężkości, które możemy obserwować i modelować. Oznacza to relację pomiędzy rozdzielczością przestrzenną danych i rozdzielczością modelu pola siły ciężkości. Związek ten jest w artykule analizowany i porównywany z problemem regularyzacji, występującym często w modelowaniu danych przestrzennych.