Bollettino di Geofisica Teorica ed Applicata Vol. 60, n. 3, pp. 457-474; September 2019
DOI 10.4430/bgta0265

On the relations between signal spectral range and noise variance in least-squares collocation and simple kriging: example of gravity reduced by EGM2008 signal

W. Jarmołowski

Faculty of Geodesy, Geospatial and Civil Engineering, University of Warmia and Mazury, Olsztyn, Poland

(Received: 24 April 2018; accepted: 23 December 2018)

ABSTRACT The study investigates the problem of uncorrelated noise variance in least-squares spatial prediction of geophysical phenomena. This issue is commonly solved by the regularisation or variance-component estimation, however, the attempts at physical explanations of the issue are scarce or very cautious. In this study, an equivalent procedure to the regularisation is performed numerically, but the meaning of noise variance level is explained, and the source of the noise is more carefully examined. The presented numerical test uses spatial prediction technique i.e. least-squares collocation and reveals its relationships with the spectral signal properties. The numerical test is based on terrestrial Bouguer anomalies, which have a large variance at higher signal frequencies, i.e. their power spectral density decreases slowly when the spatial resolution increases. The same quantity is calculated from the EGM2008 model using various maximum degrees of the harmonic expansion. The different degrees of applied harmonics remove some spectral part of the signal from the data, which leads to the observations of the suspected relation of the estimated values of noise variance with medium and high-frequency signal parts. The observed statistical quantities prove that the noise has some relations with signal spectral range and data spatial resolution. The paper provides a relevant proof that the noise is not solely dependent on the survey error and introduces some physical interpretations of the regularisation requirement.

Key words: sampling, a priori noise, regularisation, LSC, simple kriging.

1. Introduction

Least-squares collocation (LSC) is a spatial technique of data modelling, which uses a composed signal spectrum. Typically, physical fields have a multi-spectral signal, because the phenomena generating them act from various distances and directions and with different magnitudes. Therefore, geophysical signals can be represented by the series in the frequency domain (e.g. spherical harmonics or Fourier series). All these parts compose the signal, which is received by the sensors together with noise. The local residual data sets that are often applied in LSC are also multispectral, and two of their spectral bands can be found at once: a lower band, depending on the data area range, and an upper one, depending on the spatial resolution and measurement accuracy. The lower bound is usually determined by the detrending process, i.e. the

removal of some lower-frequency signal, and it is related with the size of the local data area. The lower-order trend is composed of the lowest spherical harmonics of the signal and can be derived, e.g. from any global model like spherical harmonic expansion of geopotential. The upper spectral bound of the local data will be discussed in this article with special attention.

LSC interpolation, characterised by a similarity to simple kriging (SK) (Dermanis, 1984) assumes known expectation of the stochastic field. Satellite measurements have provided many global models of the phenomena occurring in the oceans, on the Earth and in the atmosphere, which creates an opportunity to apply a very accurate long-wavelength expected values of gravity, magnetic field intensity, ocean topography, total electron content in the ionosphere, etc. The data can be reduced, which provides a substantial advantage: one has to estimate and apply a local correlation of the data, which has a smaller range and reduces the number of the observations applied in the computations. Nevertheless, we never work with the pure signal only; there are no errorless observations, because there is no errorless measuring equipment. The general assumption about the lack of spatial correlation of noise has been made in the article. This choice is made due to the simplicity of the selected option, which is very helpful in the actual recognition of the problem. The author decides to avoid literature examples with non-correlated noise and correlated noise cases, because it is obvious that both types of noise are comparably frequent in the literature and this depends for instance on measuring equipment or kind of the observational technique. Nevertheless, it is easy to notice in the following paragraphs that most of the cited literature refers to uncorrelated noise, as it is a more convenient and universal option. This is especially true when we meet lack of a more precise information about the noise (Gilardoni *et al*., 2013). An article about high-frequency random errors, however, is worth reference here, as it also suggests a need for a more insightful investigation of the gravity data noise. Saleh *et al*. (2013) analyse high-frequency gravity errors of U.S. gravity database mainly by crossover comparisons and obtain values at the level of a few milliGals, which is much more than the measurement noise of the static gravimeters. These errors can originate from different sources and the present work attempts to study medium and high-frequency signals and data spacing as some of the probable reasons, which are jointly called a second presumption in this paper.

The paper investigates two possible sources of data noise, i.e. measurement error and high frequency signal corresponding to the resolution higher than data spatial resolution. The former can be found as an old and well-established norm in the cited works, whereas the latter one occurs in the literature as uncertain and informal principle, which in practice helps, however, to decrease the estimation errors. Some researchers notice problems related to the application of the measurement error in the noise covariance matrix and apply an additional manipulation to the diagonal, calling this process a regularisation (Marchenko *et al*., 2001). They then explain the regularisation from the mathematical point of view, referring it to matrix algebra problems and the measurement noise. The question raised in this article concerns possible relationships of the regularisation with physical data properties, i.e. such high-frequency signals, which can contribute to the noise directly.

The first presumption on the possible dependency of the noise and the measuring errors is very popular amongst geodetic and geophysical community and can be easily found in the literature dealing with least-squares modelling of different data types. Denker (1998) analyses GPS/levelling geoid modelling by LSC with the application of the noise based on survey documentation. The same type of data is analysed by Kavzoglu and Saka (2005), who consider noise covariance matrix

as a matrix of measurement errors. Bouman (1997), although investigates an application of the regularisation in the development of global geopotential models, also perceives the measurement noise as the principal physical contributor to noise covariance. Vergos *et al*. (2005) also indicate a dependency between the quality of the data and the noise covariance matrix in LSC and assign a presumed error value to the gravity data, due to the limited knowledge of the measurement error. El-Fiky *et al*. (1997) state that vector of the noise directly represents measuring error in the LSC modelling of crustal movement rates from the levelling data. Egli *et al*. (2007) apply LSC to the interpolation of crustal deformation and also incorporate observational error in the noise matrix, as the main component. Darbeheshti and Featherstone (2010) similarly associate data noise with the accuracy of surveying techniques in relation to GPS and geometric levelling.

The second presumption has become an objective of this study and refers to the relation between the noise covariance matrix and signal at frequencies higher than data resolution. At the beginning, however, let us refer to some existing works, which, as opposed to those listed in the previous paragraph, find some other factors contributing to the noise covariance matrix. An extended study on the relationship between the noise, regularisation and data spacing can be found in Rummel *et al*. (1979). The data of different resolutions have been tested in the numerical test and the conclusion has been drawn, that the stability of downward continuation solutions depends on the number and geographical distribution of the data, aside from the observational error. Xu and Rummel (1994) continue investigations on the regularisation problems related to gravity and find data sampling as a factor seriously affecting the reliability and accuracy of least-squares solutions. They also see some incompleteness in the description of the influence of regularisation on the error of the solution and its physical meaning. The spatial resolution, as a factor affecting the accuracy of the results is also discussed in Eshagh and Sjöberg (2011). The same problem has been investigated using different data resolutions by the author in Jarmołowski (2016).

Klees *et al*. (2004) analyse satellite gradiometry data and existing problems in least-squares solutions based on these data. They provide some suspected spatial factors affecting the covariance matrices of the signal and noise. Aside from the problem of the coloured noise, they mention data gaps due to the orbit geometry. Such gaps make differences in the spatial resolution, which can be large in case of satellite data, especially at lower latitudes. Andersen and Knudsen (1998) investigate marine gravity field from altimetry and apply noise variances larger than the observational errors. They also introduce additional covariance function for the high-frequency noise along the satellite tracks. This function is also related with different data resolutions in along-track and cross-track directions. Kim *et al*. (2008) also investigate altimetry data and apply filtering, pointing out its direct relation with the spacing between the profiles. The studies on altimetry data filtering showing its close relation with along-track and cross-track resolution can be also found in Sandwell (1992) and Sandwell and Smith (1997). Jarmołowski and Łukasiak (2016) estimate the across-track and along-track noise characteristics by maximum likelihood (ML) method and confirm the observations from the abovementioned papers. Some research on the correlation between the data spacing choice and LSC solution accuracy is presented in Lee *et al*. (2013). Their work refers to the geoid fitting problems, where GNSS/levelling points have often different spacing than the gravimetric geoid model resolution.

Other works go further in the studies of relations between the noise and data spacing. The spacing along the satellite tracks is used for the estimation of the standard error by Paolo and

Molina (2010). They provide an approximate formula linking altimeter noise level with along-track spacing in the computation of *a priori* error. Another example of similar formulae for the noise approximation can be found in Hwang and Parsons (1995). A comprehensive study on gravity data resolution is provided by Filmer *et al*. (2013), who discuss the problem of sampling in the case of interpolation. The prediction of the gravity at the benchmarks, necessary for physical height determination, can be inaccurate due to the signal omission. A non-representative sampling and related omission of higher-frequency signal can be especially easily detected in moderate mountains and can be really significant, as stated by Filmer *et al*. (2013). This point of view is close to the hypothesis of the presented study, which presents the influence of the mentioned signal omission on the noise variance. This research provides some observations on the contribution of the signal components to the noise variance, aside from the contribution from the survey accuracy.

The problem of data sampling effect on the parameter representing variance of the noise can be also found in the works that refer to the ordinary kriging (OK); although it belongs to the same family of parametric geostatistical tools, it differs in relation to LSC and SK due to some specific assumptions related to the expected value of the stochastic field. In this case the parameter responsible for the noise is called nugget. We can also find the studies that associate the nugget with the resolution issues. Tranchant and Vincent (2000) assign uncorrelated noise to the nugget parameter in the interpolation of ozone from the satellite data. They interpret the nugget as a sampling error at points. Even more questions regarding the nugget are posed issue in Clark (2010). That paper discusses inconsistency in the treatment of the nugget parameter in the literature and existing software. It also indicates the factors contributing to the nugget, like sampling and short-scale variability of the analysed phenomenon. The short-scale variability can be also considered as high-order frequencies in terms of spatial spectral decomposition, and this will be an objective of special focus in the numerical part of this article.

A comprehensive study of the geophysical signal referred to the data sampling and accuracy of the modelling is given by Vassiliou (1988). The power spectral density functions of the continuous gravity signal, and also of selected data samples with different grid spacings, are compared. Next the conclusions on the aliasing effects related to the data resolution decrease in gravity field approximation have been drawn.

The numerical part in the current article illustrates the issue investigated by Vassiliou (1988) from the practical point of view, which is possible due to the wide range of spherical harmonic degrees of the EGM2008 model and availability of noisy point gravity data from the same area. The terrestrial gravity data have been selected in such a way that it represents a spatial resolution corresponding to the middle degree between the lowest EGM2008 trend degree applied for the selected data area and the highest used degree of this global model. The terrestrial gravity data set and nine surfaces of various trend degrees based on the EGM2008 are then applied in two numerical tests, prepared for the covariance parameters analysis. In the first test, successively increasing harmonic degrees of the signal trend have been subtracted in sequence, starting from really low trend order and finally exceeding corresponding data resolution and removing all the correlated signal from the data. In the second test the lower order trend was set to a reasonable constant value in order to maintain general LSC rule related to the zero mean. Then, the ranges of the highest spherical harmonic degrees were removed three times, keeping always some correlated signal bandwidth in the residuals. These two tests are designed in order to analyse the influence of

lower and upper frequency cutoffs on the covariance parameters, particularly on the uncorrelated *a priori* noise size.

Three estimators are analysed and compared in order to assess noise variance size. The first is uncorrelated average *a priori* noise standard deviation (σ), indicated on the σ-axis by the minimum standard deviation (SD) of data-prediction differences in Leave-One-Out (LOO) validation (MSDLOO, Figs. 3 and 6, Table 3), the second is just this MSDLOO (minimum of contours in Figs. 3 and 6, Table 3) and the last one is *a posteriori* estimate of the prediction error m_p (Eq. 8, Table 3). The third estimator is strongly dependent on σ, and since σ-values are thoroughly analysed here, the discussion on the *a posteriori* error in this article is limited to the comparison with remaining two parameters (Table 3). A special focus is put on the first two estimators: σ (Eq. 6) and MSDLOO (Eq. 7) and their relations with the subtracted lowest and highest frequencies/degrees of the signal. The conclusions explain the relation of commonly applied regularisation with data spatial resolution to a considerable extent.

2. Lower-order signal, detrending, parametrisation and cross-validation

The Gauss-Markov model is a very popular statistical way for describing the physical phenomena such as magnetic anomalies, gravity, and many others. It assumes a split of the signal into deterministic and stochastic parts. The former one is often represented by mathematically determined polynomial. The global model based on the high-order spherical harmonics is a better approximation and, therefore, is applied in this study. The latter part of Gauss-Markov signal split is the stochastic part, which is here interpolated by the LSC. Thus, the data in this article is divided into two parts:

$$\Delta g = \Delta g^{EGM} + \Delta g^{r} \tag{1}$$

where Δg^{EGM} represents spherical harmonic synthesis of the Bouguer gravity anomalies of different maximum degrees/orders and Δg^{r} includes the noise together with the residual part of the signal, which will be interpolated by LSC. The long-wavelength part of the signal is generated in the numerical study by spherical harmonic synthesis (Pavlis *et al*., 2012), using various maximum degrees/orders (*nmax* = *mmax*), i.e.

$$\Delta g^{EGM}(r,\theta,\lambda) = \frac{GM}{r^2}\sum_{n=2}^{nmax}(n-1)\left(\frac{a}{r}\right)^n \sum_{m=-n}^{n}\bar{C}^{s}_{nm}\bar{Y}_{nm}(\theta,\lambda) \tag{2}$$

where $\bar{Y}_{nm}(\theta,\lambda)$ denotes fully-normalised spherical harmonic of two spherical variables: polar angle θ and longitude λ and uses fully-normalised Legendre function i.e.

$$\bar{Y}_{nm}(\theta,\lambda) = \bar{P}_{n|m|}(cos\theta)\bullet\begin{Bmatrix} cosm\lambda \\ sin|m|\lambda \end{Bmatrix}\begin{matrix} if\, m \geq 0 \\ if\, m < 0 \end{matrix} \tag{3}$$

The set of fully-normalised coefficients of EGM2008 model is used to generate a data trend. Increasing degrees/orders of the trends provides more and more detailed approximation of the gravity, which at some degree reaches the same level of resolution as the terrestrial data. After

detrending, the residuals should follow Gaussian distribution. This is true when we deal with enough level of stationarity and a sufficiently large maximum degree/order of EGM2008 spherical harmonic expansion is applied. Then, the residuals can be interpolated in the positions of data points by LSC:

$$\tilde{\Delta g}^r = C_P{}^T \bullet \left(C_s + C_n\right)^{-1} \bullet \Delta g^r \tag{4}$$

where Δg^r is the residual data vector. The matrix C_s is the covariance matrix of data residuals, C_P is the covariance matrix between predicted residuals and data residuals and C_n represents the noise covariance matrix. The signal is spatially correlated and C_s, as well as C_P, are generated using Gauss-Markov third order planar model (Moritz, 1978). This model is selected as an example planar model with two parameters, however, signal covariance can be well approximated by the other planar models (El-Fiky *et al*., 1997; Andersen and Knudsen, 1998; Darbeheshti and Featherstone, 2010) or by spherical reproducing models (Vergos *et al*., 2005; Gilardoni *et al*., 2013; Jarmołowski, 2016). Every planar model needs specific values of the parameters in order to fit the empirical covariance, however it is proven in Jarmołowski (2013) that LOO estimation of uncorrelated noise is comparably effective with different planar covariance models. Third order Gauss-Markov planar model reads:

$$C_s\left(C_0, CL, \psi\right) = GM3\left(C_0, CL, \psi\right) = C_0\left(1 + \frac{\psi}{CL} + \frac{\psi^2}{3CL}\right) \bullet exp\left(-\frac{\psi}{CL}\right) \tag{5}$$

where C_0 denotes residual signal variance, spherical distance ψ is used as the variable and correlation length (*CL)* is also in spherical distance units. The noise is assumed to be non-correlated. Thus, since we consider uncorrelated noise, C_n becomes diagonal matrix (σ is homogeneous, average noise SD):

$$C_n = \sigma^2 \bullet I_n \,. \tag{6}$$

LSC described by Eq. 4 is quite common in regular gridding of gravity anomalies. It can be used in a regular grid over the data area, but also in sparse points. The latter option can be used in cross-validation process, which can be found in the literature in different forms. One of them is LOO validation, which removes one point from the data set being used for the prediction in the position of the removed one (Kohavi, 1995). The differences between *n* data values and the predictions made in the same positions are often used as a measure of the prediction precision. The difference in point P used in LOO validation reads:

$$LOO_P\left(C_0, CL, \sigma \middle| \Delta g^r_{n\times 1}\right) = \left\{ \begin{array}{c} \Delta g^r_P - \tilde{\Delta} g^r_P \Big| \\ \tilde{\Delta} g^r_P = C^T_{P_{(n-1)\times 1}} \bullet \left(Cs_{(n-1)\times(n-1)} + Cn_{(n-1)\times(n-1)}\right)^{-1} \bullet \Delta g^r_{(n-1)\times 1} \wedge \\ \Delta g^r_P \notin \Delta g^r_{(n-1)\times 1} \end{array} \right\} \tag{7}$$

The SD of all *n* differences calculated by Eq. 7 is a measure of the quality of the estimated model. SD of LOO is affected by the noise present in the data set, as it comes from the comparison with noisy data. Therefore, if the prediction is optimal in least-squares sense, SD of LOO can be assessed as an empirical measure of the noise. In the proposed study LOO differences are

calculated for some range of parameters *CL* and σ and for different residuals Δg^r determined by the subtraction of the gravity from EGM2008 model for various *nmax*. An additional measure of the noise in LSC is *a posteriori* error, estimated here by the formula:

$$m_P^2 = C_0 - C_P{}^T \bullet \left(C_s + C_n\right)^{-1} \bullet C_P. \tag{8}$$

3. Data

The presented numerical test applies Bouguer gravity anomalies downloaded from the U.S. gravity database available at the website of University of Texas at El Paso (Hildenbrand *et al*., 2002; Saleh *et al*., 2013) and Bouguer gravity anomalies calculated with various maximum degree/order from the EGM2008 geopotential model (Pavlis *et al*., 2012). Gravity anomalies have a significant signal variance at higher degrees of harmonic expansion in contrast to geoid, where predominant part of the signal is cumulated at lower degrees. The LSC, as a spatial technique, employs a range of signal frequencies that compose the signal. The lower limit of the residual data spectrum is always approximately defined by the subtracted trend surface. The second presumption in this paper is that the upper limit is related with the spatial resolution and, therefore, data sampling plays a crucial role in the experiment. It is obvious that the trend has to ensure the assumed distribution of the residuals, i.e. normal distribution; however, this condition depends on the maximum degree of the trend spectrum, which must be sufficiently high in order to remove all long-wavelength signal part.

The horizontal data resolution is approximately homogeneous and approximately equal 0.1°×0.1° (Figs. 1a and 2a). The original data resolution was higher, but it was decreased for the purpose of the designed experiment. The selected density corresponds to the maximum harmonic expansion degree, *nmax* = 1800, and will have a specific importance in the numerical test. The EGM trends start from *nmax* = 40 and realise maximum degrees equal to 80, 180, 360, 540, 720, 1080, 1440 and 2160, respectively. Figs. 2a to 2c show really low degrees, which can be insufficient to meet the fundamental LSC requirement, i.e. that the expected value of the stochastic field should be close to zero. Figs. 2d to 2f show more reasonable degrees of detrending, whereas Figs. 2g to 2i describe signal range, which is very close to all of the correlated signal available in terrestrial gravity data. The variable degrees/orders of the EGM2008 trends start from much lower degree than the threshold value of 1800 and end at the highest degree of EGM2008 (*nmax* = 2160). This enables to observe how the trend removes consecutive signal frequencies from those included in the data and what happens when the trend degree approaches the degree resulting from the data resolution. This will also indicate how the decrease of the signal spectrum affects covariance parameters and accuracy of the LSC solution.

Fig. 1b describes the highest degree/order of EGM2008 grid generated from its coefficients, namely *nmax* = 2160, overlaid with data points. The subtraction of the model calculated with *nmax* = 2160 removes all of the correlated signal from the data set, which will be observable in the numerical part of the paper. Therefore, the data apparently coincide with the EGM2008 surface in Fig. 1b. Fig. 1b confirms the absence of individual outlying values in the data set, because all the large peaks of the gravity anomaly surface are modelled from a few close point values. These peaks denoting high-order signal frequencies remain in the residuals in case of relatively low-

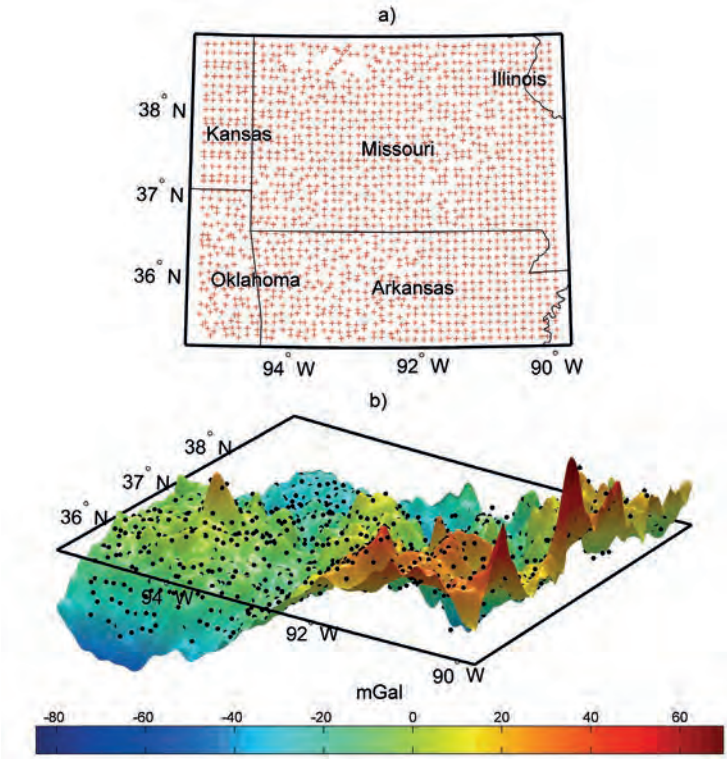

Fig. 1 - a) Location of selected Bouguer gravity data; b) data points together with model of maximum degree from EGM2008 (*nmax* = 2160). Data occurring in groups over peaks prove that no outliers can be suspected there.

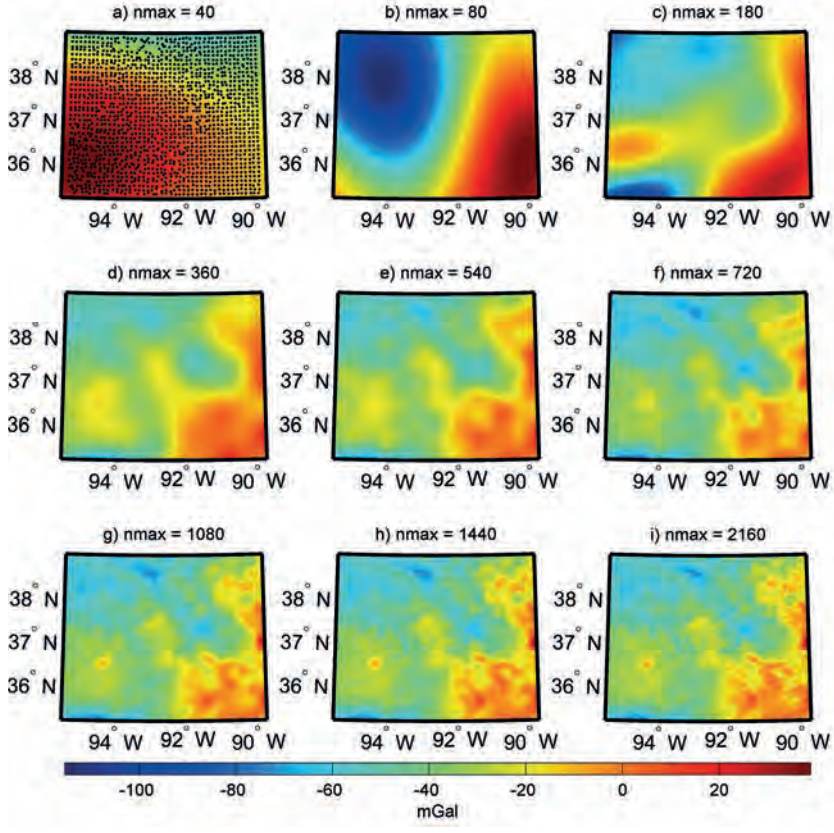

Fig. 2 - Bouguer gravity anomalies calculated with different maximum degree/order from EGM2008 and used for data detrending. Terrestrial gravity data locations used in test shown in panel a.

order trend subtraction. However, they can be removed from the data with higher order EGM2008 trends, if the harmonic degrees composing the trend include these small gravity details.

Table 1 describes statistics regarding the actual data and different degrees of the trend. The statistics, and first of all the means of the lowest degrees/orders, absolutely disagree with the data. It is obvious, therefore, that *nmax* = 40 and 80 cannot detrend the data to the level necessary in LSC for this data sample. Looking at the means of higher degrees, we can observe an apparent systematic difference between the data mean and means of EGM2008 solutions equal around 0.60 mGal. This difference can result from high-order gravity details, which can be different in point observations in comparison to EGM2008 trends, due to different cutoff degrees of the signal. The local sampling of the gravity can also contribute to this difference, as well as other factors that can result e.g. from the EGM2008 modelling process. The investigation of the bias is skipped in this study, as it becomes smaller after detrending and the residuals have mean around only 0.25 mGal, which is one order smaller than the investigated noise estimators in this study (Table 2). It is possible that some factors contributing to the bias are common for point data and EGM2008 and have been removed by detrending.

Table 1 - Statistics of data and different orders of EGM2008 trend.

Degree of detrending	Data (mGal)				Trend from EGM2008 (mGal)			
	Min.	Max.	Mean	St.Dev.	Min.	Max.	Mean	St.Dev.
40	-72.01	8.49	-34.79	16.84	-42.88	-0.14	-19.12	11.47
80					-71.06	1.19	-37.45	21.74
180					-77.94	1.84	-35.63	16.29
360					-67.23	4.64	-35.40	16.15
540					-68.85	4.13	-35.40	16.27
720					-69.75	5.15	-35.39	16.39
1080					-71.49	5.98	-35.41	16.50
1440					-73.92	8.48	-35.42	16.58
2160					-73.55	10.31	-35.41	16.64

4. The study of the noise with the use of different residual gravity signals

As it was introduced in sections 2 and 3, the study starts from the detrending of terrestrial Bouguer gravity data with the use of nine different models of Bouguer gravity anomalies generated from EGM2008 harmonic coefficients. The succeeding degrees of gravity trends generated from the EGM2008 contain increasingly more higher-order signal, which after subtraction, produces decreasing residuals that successively improve their match with the normal distribution. Fig. 3 describes the results of LOO validation for two covariance parameters: correlation length (*CL*) related to the signal covariance and average uncorrelated noise SD (σ). The variance of the signal C_0 (Eq. 5) is set to the data variance in the calculations, which limits the number of estimated parameters to the abovementioned two (Eqs. 4 to 6). The contours of LOO SD in Fig. 3 show LOO validation and MSDLOO indicates proper parameters on *CL* and σ axes, for optimal LSC prediction. Fig. 3f shows LOO for the last case of the residuals, i.e. the largest *nmax* for which

the covariance parameters can be estimated. This level of detrending is based on *nmax* = 720, which corresponds to 0.25° in horizontal resolution. The next trend is calculated for *nmax* = 1080, adequate to the resolution which approximately equals 0.166°, which is close to the resolution of the point data equaling 0.1°. This means that *nmax* = 1080 removes those signal frequencies that could be represented by 0.166°×0.166° grid, and the amount of the remaining signal in the residuals becomes very small, which significantly impedes the estimation of the parameters. Thus, the test with the use of LOO (Eq. 7) allows to observe six cases when the detrended signal remains correlated and signal covariance can be efficiently parametrised (Figs. 3a to 3f). In the remaining, opposite three cases for three highest maximum degrees of the EGM2008 trend, we observe no estimable correlation parameters of the residuals (Figs. 3g to 3i). The MSDLOO in Figs. 3g to 3i can be found on the plot edges and it is clear that unreasonably large values of *CL* and σ are simply false. After LOO, the LSC solutions based on the best estimated *CL*'s and σ's are selected to show a representative statistical quantities (Tables 2 and 3), except for the cases from Figs. 3g to 3i, where the last estimable parameters shown in Fig. 3f are preserved in order to compute statistical quantities. All the applied parameters are indicated by the white squares in Fig. 3.

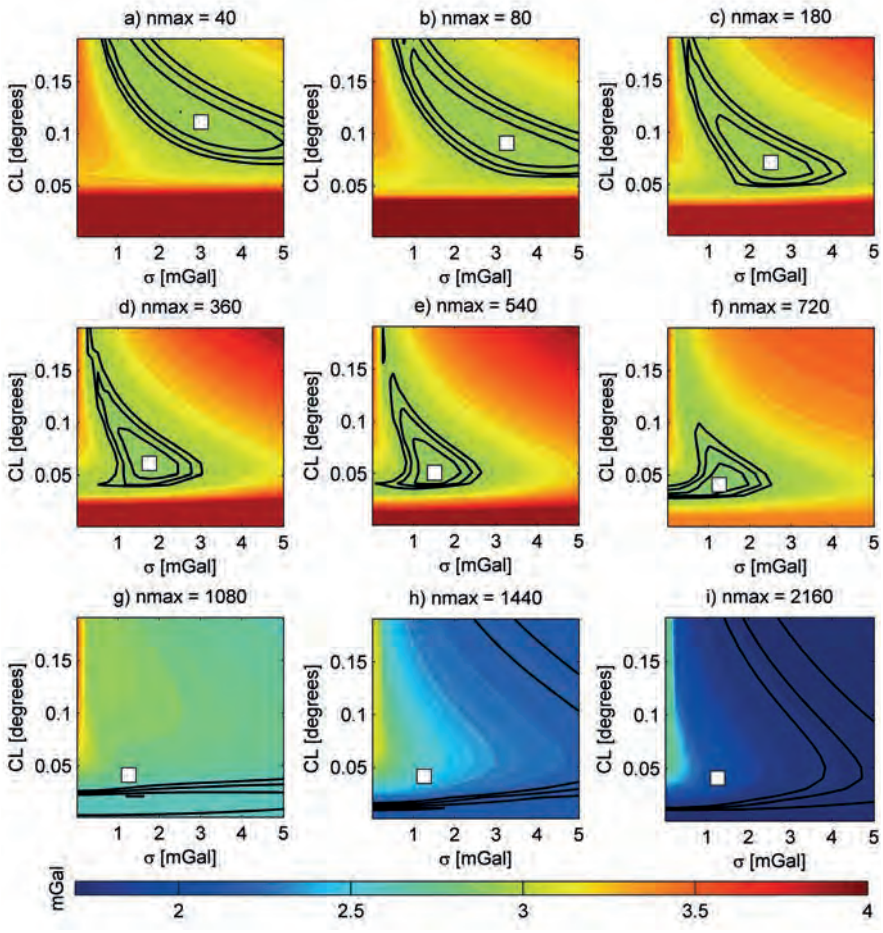

Fig. 3 - SD of LOO differences calculated for varying *CL* and σ and for different detrending degrees/orders. White square indicates MSDLOO (panels a to f) or the same MSDLOO as for panel f when no reasonable minima can be found (panels g to i).

The mean of LOO differences in all cases of trends is close to zero and therefore SD can be equivalent dispersion parameter of these differences to RMS. In general, the empirical estimate of noise, i.e. MSDLOO, remains almost unchanged for the residuals analysed in Figs. 3a to 3f and equal to around 2.95 mGal despite slightly decreasing σ (Table 3). A small increase of MSDLOO is noticeable only for *nmax* = 720 (Table 3), which can be related with small SD of the residuals approaching the noise SD and related signal aliasing. MSDLOO decreases rapidly only after highest-degree trends removal starting from Fig. 3g (Table 3), i.e. when the spatial resolution of EGM2008-based trend model reaches and exceeds the data resolution. The degrees of the trend in the cases applied for LOO in Figs. 3e to 3f provide the residuals having SD equal to 4.25 mGal and 3.45 mGal, respectively (Table 2), which is getting closer and closer to MSDLOO (Table 3). The predicted amounts of the signals in these sets have SD's equal to 3.26 mGal and 2.01 mGal, respectively (Table 2), which is close to or even below the noise approximated by MSDLOO. Therefore, in practice, one should select a degree of the trend that provides normal distribution of the residuals and preserves optimal range of the frequencies for the correct parametrisation of the LSC. The optimum trend avoiding non-Gaussian or aliasing signal in the analysed residuals, probably exists somewhere between *nmax* = 360 (Fig. 3d) and *nmax* = 540 (Fig. 3e).

Table 2 - Statistics of residuals at points, and predictions with parameters inferred from MSDLOO, and *CL* estimated by LOO.

Degree of detrending *(nmax)*	Residual gravity (mGal)				Predicted gravity (mGal)				Estimated *CL* (degrees)
	Min.	Max.	Mean	St.Dev.	Min.	Max.	Mean	St.Dev.	
40	63.61	33.28	-15.61	19.93	62.60	29.67	15.60	19.76	0.12
80	43.94	44.04	2.23	15.78	42.51	42.57	2.24	15.52	0.10
180	27.42	30.44	0.55	8.42	25.65	26.55	0.54	7.97	0.08
360	19.36	32.01	0.28	5.34	18.79	19.36	0.27	4.43	0.06
540	16.43	23.50	0.27	4.25	11.98	14.52	0.26	3.26	0.06
720	13.84	19.72	0.24	3.45	-7.27	10.87	0.21	2.01	0.04
1080	11.20	14.64	0.26	2.62	-4.47	6.93	0.23	1.13	-
1440	10.14	13.19	0.29	2.13	-4.13	5.54	0.24	0.85	-
2160	-8.80	9.42	0.26	1.70	-3.76	2.92	0.20	0.67	-

The parameters C_0 and *CL* always decrease with an increasing order of the subtracted polynomial trend, as the variance of the residuals decreases. This is also the case for the EGM2008 trends in this study. The noise parameter σ, that represents uncorrelated parts of the data, decreases slightly with growing trend degree, as well as *CL*, which can be found in Figs. 3a to 3f. Figs. 3a and 3b depicts LOO contours for the residuals detrended improperly, as the histograms of residuals in Figs. 4a and 4b are non-Gaussian at all. The residuals in Fig. 4c can be assessed as approaching the normal distribution, but still slightly biased. The estimated σ in Fig. 3c equals to 2.50 mGal (Table 3), which is close to MSDLOO (around 2.90 mGal). Figs. 4d to 4f show a relatively good normal distribution. The corresponding minima of the contours in Figs. 3d to 3f indicate σ values decreasing from 1.75 mGal to 1.25 mGal, which is twice lower than corresponding MSDLOO. The unexpected change of σ can be related with an influence of some signal parts from EGM2008

Table 3 - Statistics of estimated errors at points, and LOO differences at points (used parameters inferred from MSDLOO), and σ estimated by LOO.

Degree of detrending	A posteriori errors (mGal)				LOO differences (mGal)				Estimated σ (mGal)
	Min.	Max.	Mean	St.Dev.	Min.	Max.	Mean	St.Dev. (MSDLOO)	
40	2.28	5.79	3.25	0.40	-17.81	21.00	-0.01	2.95	3.00
80	2.54	6.30	3.70	0.42	-17.47	20.94	-0.00	2.94	3.25
180	2.02	4.72	3.00	0.31	-17.24	20.96	0.00	2.92	2.50
360	1.51	3.50	2.32	0.23	-17.20	21.05	0.01	2.91	1.75
540	1.42	3.25	2.28	0.21	-17.05	20.66	0.01	2.91	1.50
720	1.39	3.01	2.31	0.19	-14.86	19.82	0.13	3.09	1.25
1080	1.98	2.61	2.53	0.07	-12.82	16.64	0.14	2.61	-
1440	2.03	2.13	2.13	0.01	-10.15	15.52	0.17	2.25	-
2160	0.90	1.25	0.94	0.06	-8.51	10.68	0.16	1.85	-

due to the following reasons: Figs. 2d to 2f, which correspond to Figs. 3d to 3f, show some interesting anomalies of the gravity field that are spatially small and cannot be observed in the previous Fig. 2c. These details are really local in relation to the data resolution. Even if these local anomalies extend to the area of a few data points, their correlation parameters differ from the average parameters that we estimate with LOO and can be hard to handle from 0.1° data resolution. It is known that even 3-4 points are insufficient to assess their spatial covariance by any method as it is statistically an inefficient sample. Therefore, these anomalies are suspected to contribute to σ, rather than to the signal. To summarise, in Fig. 3 we can find:

- LOO validation used in the search of *CL* and σ parameters for nine degrees of EGM2008 used in the detrending;
- six cases of efficient LOO parametrisation (parameters in Tables 2 and 3), where the removed trend preserves some correlated signal in the residuals, and three cases when no correlation can be found after detrending;
- easily explainable decrease of *CL* with the decrease of *nmax* and more difficult to explain decrease of σ, which is suspected to be related with the single peaks in gravity anomalies in Fig. 1b. These peaks are part of the signal, however, can contribute to σ, as they are spatially small and their covariance parameters differ from the average covariance parameters for the area. Their influence on σ estimation decreases (Figs. 3d to 3f), because we remove them with the higher EGM2008 degrees (Figs. 2d to 2f).

Fig. 4 describes the histograms of residual data referred to all nine degrees of the trend and LOO differences based on these residuals. Figs. 4a and 4b reveal the distribution of the residuals significantly different from Gaussian, which indicates that trends up to degree *nmax* = 80 insufficiently detrend Bouguer gravity data in the area of size 8°×11°. Figs. 4c to 4f show continuously decreasing variance of the residuals. Figs. 4g to 4i represent residual data and LOO differences with the use of the parameters estimated for the residuals at *nmax* = 720, which is intentional due to the unsuccessful parametrisation for *nmax* > 720 (Figs. 3g to 3i). The residuals, after detrending with the last three EGM2008 trends (Figs. 4g to 4i, Table 2), have small SD, at the level of MSDLOO (Table 3). The predictions from these residuals with the parameters inferred from Fig. 3f, indicated by the white squares in Figs. 3g to 3i (*CL* = 0.04° and σ = 1.25

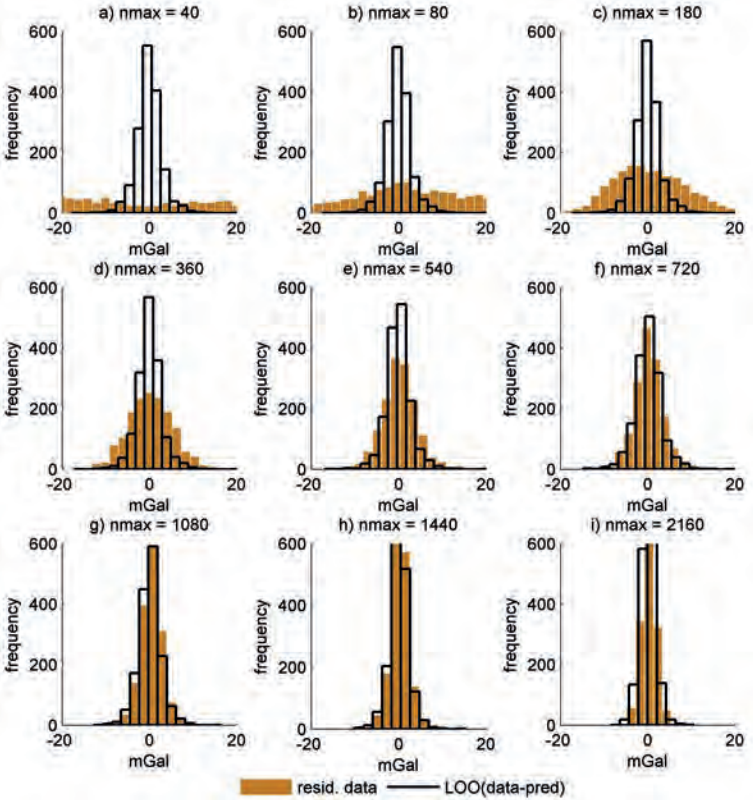

Fig. 4 - Histograms of residual data and LOO differences for different degrees/orders of trend.

mGal, Tables 2 and 3) have SD even below MSDLOO, which may indicate that these predictions contribute little to LOO SD.

Fig. 5 presents the histograms of LOO differences and histograms of the predictions at points. LOO validation assumes a comparison between noisy data and the predicted signal, knowing that the noise contributes to LOO differences. A relatively small number of larger values of LOO differences reaching 20 mGal can be observed in the black histograms in Figs. 5a to 5d. This number decreases in Figs. 5e and 5f and starts to disappear from Fig. 5g becoming lost in Fig. 5i. These large LOO differences and their removal have a confirmation in extremum values of LOO in Table 3, but it is easier to discuss them using a single individual case. The extrema of LOO differences in Table 3 (*nmax* = 720) are equal respectively to min. = -14.86 mGal and max. = 19.82 mGal. The extrema of the corresponding residual data (Table 2) are respectively min. = -13.84 mGal and max. = 19.72 mGal, which is quite close to LOO differences. The prediction however (Table 2), has smaller extrema, i.e. min. = -7.27 mGal and max. = 10.87 mGal, and this indicates that such large LOO discrepancies must be related with the data residuals. This similarity of LOO differences to the residuals and dissimilarity from the prediction, repeats for higher degrees of the trend. The residuals can include a contribution from at least three components: signal, noise, and outliers. The spherical harmonic model, however, cannot remove uncorrelated noise or outliers from these residuals and, therefore, the larger LOO values visible in Figs. 5a to 5f, which are lost in Figs. 5g to 5i, must be associated with the high-frequency signal. Additionally, MSDLOO for the last three trends in Table 3, coinciding with SD of the last three residual data sets in Table 2, proves that SD of the residuals determines directly MSDLOO if the residuals contain no

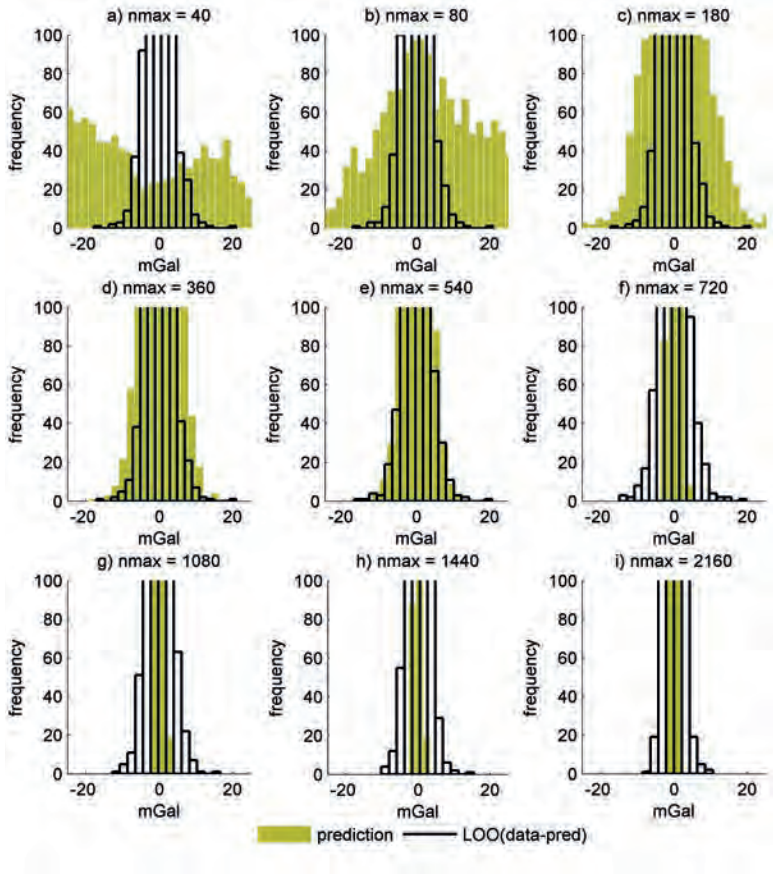

Fig. 5 - Histogram of prediction values at data points and LOO differences for different degrees/ orders of trend (zoom view to frequency = 100). The view is zoomed to reveal small numbers of large LOO differences.

correlated signal. To summarise, Figs. 4 and 5 show:

- non-normal distribution of the residuals for lower order trends;
- aliasing of the correlated residual signal in case of detrending with *nmax* > 720;
- single large LOO differences (e.g. Figs. 5d to 5f) that are removable by higher-degree EGM trend (see Figs. 5g to 5i) and, therefore, are suspected to be based on single signal peaks in the data (Fig. 1b). These single peaks can affect LOO and, therefore, contribute to σ estimation at lower degrees of detrending.

The detrending with the high-degree (*nmax* > 720) EGM2008 trend rapidly decreases MSDLOO, whereas it remains approximately constant up to *nmax* = 720 (Table 3). A similar behaviour can be observed for the mean of *a posteriori* errors (Table 3), which are also noise estimators. The LOO validation applies differences between the prediction and residuals including correlated signal and uncorrelated part, called usually noise. Therefore, MSDLOO cannot be smaller than uncorrelated data part. Thus, a noticeable decrease of MSDLOO when *nmax* >720 suggests that the increasing trend degrees remove high-degree gravity signal that contribute to the uncorrelated part. This can be observable in Figs. 3g to 3i, as the colours of MSDLOO changing from light green to blue. However, this dependence needs some confirmation, as the parametrisation of the signal fails for *nmax* >720 and LSC prediction is uncertain there. Thus, in order to confirm the relation of the highest signal degrees with σ, MSDLOO and *a posteriori* error, an additional test was designed. This test investigates a subtraction of the highest signal degrees, keeping some lower degrees of the signal at the same time. This test is started with the removal of low-order

EGM2008 gravity up to *nmax* = 180, in order to assure LSC unbiasedness condition related with the expected value of the residuals, which must tend to zero. Then, the highest frequencies are also removed, respectively: degrees between n = 1440 and n = 2160 (Fig. 6b), degrees between n = 1080 and n = 2160 (Fig. 6c), and degrees between n = 720 and n = 2160 (Fig. 6d). Fig. 6a repeats Fig. 3c for the comparison, as it shows the residuals after the removal of the lower-order gravity trend only (*nmax* = 180). The high-pass filtered case of residuals applied in Fig. 6b, shows, together with Fig. 6c, a continuous subtraction of the high-order signal. This removal decreases σ, but contrary to Figs. 3a to 3f, the change does not follow in the direction of decreasing *CL*. In this case, when the highest-frequency signal is removed from the residuals, the MSDLOO indicates increasing *CL* and slides in the NW direction (Figs. 6b and 6c). Fig. 6d, on the contrary, shows an increase of σ along σ axis. The suspected reason is a wide range of the high-order signal subtracted from the residuals. This range includes higher-order frequencies together with the ones at degrees lower than corresponding point data resolution. These subtracted lower frequencies produce the so-called "omission error", i.e. loss of the frequencies that could be theoretically interpolated as correlated. The additional effect is a general decrease of LOO SD (Figs. 6b and 6c), which is obvious if we know that LOO differences are dependent on the existing data noise and this noise is composed of the measurement errors, and also of the high-frequency signal, which was removable here. To conclude, the additional test in Fig. 6 reveals:

- a direct relation of the removed high-frequency signal with σ parameter, which decreases along with the loss of the signal highest degrees;
- a direct relation of the removed high-frequency signal with MSDLOO, which is an empirical estimator of the noise;
- an omission error, occurring when we remove a wide range of higher-order signal including lower-order correlated signal.

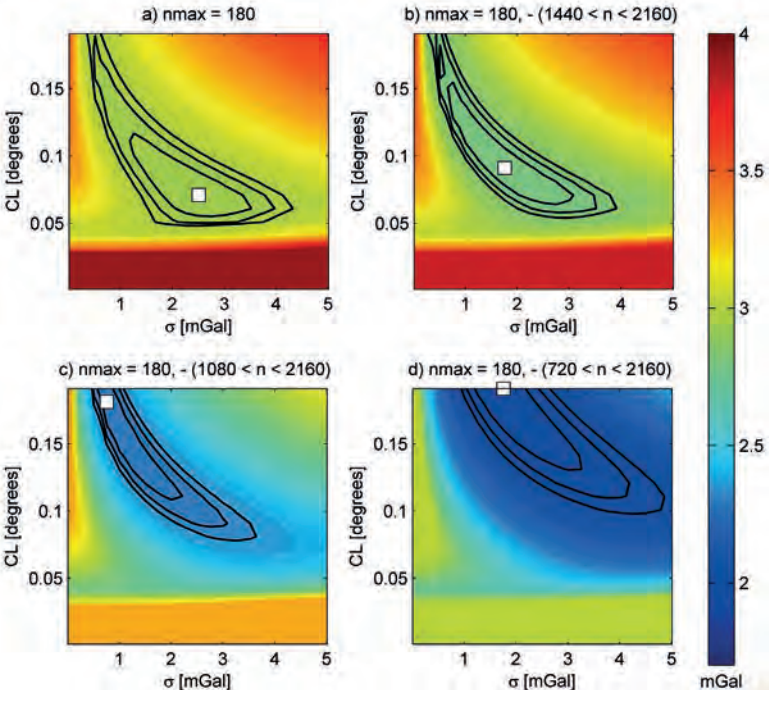

Fig. 6 - SD of LOO differences calculated for different *CL* and σ and for the same detrending degree/order = 180, but for different high-frequency cutoffs. White square indicates optimal parameters inferred from MSDLOO. The residuals, aside from detrending, have removed also highest part of gravity spectrum obtainable from EGM2008 (panels b to d), whereas panel a is repeated from Fig. 3c for comparison.

5. Conclusions

It is clear from Figs. 3a to 3f that two covariance parameters, *CL* of the signal covariance and σ representing *a priori* noise SD decrease and get closer to axes together with the increase of EGM2008 trend order. A decrease of σ is also observable when removing high degrees of the harmonic expansion in Figs. 6a to 6c. However, in the second case *CL* tends to increase. All panels in Figs. 3 and 6 present graphically two directions of σ decrease, where the one directed towards the frame origin is clearly related with the subtraction of lower degrees and the other, directed upwards the frame of Fig. 6 and more towards *CL* increase is related with high degrees removal ($nmax > 720$). It is, therefore, obvious that the manipulation of the signal band, from any side, affects σ estimate.

The variation of MSDLOO when we detrend with $nmax < 720$ (Table 3) is very small. The parameter σ in case of the test illustrated in Fig. 3 varies from the value comparable to MSDLOO from the successful parametrisations (around 2.90 mGal) to the half of its value (Table 3). It must be noted that LOO differences are differences of the predictions with assumed σ and data having its own noise, and the error propagation rules for the difference can primarily explain a larger MSDLOO in relation to σ. The other suspected factor that decreases σ is inferred from Figs. 2d to 2f. The gravity trends for medium degrees of the harmonic expansion, let us say $360 < nmax < 720$, are characterised by some small anomalies, which are very local in relation to data spatial distribution. In other words, it is easy to find a local peak in the gravity field, e.g. in Fig. 2e, which is extended over a few points of the data, which is too little to find the correlation in LSC process and in LOO validation. Therefore, these anomalies are suspected to contribute to σ.

LOO differences are strongly related with high-frequency signal parts, which can be observed in Figs. 5g to 5i, where the EGM2008 signal removed for various $nmax > 720$, removes large LOO differences. For $nmax > 720$ the signal is still removable and MSDLOO decreases rapidly (Figs. 3g to 3i). The additional proof has been also presented in Figs. 6b and 6c, where the subtraction of selected high-degree signals ($1440 < n = m < 2160$ and $1080 < n = m < 2160$) decreases MSDLOO as well. The decrease of σ, due to the removal of upper degrees higher than 720, was expected and is related with the signal at the frequencies higher than those corresponding to point data resolution (Figs. 6b and 6c). Fig. 6d illustrates the case where some more selected high frequency signal from the band between degrees 720 and 2160 is subtracted from the data, previously detrended with $nmax = 180$. The contours of LOO SD, as well as MSDLOO, move to the right, i.e. an increase of σ is observable contrary to the previous Figs. 6b and 6c. The maximum spherical harmonic degree of the reduced residual signal equals here 720 and corresponds to the resolution 0.25°, whereas average terrestrial data resolution is around 0.1°. This means that we have removed a substantial part of the high-order correlated signal from the residuals, namely that between the lower degree of the subtracted part equal to 720 and the degree corresponding to maximum data resolution. This way we limit upper degree of the correlated remaining residual signal to the degree 720. The data resolution and gravimeter precision provide originally more degrees of the signal than 720, and, therefore, we introduce the so-called "omission error", detected as an additional contribution to noise and an increase of σ in the LSC.

The high-degree part of the signal, which does not contribute to the prediction is not really random, but it is perceived as random by the spatial techniques like LSC or kriging, due to the limited data resolution. The physical field can be homogeneously interpolated only up to the

frequencies equivalent to the data resolution, even if more signal is available at points from the measurement instrument precision. A special case occurs when the survey precision is low. Then, it can limit the upper frequency of the signal even at degrees lower than those corresponding to data resolution and we obtain the same effect as "omission error" simulated here in Fig. 6d. This is rare in case of terrestrial gravity data due to the high accuracy of static gravimeters, but may happen for other data.

Acknowledgements. Gravity data originates 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). I am grateful to two anonymous reviewers for their valuable remarks and the improvement of the manuscript.

REFERENCES

Andersen O.B. and Knudsen P.; 1998: *Global marine gravity field from the ERS-1 and Geosat geodetic mission altimetry*. J. Geophys. Res.**, 103,** 8129-8137.

Bouman J.; 1997: *Quality assessment of geopotential models by means of redundancy decomposition*. DEOS Prog. Lett., **97**, 49-54.

Clark I.; 2010: *Statistics or geostatistics? Sampling error or nugget effect?* J. S. Afr. Inst. Min. Metall., **110**, 307-312.

Darbeheshti N. and Featherstone W.E.; 2010: *Tuning a gravimetric quasigeoid to GPS levelling by non-stationary least-squares collocation*. J. Geod., **84**, 419-431.

Denker H.; 1998: *Evaluation and improvement of the EGG97 quasigeoid model for Europe by GPS and leveling data*. In: Vermeer M. and Adam J. (eds), Proc. Continental Workshop on the Geoid in Europe, Budapest, Hungary, Reports of the Finnish Geodetic Institute, **98**, 53-61.

Dermanis A.; 1984: *Kriging and collocation - A comparison*. Manuscr. Geod., **9**, 159-167.

Egli R., Geiger A., Wiget A. and Kahle H.-G.; 2007: *A modified least-squares colocation method for the determination of crustal deformation: first results in the Swiss Alps*. Geophys. J. Int., **168**, 1-12.

El-Fiky G.S., Kato T. and Fujii Y.; 1997: *Distribution of vertical crustal movement rates in the Tohoku district, Japan, predicted by least-squares collocation*. J. Geod., **71**, 432-442.

Eshagh M. and Sjöberg L.E.; 2011: *Determination of gravity anomaly at sea level from inversion of satellite gravity gradiometric data*. J. Geodyn., **51**, 366-377

Filmer M.S., Hirt C. and Featherstone W.; 2013: *Error sources and data limitations for the prediction of surface gravity: a case study using benchmarks*. Stud. Geophys. Geod., **57**, 47-66.

Gilardoni M., Reguzzoni M. and Sampietro D.; 2013: *A least-squares collocation procedure to merge local geoids with the aid of satellite-only gravity models: the Italian/Swiss geoids case study*. Boll. Geof. Teor. Appl., **54**, 303-319.

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. and Webring M.; 2002: *Rationale and operational plan to upgrade the U.S. Gravity Database*. U.S. Geological Survey, Reston, VA, USA, Open-File Report 02-463, 12 pp.

Hwang C. and Parsons B.; 1995: *Gravity anomalies derived from Seasat, Geosat, ERS-1 and TOPEX/POSEIDON altimetry and ship gravity: a case-study over the Reykjanes Ridge*. Geophys. J. Int., **122**, 511-568.

Jarmołowski W.; 2013: *A priori noise and regularization in least squares collocation of gravity anomalies*. Geod. Cartography, **62**, 199-216.

Jarmołowski W.; 2016: *Estimation of gravity noise variance and signal covariance parameters in least squares collocation with considering data resolution*. Ann. Geophys., **59**, S0104.

Jarmołowski W. and Łukasiak J.; 2016: *A study on along-track and cross-track noise of altimetry data by maximum likelihood: Mars Orbiter Laser Altimetry (MOLA) example*. Artif. Satell. (J. Planet. Geod.), **50**, 143-155.

Kavzoglu T. and Saka M.H.; 2005: *Modelling local GPS/levelling geoid undulations using artificial neural networks*. J. Geod., **78**, 520-527.

Kim J.W., Roman D.R., Lee B.Y. and Kim Y.; 2008: *Altimetry enhanced free-air gravity anomalies in the high latitude region*. Terr. Atmos. Ocean. Sci., **19**, 111-116.

Klees R., Ditmar P. and Kusche J.; 2004: *Numerical techniques for large least-squares problems with applications to GOCE*. In: Sansò F. (ed), International Association of Geodesy Symposia, Proc. V Hotine-Marussi Symposium on Mathematical Geodesy, Matera, Italy, vol. 127, pp. 12-21.

Kohavi R.; 1995: *A study of cross-validation and bootstrap for accuracy estimation and model selection*. In: Melish C.S. (ed), Proc. 14th International Joint Conference on Artificial Intelligence, Montreal-Quebec, Canada, vol. 2, pp. 1137-1143, <ijcai.org/Past%20Proceedings/IJCAI-95-VOL2/PDF/016.pdf>.

Lee S., Choi C. and Kim J.; 2013: *Evaluating the suitability of the EGM2008 geopotential model for the Korean peninsula using parallel computing on a diskless cluster*. Comput. Geosci., **52**, 132-145.

Marchenko A.N., Barthelmes F., Meyer U. and Schwintzer P.; 2001: *Regional geoid determination: an application to airborne gravity data in the Skagerrak*. Deutsches GeoForschungsZentrum, Potsdam, Germany, Scientific Technical Report, STR01/07.

Moritz H.; 1978: *Least-squares collocation*. Rev. Geophys., **16**, 421-430.

Paolo F. and Molina E.; 2010: *Integrated marine gravity field in the Brazilian coast from altimeter-derived sea surface gradient and shipborne gravity*. J. Geodyn., **50**, 347-354.

Pavlis N.K., Holmes S.A., Kenyon S.C. and Factor J.F.; 2012: *The development and evaluation of Earth Gravitational Model (EGM2008)*. J. Geophys. Res., **117**, B04406, doi: 10.1029/2011JB008916.

Rummel R., Schwarz K.P. and Gerstl M.; 1979: *Least squares collocation and regularization*. Bull. Geod., **53**, 343-361.

Saleh J., Li X., Wang Y., Roman D.R. and Smith D.A.; 2013: *Error analysis of the NGS' surface gravity database*. J. Geod., **87**, 203-221.

Sandwell D.T.; 1992: *Antarctic marine gravity field from high-density satellite altimetry*. Geophys. J. Int., **109**, 437-448.

Sandwell D.T. and Smith W.H.F.; 1997: *Marine gravity anomaly from Geosat and ERS-1 satellite altimetry*. J. Geophys. Res., **102**, 10039-10054.

Tranchant B.J.S. and Vincent A.P.; 2000: *Statistical interpolation of ozone measurements from satellite data (TOMS, SBUV and SAGE II) using the kriging method*. Ann. Geophys., **18**, 666-678.

Vassiliou A.A.; 1988: *The computation of aliasing effects in local gravity field approximation*. Bull. Geod., **62**, 41-58.

Vergos G.S., Tziavos I.N. and Andritsanos V.D.; 2005: *Gravity database generation and geoid model estimation using heterogeneous data*. In: Jekeli C., Bastos L. and Fernandes J. (eds), International Association of Geodesy Symposia, Gravity Geoid and Space Missions, **129**, 155-160.

Xu P. and Rummel R.; 1994: *A simulation study of smoothness methods in recovery of regional gravity fields*. Geophys. J. Int., **117**, 472-486.

Corresponding author: Wojciech Jarmołowski
Faculty of Geodesy, Geospatial and Civil Engineering, University of Warmia and Mazury
Ul. Oczapowskiego 2, 10-719 Olsztyn, Poland
Phone: + 48 895234764; e-mail: wojciech.jarmolowski@uwm.edu.pl