Journal of Geodesy (2018) 92:1329–1350
https://doi.org/10.1007/s00190-018-1122-6

REVIEW

Simultaneous estimation of cross-validation errors in least squares collocation applied for statistical testing and evaluation of the noise variance components

Behzad Behnabian[1,2] · Masoud Mashhadi Hossainali[1] · Ahad Malekzadeh[3]

Received: 2 July 2016 / Accepted: 30 January 2018 / Published online: 26 February 2018

Abstract
The cross-validation technique is a popular method to assess and improve the quality of prediction by least squares collocation (LSC). We present a formula for direct estimation of the vector of cross-validation errors (CVEs) in LSC which is much faster than element-wise CVE computation. We show that a quadratic form of CVEs follows Chi-squared distribution. Furthermore, a posteriori noise variance factor is derived by the quadratic form of CVEs. In order to detect blunders in the observations, estimated standardized CVE is proposed as the test statistic which can be applied when noise variances are known or unknown. We use LSC together with the methods proposed in this research for interpolation of crustal subsidence in the northern coast of the Gulf of Mexico. The results show that after detection and removing outliers, the root mean square (RMS) of CVEs and estimated noise standard deviation are reduced about 51 and 59%, respectively. In addition, RMS of LSC prediction error at data points and RMS of estimated noise of observations are decreased by 39 and 67%, respectively. However, RMS of LSC prediction error on a regular grid of interpolation points covering the area is only reduced about 4% which is a consequence of sparse distribution of data points for this case study. The influence of gross errors on LSC prediction results is also investigated by lower cutoff CVEs. It is indicated that after elimination of outliers, RMS of this type of errors is also reduced by 19.5% for a 5 km radius of vicinity. We propose a method using standardized CVEs for classification of dataset into three groups with presumed different noise variances. The noise variance components for each of the groups are estimated using restricted maximum-likelihood method via Fisher scoring technique. Finally, LSC assessment measures were computed for the estimated heterogeneous noise variance model and compared with those of the homogeneous model. The advantage of the proposed method is the reduction in estimated noise levels for those groups with the fewer number of noisy data points.

Keywords Cross-validation errors · Least squares collocation · Statistical tests · Blunder detection · Estimation of noise variance components

1 Introduction

Least squares collocation is known as an estimation method introduced by Moritz (1962) and Krarup (1969) for interpolation and extrapolation of the Earth gravity field using observations measured at discrete points. LSC can be generalized to arbitrary data as a purely analytical approximation method (Grafarend 1976). Assessment of the quality of predictions in terms of both accuracy and precision is a major concern especially when the distribution of data is sparse. This is seen for example in the application of LSC for interpolation of vertical crustal deformation using data from leveling, tide gauges and continuous GPS measurements (e.g., El-Fiky et al. 1997; Vestøl 2006). The two commonly used criteria for measuring the quality of prediction by LSC are the LSC prediction and the cross-validation errors (Darbeheshti and Featherstone 2009). The LSC prediction error is the standard measure to estimate error at the prediction locations (Moritz 1980, p. 154). Leave-one-out (LOO) cross-validation error is a criterion to assess the reliability of prediction around data points, in which one point is omit-

✉ Behzad Behnabian
behzadbehnabian@gmail.com

[1] Faculty of Geodesy and Geomatics, K. N. Toosi University of Technology, Tehran 19967-15433, Iran

[2] Department of Geodesy and Surveying Engineering, Tafresh University, Tafresh 39518-79611, Iran

[3] Department of Mathematics, K. N. Toosi University of Technology, P.O. Box 16315-1618, Tehran, Iran

ted from each LSC prediction before the predicted value at that point is compared with the observed value. CVEs are not only used as a measure to assess the quality of prediction with LSC but also employed for some further applications. The cross-validation technique has been used to improve prediction results by LSC (e.g., Featherstone and Sproule (2006)). One way to increase the quality of prediction by LSC is the precise estimation of covariance parameters that is essential when increased accuracy is required. The usual method for determination of covariance parameters is to compute an empirical covariance model from observed data. This is followed by the inference of covariance model parameters that fit the empirical covariances based on some criteria such as weighted least squares (Mikhail and Ackermann 1976, p. 399; El-Fiky et al. 1997; Arabelos et al. 2007). We can briefly call this method as model fitting to empirical covariances (MFEC). Jarmołowski (2013) used the LOO cross-validation technique for estimating the covariance parameters and a priori noise variance. Accordingly, although the LOO validation is a time-consuming process, the method is very straightforward and accurate. Jarmołowski (2015) also used LOO CVEs for grouping observations when applying a heterogeneous noise model.

A priori noise variance estimation in LSC can be considered as or compared to the regularization problem. The LSC prediction equation is equivalent to the Tikhonov regularization (Koch and Kusche 2002; Kusche and Klees 2002; Eshagh and Sjöberg 2011). In LSC problems, a priori noise variance plays the role of regularization factor (Rummel et al. 1979). This clarifies the importance of precise estimation of a priori noise variance in LSC especially when the condition number of the covariance matrix of observations is large. By showing that LSC is equivalent to the special case of least squares adjustment (LSA), variance of the unit weight is estimated for the collocation by Koch (1977). This estimate gives the scale factor for the covariance matrices being used in the collocation. Koch (1977) also suggests the methods of testing hypotheses and establishing confidence intervals used for the parameters of LSA to be applied for the collocation. Schaffrin (2001) also presents a formula for estimating the variance factor of residuals for random effects model in which the expected value of trend parameters (or random effects vector) is assumed to be known, the formula may also be used for LSC (Snow 2012, p. 40). A reliable predication by LSC requires that we do not have any blunders in the observations. By selecting an appropriate trend for the model together with choosing a suitable covariance function for the data, the covariance parameters could be estimated precisely. Statistical testing is one of the most trustful approaches to ensure such requirements (Wei 1987). The problem of hypothesis testing in relation to LSC has been studied by Krakiwsky and Biacs (1990). That paper commences with an overview of the development of LSC and its relation to LSA. They derived expressions for various random variables and their corresponding covariance matrices. However, some of the required covariance matrices are either ambiguously hidden or unavailable (Krakiwsky and Biacs 1990). Tscherning (1991a, b) introduced a technique for gross error detection based on some type of CVEs using neighboring points (e.g., 10 points around the purposed point). According to Tscherning (1991b) "A comparison of difference between the observed and the predicted value with the error estimate may be used to identify a possible gross error." He proposes implementing neighboring data points for the prediction using LSC, in addition to recommending the trend from the dataset before outlier detection. Referring to this method, Sadiq et al. (2009) state that: "The validity of a priori error estimates as well as the occurrence of outliers may be investigated by comparing the a posteriori determined differences between observed and predicted values with the error estimate of the differences obtained using least squares collocation."

Incomplete knowledge of the noise of observations may occur in many geodetic applications. In such cases the covariance matrix of observations is only partly known, while the unknown part needs to be estimated from the data. Estimating the unknown components of a covariance matrix of observations is generally referred to the variance component estimation (VCE). Many VCE methods are applied in LSA such as minimum norm quadratic unbiased estimator (MINQUE), best invariant quadratic unbiased estimator (BIQUE) and restricted maximum likelihood (REML) (Amiri-Simkooei 2007). Some of these methods have already been used in LSC for certain applications. Yang et al. (2009) introduced an adaptive collocation estimator to balance the covariance matrices of the signals and the observations. The corresponding adaptive factor of this method is constructed by the ratio of the variance components of the signals and the observations. They also present a simplified Helmert-type estimator of the variance components for the collocation. The method is reported to be effective in balancing the contribution of observations and the signals in the collocation model (Yang et al. 2009). Jarmołowski (2015) estimates a priori error associated with heterogeneous non-correlated noise in a dataset. The errors are estimated by the REML, while the solution is composed of a LOO cross-validation technique and REML estimation of a priori noise for different data groups. Jarmołowski (2015) also used the LOO validation for outlier detection. He concludes that group noise estimation by REML is practically equivalent to the removal of outliers and states "REML is a helpful tool in the empirical search of the threshold for outliers and an estimator of sufficiently large noise for outliers."

In this paper we derive equations for simultaneous estimation of CVEs and the corresponding covariance matrix in the trend–signal–noise model of LSC. Moreover, we employ

a hypothesis test on the quadratic form of CVEs which is adopted from the global model test in LSA. We suggest two statistical tests for outlier detection in terms of CVEs which are also adopted from Baarda's data snooping theory (Baarda 1968) and Pope's outlier detection test (Pope 1976). These statistical tests can be used to check the validity and reliability of the LSC results. Depending on the information about the existing noise, a homogeneous or a heterogeneous noise model may be used. For a homogeneous noise model, we need to estimate the noise variance for all of the observations. For this purpose, Jarmołowski and Bakuła (2014) proposed and used the REML method. As an alternative, we present a simple iterative technique for fast estimation of the noise variance factor when a priori noise variance of the observations is unknown. Employing a heterogeneous noise model requires a priori information to classify observations into different groups. In the case when the required information for classification of observations is not accessed, we propose a method using diagram of standardized CVEs to separate observations into different groups with presumed distinct noise variances. Finally, we test and validate the methods discussed in this paper using computed rates of vertical displacements in northern coast of the Gulf of Mexico and assess the results using some various measures indicating accuracy and precision of LSC predictions.

2 Simultaneous estimation of CVEs

The trend–signal–noise model or so-called generalized model of LSC is given by (Moritz 1980, p. 111)

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{y}$ is a $m \times 1$ vector of known observations, $\mathbf{x}$ is the $u \times 1$ vector of unknown trend parameters, $\mathbf{A}$ is a given design matrix of the trend, $\mathbf{s}$ is a zero-mean signal vector at observation points, and $\mathbf{n}$ is a zero-mean noise vector which represents the measuring error. $\mathbf{s}$ and $\mathbf{n}$ are uncorrelated random variates and so is $\mathbf{y}$. Equation (1) may also be written as the linear form of the Gauss–Markov model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{2}$$

where $\mathbf{w} = \mathbf{s} + \mathbf{n}$ is the vector of de-trended observations or residuals which is actually unknown and we can estimate it by

$$\hat{\mathbf{w}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}} = \mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y} \tag{3}$$

where $\hat{\mathbf{x}}$ denotes the vector of estimated parameters and $\mathbf{P}_{\mathbf{A}}^{\perp}$ projects $\mathbf{y}$ into the left null space of $\mathbf{A}$ that would be computed as follows

$$\mathbf{P}_{\mathbf{A}}^{\perp} = \mathbf{I}_m - \mathbf{A}\left(\mathbf{A}^T\mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{C}_{\mathbf{ww}}^{-1} \tag{4}$$

where $\mathbf{C_{ww}} = \mathbf{C_{ss}} + \mathbf{C_{nn}}$ denotes the auto-covariance matrix of residuals, $\mathbf{C_{ss}}$ is the auto-covariance matrix of signals, and $\mathbf{C_{nn}}$ represents the noise covariance matrix. Leave-one-out cross-validation error (LOO-CVE) is defined by (e.g., Darbeheshti and Featherstone 2009)

$$\hat{\varepsilon}_i = y_i - \hat{y}_{i,-i} \quad \forall i = 1, 2, \ldots, m \tag{5}$$

where y_i is ith observation and $\hat{y}_{i,-i}$ is prediction of the ith observation by removing it from the vector of observations. If we assume that the ith observation is already used in estimation of unknown trend parameters, then $\hat{y}_{i,-i}$ is predicted by

$$\hat{y}_{i,-i} = \mathbf{a}_i\hat{\mathbf{x}} + \hat{w}_{i,-i} \tag{6}$$

$\mathbf{a}_i$ is the ith row of the design matrix and $\hat{w}_{i,-i}$ is prediction of the ith residual using the rest of residuals. Replacing $\hat{y}_{i,-i}$ from Eq. (6) into Eq. (5) yields

$$\hat{\varepsilon}_i = y_i - \mathbf{a}_i\hat{\mathbf{x}} - \hat{w}_{i,-i} \tag{7}$$

By using Eq. (3), we can simply see that $\hat{w}_i = y_i - \mathbf{a}_i\hat{\mathbf{x}}$; therefore, the LOO-CVE can be written in terms of residuals

$$\hat{\varepsilon}_i = \hat{w}_i - \hat{w}_{i,-i} \tag{8}$$

$\hat{w}_{i,-i}$ is computed as follows

$$\hat{w}_{i,-i} = \mathbf{c}_{i,-i}\mathbf{C}_{-i,-i}^{-1}\hat{\mathbf{w}}_{-i} \tag{9}$$

where $\mathbf{c}_{i,-i}$ denotes the cross-covariance vector between the signal of the ith observation and the rest of observations. Moreover, $\mathbf{C}_{-i,-i}$ is the auto-covariance matrix of residuals in which the ith row and ith columns have been removed. $\hat{\mathbf{w}}_{-i}$ is the vector of residuals which their ith element has been removed. It also should be noted that if $\mathbf{C_{nn}}$ is a diagonal matrix then $\mathbf{c}_{i,-i}$ also denotes cross-covariance vector between the residual of the ith observation and residual of the rest observations. In other words, $\mathbf{c}_{i,-i}$ is the ith row of $\mathbf{C_{ww}}$ that its ith element has been removed. Computation of the vector of CVEs ($\hat{\boldsymbol{\varepsilon}}$) by Eq. (5) would be a time-consuming process, especially when a large number of observation points are concerned (Darbeheshti and Featherstone 2009; Jarmołowski 2015). Here, we show that it is possible to find a simple relation for efficient calculation of the LOO predictions. This makes the simultaneous computation of $\hat{\boldsymbol{\varepsilon}}$ and its covariance matrix feasible. Let's define $\mathbf{G} = \mathbf{C}_{\mathbf{ww}}^{-1}$ and assume that $\mathbf{C_{nn}}$ is a diagonal matrix, in appendix A it is proved that

$$\mathbf{c}_{i,-i}\mathbf{C}_{-i,-i}^{-1} = -\frac{1}{g_{ii}}\mathbf{g}_{i,-i} \tag{10}$$

where $\mathbf{g}_{i,-i}$ denotes the ith row of $\mathbf{G}$ without the ith element. Substituting Eq. (10) in Eq. (9) gives

$$\hat{w}_{i,-i} = -\frac{1}{g_{ii}}\mathbf{g}_{i,-i}\hat{\mathbf{w}}_{-i} = -\frac{1}{g_{ii}}\sum_{j\neq i} g_{ij}\hat{w}_j \tag{11}$$

Replacing Eq. (11) in Eq. (8) yields

$$\hat{\varepsilon}_i = \hat{w}_i - \hat{w}_{i,-i} = \frac{1}{g_{ii}}g_{ii}\hat{w}_i + \frac{1}{g_{ii}}\sum_{j\neq i} g_{ij}\hat{w}_j = \frac{1}{g_{ii}}\sum_{j=1}^{m} g_{ij}\hat{w}_j \tag{12}$$

Defining $\mathbf{\Lambda} = diag(\mathbf{G})$ in which $diag(\cdot)$ represents a diagonal matrix with diagonal elements of the input matrix. Eq. (12) can be generalized to the following vector form

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{\Lambda}^{-1}\mathbf{C}_{\mathbf{ww}}^{-1}\hat{\mathbf{w}} = \mathbf{\Lambda}^{-1}\mathbf{R}\mathbf{y} \tag{13}$$

where $\mathbf{R} = \mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}$ is usually used for implementation of the trend (Koch 1986). Considering $\mathbf{R} = \mathbf{R}^T$ and $\mathbf{R}\mathbf{C}_{\mathbf{ww}}\mathbf{R} = \mathbf{R}$, one gets the covariance matrix of CVEs using the *law of error propagation* (Koch 1999, p. 99f)

$$\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}} = \mathbf{\Lambda}^{-1}\mathbf{R}\mathbf{\Lambda}^{-1} \tag{14}$$

Equations (13) and (14) also work in the standard form of LSC in which no trend is defined. To this end, $\mathbf{P}_{\mathbf{A}}^{\perp}$ is removed and $\mathbf{R}$ is replaced by $\mathbf{C}_{\mathbf{ww}}^{-1}$. Using Eq. (13), the cpu time for computation of CVEs is drastically reduced in contrast with the element by element computation. This is because instead of solving m systems of $m-1\times m-1$ linear equations (Eq. 9), it only requires the inversion of one $m \times m$ matrix. Diagonal elements of $\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}}$ in Eq. (14) simply give variance of CVEs. Nevertheless, a single CVE variance may also be computed from $\hat{\varepsilon}_i = \hat{w}_i - \hat{w}_{i,-i}$ using the law of error propagation and the formula of LSC prediction error, while the covariances between $\hat{w}_i$ and $\hat{w}_{i,-i}$ should also be considered.

3 Quadratic form of CVEs and statistical tests

Substitution of $\mathbf{w}$ instead of $\hat{\mathbf{w}}$ in Eq. (13) yields a type of cross-validation error which is free from the trend estimation and can be denoted by $\boldsymbol{\varepsilon}$ and presented as follows

$$\boldsymbol{\varepsilon} = \mathbf{\Lambda}^{-1}\mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{w} \tag{15}$$

The covariance matrix of $\boldsymbol{\varepsilon}$ is given by

$$\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} = \mathbf{\Lambda}^{-1}\mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{\Lambda}^{-1} \tag{16}$$

By means of the above relation, we present a quadratic form of CVEs as follows

$$\hat{\Omega} = \hat{\boldsymbol{\varepsilon}}^T\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\hat{\boldsymbol{\varepsilon}} \tag{17}$$

Using Eqs. (13) and (16), one can prove that $\hat{\Omega}$ is equal to the weighted norm of the vector of estimated residuals

$$\begin{aligned}\hat{\Omega} &= \hat{\boldsymbol{\varepsilon}}^T\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\hat{\boldsymbol{\varepsilon}} = \hat{\mathbf{w}}^T\mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}\mathbf{C}_{\mathbf{ww}}\mathbf{\Lambda}\mathbf{\Lambda}^{-1}\mathbf{C}_{\mathbf{ww}}^{-1}\hat{\mathbf{w}} \\ &= \hat{\mathbf{w}}^T\mathbf{C}_{\mathbf{ww}}^{-1}\hat{\mathbf{w}}\end{aligned} \tag{18}$$

Employing Eq. (3) also yields

$$\hat{\Omega} = \mathbf{y}^T(\mathbf{P}_{\mathbf{A}}^{\perp})^T\mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y} = \mathbf{y}^T\mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y} = \mathbf{y}^T\mathbf{R}\mathbf{y} \tag{19}$$

All estimates $\hat{\mathbf{x}}$, $\hat{\mathbf{w}}$, $\hat{\boldsymbol{\varepsilon}}$ and $\hat{\Omega}$ are functions of $\mathbf{y}$ and consequently random variates or random vectors. $\boldsymbol{\varepsilon}$ which is a function of $\mathbf{w}$ is also a random vector.

3.1 Global test on the quadratic form of CVEs

Assuming that the data obeys the multivariate Gaussian distribution, i.e., $\mathbf{y} \sim N(\mathbf{A}\mathbf{x}, \mathbf{C}_{\mathbf{yy}})$, and using Eq. (13), one concludes $\hat{\boldsymbol{\varepsilon}} \sim N(\mathbf{0}, \mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}})$. According to Koch (1999, p. 135), it can be proved that we have $\hat{\Omega} \sim \chi^2_{df}$ with $df = tr(\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}})$, if and only if $\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}}$ is idempotent. From Eqs. (14) and (16), one can see that $\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}} = \mathbf{\Lambda}\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{\Lambda}^{-1}$. Therefore, $\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}}$ is idempotent. Moreover, the degree of freedom df is given by

$$df = tr(\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}}) = tr(\mathbf{P}_{\mathbf{A}}^{\perp}) = m - u \tag{20}$$

where $tr(\cdot)$ is the trace of matrix. Mathematical expectation of $\hat{\Omega}$ is (Koch 1999, p. 134)

$$E\left\{\hat{\boldsymbol{\varepsilon}}^T\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\hat{\boldsymbol{\varepsilon}}\right\} = tr(\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}}) = m - u \tag{21}$$

Following LSA a priori variance factor of noises σ^2 in LSC may be presented as

$$\mathbf{C}_{\mathbf{ww}} = \sigma^2\mathbf{Q}_{\mathbf{ww}} = \mathbf{C}_{\mathbf{ss}} + \sigma^2\mathbf{Q}_{\mathbf{nn}} \tag{22}$$

where $\mathbf{Q}_{\mathbf{ww}}$ is the cofactor matrix of residuals and $\mathbf{Q}_{\mathbf{nn}}$ is the cofactor matrix of noises which is assumed known. It should be mentioned that $\mathbf{Q}_{\mathbf{ww}}$ is known only if σ^2 is known. The a posteriori noise variance factor $\hat{\sigma}^2$ is an unbiased estimator of a priori noise variance factor σ^2. For the Gauss–Markov model, it is given by (Koch 1999, p. 162)

$$\hat{\sigma}^2 = \frac{\mathbf{y}^T\mathbf{Q}_{\mathbf{ww}}^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y}}{m-u} \tag{23}$$

As expected, $\hat{\sigma}^2$ is also a random variable. Observing Eq. (22) reveals that in LSC, computation of $\mathbf{Q_{ww}}$ and consequently $\hat{\sigma}^2$ is possible only when σ^2 is known. From Eqs. (19), (22) and (23), one concludes that

$$\hat{\Omega} = \frac{1}{\sigma^2}\left(\mathbf{y}^T\mathbf{Q}_{\mathbf{ww}}^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y}\right) = \frac{(m-u)\hat{\sigma}^2}{\sigma^2} \tag{24}$$

So if σ^2 is known, then from Eq. (21) we note that $E\left\{\hat{\Omega}\right\} = m-u$; therefore, $E\left\{\hat{\sigma}^2\right\} = \sigma^2$. In LSC the global test will be used to examine the compatibility of observations with the prediction model. Commonly alike LSA the corresponding hypotheses are

$$H_0 : E\left\{\hat{\sigma}^2\right\} = \sigma^2 \text{ versus } H_a : E\left\{\hat{\sigma}^2\right\} \neq \sigma^2 \tag{25}$$

where $\sigma^2 = 1$ is commonly assumed. The test statistic $\hat{\Omega}$ follows the χ^2_{m-u} distribution. In practice, the realization $\underset{\sim}{\mathbf{y}}$ of $\mathbf{y}$ is employed and realization of the vector of CVEs $\underset{\sim}{\hat{\boldsymbol{\varepsilon}}}$ is computed by $\boldsymbol{\Lambda}^{-1}\mathbf{R}\underset{\sim}{\mathbf{y}}$. In terms of CVEs, the null hypothesis is accepted if

$$\chi^2_{\alpha/2;m-u} < \underset{\sim}{\hat{\boldsymbol{\varepsilon}}}^T\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\underset{\sim}{\hat{\boldsymbol{\varepsilon}}} < \chi^2_{1-\alpha/2;m-u} \tag{26}$$

Here, α is the significance level. Possible reasons for failure of this test are: incorrect mathematical modeling, the covariance matrices are not properly assigned (the covariance model is not suitable for the data, the covariance parameters are not correctly estimated, incorrect estimate of the noise variance of observations), and/or blunders are present in the data. We may not know which one of the above reasons causes the test rejection, and the test does not give any additional information. Whatever the reason is, it should be investigated and not ignored. In Sect. 3.3 we confine ourselves to the third possible cause for the test failure, i.e., blunders in the observations.

3.2 Estimation of the noise variance factor when a priori noise variance is unknown

In practice when σ^2 is unknown, computations of $\mathbf{Q_{ww}}$ from Eq. (22) and consequently $\underset{\sim}{\hat{\sigma}}^2$ as realization of $\hat{\sigma}^2$ from Eq. (23) would not be possible. In such a case, one may use the MFEC method to estimate $\underset{\sim}{\hat{\sigma}}^2$. This method will be applied in Sect. 5.1 for the case study of this research. The alternative is to implement a REML method that is discussed by Jarmołowski and Bakuła (2014). Here, we propose a fixed point iteration method (Burden and Faires 2011, p. 56) which is very simple to use. In this method the unknown parameter is written as a function of itself. Derivation of $\mathbf{Q_{ww}}$ from Eq. (22) and substituting it in Eq. (23) for the realized observations $\underset{\sim}{\mathbf{y}}$ yields

$$\underset{\sim}{\hat{\sigma}}^2 = \frac{\underset{\sim}{\mathbf{y}}^T\left(\sigma^{-2}\mathbf{C_{ss}} + \mathbf{Q_{nn}}\right)^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{y}}}{m-u} \tag{27}$$

In the above relation we have two unknowns $\underset{\sim}{\hat{\sigma}}^2$ and σ^2. We should also note that $\underset{\sim}{\hat{\sigma}}^2$ is an estimate of σ^2. By substituting $\underset{\sim}{\hat{\sigma}}^2$ with σ^2 on the right hand side of Eq. (27), the required condition for the fixed point iteration is provided. In order to use this method, the following formula is employed

$$\underset{\sim}{\hat{\sigma}}^2_{(k)} = \frac{\underset{\sim}{\mathbf{y}}^T\mathbf{Q}^{-1}_{\mathbf{ww}(k-1)}\mathbf{P}^{\perp}_{\mathbf{A}(k-1)}\underset{\sim}{\mathbf{y}}}{m-u} \tag{28}$$

where $k = 1, 2, 3, \ldots$ is the iteration number, $\underset{\sim}{\hat{\sigma}}^2_{(k)}$ is our final estimation of $\underset{\sim}{\hat{\sigma}}^2$, $\mathbf{Q}_{\mathbf{ww}(k-1)} = \underset{\sim}{\hat{\sigma}}^{-2}_{(k-1)}\mathbf{C_{ss}} + \mathbf{Q_{nn}}$, and $\mathbf{P}^{\perp}_{\mathbf{A}(k-1)} = \mathbf{I}_m - \mathbf{A}\left(\mathbf{A}^T\mathbf{Q}^{-1}_{\mathbf{ww}(k-1)}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{Q}^{-1}_{\mathbf{ww}(k-1)}$. For $k = 1$, $\mathbf{Q}_{\mathbf{ww}(0)}$ is derived from Eq. (22) as follows

$$\mathbf{Q}_{\mathbf{ww}(0)} = \left(\frac{1}{\sigma_0^2}\right)\mathbf{C_{ss}} + \mathbf{Q_{nn}} \tag{29}$$

where σ_0^2 is a starting value for $\underset{\sim}{\hat{\sigma}}^2$. The iteration process continuous unless $\underset{\sim}{\hat{\sigma}}^2_{(k)} - \underset{\sim}{\hat{\sigma}}^2_{(k-1)}$ is less than a specified threshold value.

3.3 Blunder detection tests

Outliers are a main concern in LSC. Here we adopt Baarda's data snooping theory (Baarda 1968). The null hypothesis assumes that no blunder exists in the observations vector, while the alternative hypothesis assumes that only one observation at a time is erroneous

$$H_0 : E\left\{\hat{w}_i\right\} = 0 \text{ versus } H_a : E\left\{\hat{w}_i\right\} = \nabla w_i \neq 0 \tag{30}$$

The alterative hypothesis can also be written as

$$H_a : \hat{\mathbf{w}} \sim N(\nabla\mathbf{w}_i, \mathbf{C_{ww}}) \tag{31}$$

with $\nabla\mathbf{w}_i = \boldsymbol{\eta}\nabla w_i$, in which $\boldsymbol{\eta}$ is an m-dimensional vector that contains 1 for the ith element, while the rest of the elements are equal to zero. When the a priori noise variance factor is known, we can apply the Baarda's statistic for testing

transformed residuals for the Gauss–Markov model (Eq. 2) that is derived as (Teunissen 2000)

$$T_1 = \frac{\boldsymbol{\eta}^T \mathbf{C}_{\mathbf{ww}}^{-1} \hat{\mathbf{w}}}{\sqrt{\boldsymbol{\eta}^T \mathbf{C}_{\mathbf{ww}}^{-1} \mathbf{C}_{\hat{\mathbf{w}}\hat{\mathbf{w}}} \mathbf{C}_{\mathbf{ww}}^{-1} \boldsymbol{\eta}}} = \frac{\boldsymbol{\eta}^T \mathbf{Q}_{\mathbf{ww}}^{-1} \hat{\mathbf{w}}}{\sigma \sqrt{\boldsymbol{\eta}^T \mathbf{Q}_{\mathbf{ww}}^{-1} \mathbf{Q}_{\hat{\mathbf{w}}\hat{\mathbf{w}}} \mathbf{Q}_{\mathbf{ww}}^{-1} \boldsymbol{\eta}}} \sim N(0,1) \tag{32}$$

where T_1 is the Baarda's test statistic. In LSC $\mathbf{C}_{\mathbf{ww}}^{-1}$ is not a diagonal matrix. Therefore, the corresponding statistic is not the standardized residual of the ith element. By substituting $\mathbf{C}_{\hat{\mathbf{w}}\hat{\mathbf{w}}} = \mathbf{P}_{\mathbf{A}}^{\perp} \mathbf{C}_{\mathbf{ww}} (\mathbf{P}_{\mathbf{A}}^{\perp})^T$ in Eq. (32), multiplying $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}$ to $\mathbf{C}_{\mathbf{ww}}^{-1}$ and employing Eq. (13) we obtain

$$T_1 = \frac{\boldsymbol{\eta}^T \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1} \mathbf{C}_{\mathbf{ww}}^{-1} \hat{\mathbf{w}}}{\sqrt{\boldsymbol{\eta}^T \mathbf{C}_{\mathbf{ww}}^{-1} \mathbf{P}_{\mathbf{A}}^{\perp} \mathbf{C}_{\mathbf{ww}} (\mathbf{P}_{\mathbf{A}}^{\perp})^T \mathbf{C}_{\mathbf{ww}}^{-1} \boldsymbol{\eta}}} = \frac{\boldsymbol{\eta}^T \boldsymbol{\Lambda} \hat{\boldsymbol{\varepsilon}}}{\sqrt{\boldsymbol{\eta}^T \mathbf{R} \mathbf{C}_{\mathbf{ww}} \mathbf{R}^T \boldsymbol{\eta}}} = \frac{\lambda_{ii}}{\sqrt{r_{ii}}} \hat{\varepsilon}_i \tag{33}$$

Using Eq. (14) one concludes that $\sigma_{\hat{\varepsilon}_i}^2 = r_{ii}/\lambda_{ii}^2$, where r_{ii} is the ith diagonal element of $\mathbf{R}$ and λ_{ii} is the ith diagonal element of $\boldsymbol{\Lambda}$. Applying this together with Eq. (33) gives

$$T_1 = \frac{\hat{\varepsilon}_i}{\sigma_{\hat{\varepsilon}_i}} = \hat{e}_i \sim N(0,1) \tag{34}$$

where $\hat{e}_i$ is the standardized CVE of the ith observation. By employing Eqs. (14) and (13), the vector of standardized CVEs is estimated as

$$\hat{\mathbf{e}} = (diag(\mathbf{C}_{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}}))^{-1/2} \hat{\boldsymbol{\varepsilon}} = \boldsymbol{\Lambda}^{1/2} \boldsymbol{\Delta}^{-1/2} \boldsymbol{\Lambda}^{1/2} \hat{\boldsymbol{\varepsilon}} = \boldsymbol{\Lambda}^{1/2} \boldsymbol{\Delta}^{-1/2} \boldsymbol{\Lambda}^{1/2} \boldsymbol{\Lambda}^{-1} \mathbf{R}\mathbf{y} = \boldsymbol{\Delta}^{-1/2} \mathbf{R}\mathbf{y} \tag{35}$$

where $\boldsymbol{\Delta} = diag(\mathbf{R})$. The null hypothesis is accepted if the computed test statistic $\hat{e}_i$ follows the standard normal distribution at the α_0 significance level. In practice, the realized test statistic $\underset{\sim}{\hat{e}}_i$ should satisfy the following condition

$$\left| \underset{\sim}{\hat{e}}_i \right| < N_{1-\alpha_0/2} \quad \forall i = 1, 2, \ldots, m \tag{36}$$

For the case that σ^2 is unknown and the a posteriori noise variance factor is used instead, the Pope's test statistic is employed (Pope 1976; Koch 1999, pp. 304–305)

$$T_2 = \frac{\boldsymbol{\eta}^T \mathbf{Q}_{\mathbf{ww}}^{-1} \hat{\mathbf{w}}}{\hat{\sigma} \sqrt{\boldsymbol{\eta}^T \mathbf{Q}_{\mathbf{ww}}^{-1} \mathbf{Q}_{\hat{\mathbf{w}}\hat{\mathbf{w}}} \mathbf{Q}_{\mathbf{ww}}^{-1} \boldsymbol{\eta}}} \sim \tau(1, m-u-1) \tag{37}$$

where τ denotes Thompsons τ-distribution (Koch 1999). In a similar manner to Eqs. (33), (37) can also be written as

$$T_2 = \frac{\hat{\varepsilon}_i}{\hat{\sigma}_{\hat{\varepsilon}_i}} = \tau_i \sim \tau(1, m-u-1) \tag{38}$$

The null hypothesis is accepted if

$$\left| \underset{\sim}{\tau}_i \right| < \tau_{1-\alpha_0/2, m-u-1} \tag{39}$$

where $\underset{\sim}{\tau}_i$ is the realized standardized CVE of the ith observation. The significance level for outlier test in LSA is customarily selected as $\alpha_0 = 0.001$, where this value may also be adopted for LSC. Bonferroni inequality can also be applied for the computation of the significance level α_0 (Vaníček and Krakiwsky 1986, p. 231). This theorem states that if the residuals are uncorrelated, the probability of type I error for the entire set of tested residuals will be α. Now if the residuals are correlated, this probability will be less than α. In practice α_0 can be obtained from $\alpha_0 = \alpha/m$ (Koch 1999, p. 306).

4 Estimation of the noise variance components using REML

An ultimate generalization of the covariance model of residuals in Eq. (22) is achieved by introducing a variance component model for noise of the observations which may be written as

$$\mathbf{C}_{\mathbf{ww}} = \mathbf{C}_{\mathbf{ss}} + \sum_{k=1}^{p} \sigma_k^2 \mathbf{Q}_k \tag{40}$$

where $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$ are unknown noise variance components that are to be estimated and $\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_p$ are linearly independent and symmetric cofactor matrices that are assumed to be known. Different NVCs are gathered in the vector of unknown parameters $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta} = \left[\sigma_1^2 \ \sigma_2^2 \cdots \sigma_p^2 \right]^T$. The covariance matrix of residuals is also regarded as $\mathbf{C}_{\boldsymbol{\theta}}$. The problem is completely similar to variance component estimation (VCE); therefore, VCE methods may be employed. REML is a method of variance component estimation for a Gauss–Markov model when the probability density function of observations possesses a normal distribution. Koch (1986) derives an iterative procedure for the maximum-likelihood estimates of the variance components using the orthogonal complement likelihood function. This approach is equivalent to the REML estimation. He also shows that these estimators are identical to the best invariant quadratic unbiased estimators (BIQUE) of variance components and also to the minimum norm quadratic unbiased (MINQUE) estimators.

In statistics, one of the methods applied for estimating spatial covariances is REML (Kitanidis 1983; Stein 1999, pp. 170–171). The following negative log-likelihood (NLLF) function is minimized in REML (Koch 2007; Jarmołowski 2015)

$$\mathrm{NLLF}_{\mathrm{REML}}(\boldsymbol{\theta}) = \frac{1}{2}\ln\det\left(\mathbf{C}_{\boldsymbol{\theta}}\right) + \frac{1}{2}\ln\det\left(\mathbf{A}^T\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{A}\right) + \frac{1}{2}\left(\mathbf{y}^T\mathbf{R}_{\boldsymbol{\theta}}\mathbf{y}\right) \quad (41)$$

The third term in the above equation is equal to $\hat{\boldsymbol{\varepsilon}}^T\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1}\hat{\boldsymbol{\varepsilon}}$ (see Eqs. 18, 19). This implies that the REML method minimizes the weighted norm of the vector of CVEs. In practice, the vector of unknown parameters $\boldsymbol{\theta}$ is computed using Fisher scoring (Koch 1986; Grodecki 1999). One of the main challenges for the problem is how to classify observations into distinct groups with presumed different NVCs in those cases which no information about the noise model is available. Jarmołowski (2015) uses CVEs to split observations in two groups; one with better and the other with worse NVCs. In this paper we will use standardized CVEs for classification of the data into groups with distinct NVCs. The method will be described in the next section.

5 Application of the theory using a dataset of vertical displacement rates in northern coast of the Gulf of Mexico with unknown noise variances

In order to examine the methods presented in this paper, we selected a dataset which contains rates of vertical displacement at benchmarks in a part of the northern coast of the Gulf of Mexico and lower Mississippi valley. The data are given in NOAA technical report (Shinkle and Dokka 2004). The selected area contains 1683 NGS benchmarks for which the rates of vertical displacements are computed by Shinkle and Dokka (2004). Computation of the rates of vertical displacements had been accomplished using first-order leveling data collected between 1964 and 1995. The "unadjusted heights" are used for computation of rates of vertical displacements. These values are derived from precise elevation differences observed in the leveling survey process. Corrections for known systematic errors (i.e., orthometric, rod, level, temperature, astronomic, refraction, magnetic) have been applied automatically to the observations (Shinkle and Dokka 2004). Figure 1 depicts the spatial distribution of the benchmarks that comprise the subsidence network in the study area.

Here, we use a linear trend defined by

$$\mathbf{A} = \begin{bmatrix} 1 & \phi_1 & \lambda_1 \\ \vdots & \vdots & \vdots \\ 1 & \phi_m & \lambda_m \end{bmatrix} \quad (42)$$

The exponential covariance function is applied for constitution of the covariance model. It is described by

$$C(d) = C_0\exp(-kd) \quad (43)$$

where the distance between the points is represented with d. C_0 is the signal variance and k is a correlation parameter which is reciprocal of the correlation distance d_0, i.e., $d_0 = 1/k$. As it will be seen in the next subsection, this covariance function has a proper adaptation with the empirical covariances of our case study data. We use spherical distances to determine the covariances between data points. In order to estimate covariance parameters (C_0, k), the MFEC method was employed. Noise variances of data points are actually unknown and should be estimated. We tested two

Fig. 1 Plot of the network of benchmarks used for interpolation of rates of vertical displacement in the selected area

Table 1 Basic statistics for the dataset of rate of subsidence in selected region from the northern coast of the Gulf of Mexico and its residuals after de-trending data, unit: mm/a

Dataset	Min	Max	Mean	Median	SD	RMS
Observations	− 51.94	2.43	− 9.20	− 9.07	4.87	10.41
Residuals	− 42.03	9.64	0.00	0.59	3.88	3.88

Table 2 Estimated trend parameters using the trend defined by Eq. 42 and their estimated SDs

$\hat{\underset{\sim}{x}}_1$ (mm/a)	$\hat{\sigma}_{\hat{x}_1}$ (mm/a)	$\hat{\underset{\sim}{x}}_2$ (mm/a)/rad	$\hat{\sigma}_{\hat{x}_2}$ (mm/a)/rad	$\hat{\underset{\sim}{x}}_3$ (mm/a)/rad	$\hat{\sigma}_{\hat{x}_3}$ (mm/a)/rad
− 361.88	25.10	354.97	11.47	35.37	4.87

different ideas for modeling the noises of data points. At first we applied a homogeneous noise model for the whole dataset and estimated the a posteriori noise variance. Blunder detection test was applied and the outliers were removed. Then, using standardized CVEs for the data points, we divided the data in three different groups and estimated the noise variance component of each group by the REML method. Table 1 shows basic statistics of the dataset and its estimated residuals using the defined trend, while Table 2 illustrates the estimated values of trend parameters and their estimated standard deviations. Here, for computing the trend parameters and the residuals, we have used a unit weight for all of observations.

5.1 Interpolation and testing, using a homogeneous noise model

The entire dataset is derived from the precise leveling data. At first, we assume that noises for all of data points are the same. We used Eq. (22) for noise modeling in which $\mathbf{Q_{nn}}$ is assumed to be $\mathbf{I}_m$, i.e.,

$$\mathbf{C_{ww}} = \mathbf{C_{ss}} + \sigma^2 \mathbf{I}_m \tag{44}$$

Therefore, we have a homogeneous noise model. Using MFEC method (Mikhail and Ackermann 1976, p. 403), covariance parameters (C_0, k) were estimated for signals. Figure 2 depicts the computed empirical covariances of residuals which are assigned to equally spaced 5 km distances and the fitted model. In this figure $C_w(0)$ is the variance of residuals and $C_s(0)$ is the signal variance. The difference between these two values is an estimation of the noise variance which can be derived by the MFEC method (Mikhail and Ackermann 1976, p. 399; El-Fiky et al. 1997).

The estimated value of the noise variance is derived by $C_n(0) = C_w(0) - C_s(0)$ which is actually computed equal to 5.3916 $(\text{mm/a})^2$ for this dataset. Since we assume $\mathbf{Q_{nn}} = \mathbf{I}_m$, the estimated noise variance factor by the MFEC method would be the same $\hat{\underset{\sim}{\sigma}}^2 = 5.3916$. On the other hand, a posteriori noise variance factor computed by the equations explained in Sect. 3.2 is equal to 5.7329. The difference between the standard deviation of noise computed by these two methods is only about 3.1% of $\hat{\underset{\sim}{\sigma}}$ by the proposed method. These two values do not significantly differ from each other. However, we need to select one of them to work with, which should be preferred. Table 3 shows the realization of the quadratic form of CVEs $\hat{\underset{\sim}{\Omega}}$ from Eq. (18) for each of these two estimates.

As it is seen, $\hat{\underset{\sim}{\Omega}}$ computed using $\hat{\underset{\sim}{\sigma}}^2 = 5.7329$ by the proposed method is exactly fixed on 1680 which is the degree of freedom, df, for this dataset with 3 trend parameters, while $\hat{\underset{\sim}{\Omega}}$ computed by the other $\hat{\underset{\sim}{\sigma}}^2$ has a considerable difference with df. Regarding Eq. (21) in which we have $E\left\{\hat{\underset{\sim}{\Omega}}\right\} = df$, one concludes that $\hat{\underset{\sim}{\sigma}}^2$ estimated by the proposed method has a priority to the other one for this case study. Therefore, we use $\hat{\underset{\sim}{\sigma}}^2 = 5.7329$ in the consequent computations. We applied the outlier detection test on the data set based on Eq. (38). The significance level has been selected as $\alpha = 0.05$, and therefore, considering the ending explanation in Sect. 3.3 α_0 is set as $\alpha_0 = 0.05/1683$ at the beginning of outlier detection process. Accordingly, the rejection level for the maximum value of standardized CVEs is computed as 4.165. The outliers are detected by an iterative approach in which each time only one data point with realized maximum absolute standardized CVE, $\max\left|\hat{\underset{\sim}{e}}_i\right|$, is removed and the covariance parameters (C_0, k) together with $\hat{\underset{\sim}{\sigma}}^2$ are re-estimated with the rest of data. Figure 3 depicts the flowchart of outlier detection and remove algorithm.

Through this process, 26 data points were detected as outliers and removed. Figure 4 illustrates the computed empirical covariances and the fitted model in pursuit of removing blunders. In comparison with Fig. 2 a remarkable decrease in the estimated value of $C_w(0)$ is observed. As a result, this indicates that by removing the outliers the estimated values of residuals decrease. Many computed empirical covariances are not significantly changed. The maximum change is 4.123 $(\text{mm/a})^2$ for the zero distance, while the other notable changed values are 0.74, 0.26, 0.34 and 0.11 $(\text{mm/a})^2$ for the distances of 2.5, 12.5 17.5 and 42.5 km, respectively. These changes do not have a serious

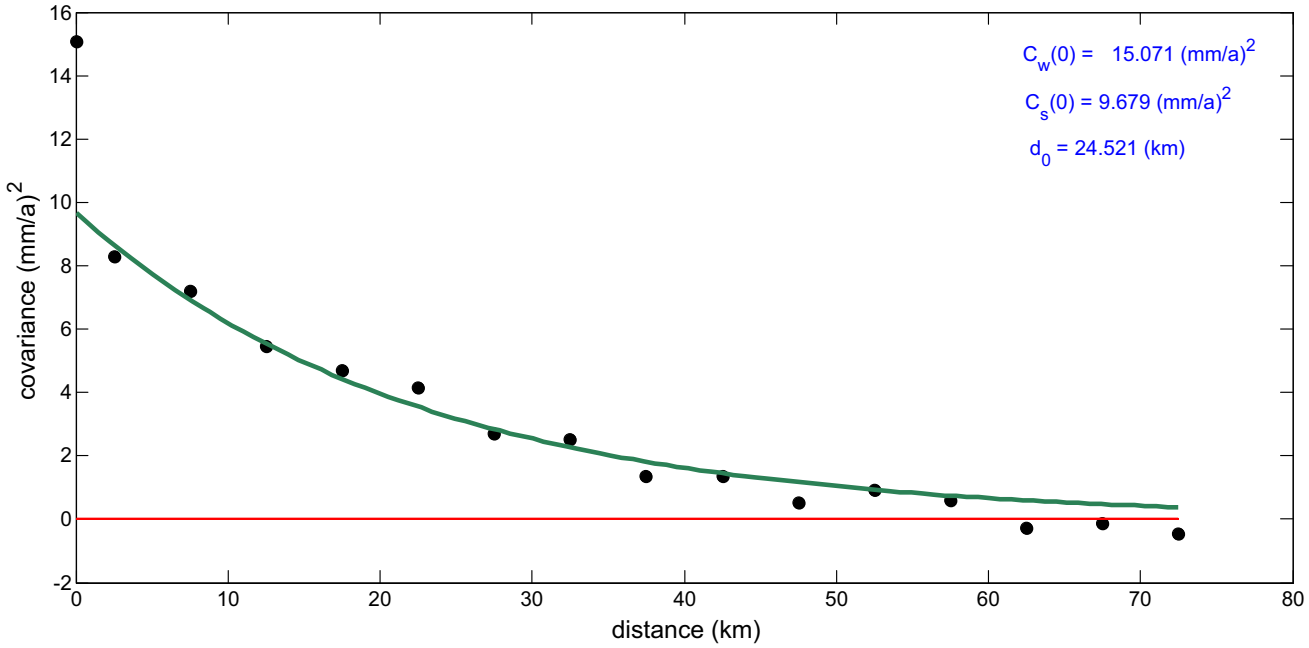


Fig. 2 Empirical covariance of residuals and the fitted exponential covariance model

Table 3 Noise variance factors computed using MFEC method and the proposed method in Sect. 3.2 and the respective quadratic forms of CVEs

Estimated by the MFEC method		Estimated by the proposed method	
$\hat{\sigma}^2$	$\hat{\Omega}$	$\hat{\sigma}^2$	$\hat{\Omega}$
5.39	1762.7	5.73	1680

influence on the estimated signal variance and the correlation distance. Because of the extreme reduction in $C_w(0)$ and the insignificant change of $C_s(0)$, the noise variance factor derived by the MFEC method is intensely decreased to 1.061. On the other hand, the value of the noise variance factor computed by the proposed method in Sect. 3.2 is 0.970. Again, it implies that for the dataset of this case study, there is not a remarkable difference between outcomes of these two methods.

Table 4 illustrates estimated $\hat{\Omega}$ for $\hat{\sigma}^2$ derived by each of the discussed methods. We preferred to work with $\hat{\sigma}^2$ derived by the method proposed in Sect. 3.2 which yields a more reliable result for $\hat{\Omega}$.

Table 5 shows changes of computed covariance parameters and the root of a posteriori noise variance factor together with the maximum absolute standardized CVEs and their RMS values after removing outliers. $\hat{\mathbf{e}}$ is computed by Eq. (35) using $\mathbf{y}$, the realization of observations vector.

In Table 5 it is seen that due to the omission of detected gross errors, $\hat{\sigma}$ as a measure of data precision or estimated measurement error for each of the observations is significantly decreased; about 59%. This implies that removing outliers will improve precision of the data used for prediction by LSC. Another estimated quantity to be noted in this table is standardized CVE. Since we assume $\hat{e}_i \sim N(0, 1)$ (Eq. 34), RMS($\hat{e}_i$) which is equal to the standard deviation of $\hat{e}_i's$ is expected to be near 1, but in Table 5 we see that after removing outliers it decreases from 0.97 to 0.95. The reason is not exactly known but one can guess that it may be because of the need to employ a heterogeneous noise model for the observations that will be discussed in the next section. Now we are interested to see the impact of the elimination of outliers from the dataset and reduction in estimated noise variance on LSC prediction results. In order to assess the quality of prediction by LSC, we used three different types of errors and estimated noise of the observations. The first is CVE which is computed by Eq. (13). CVEs indicate the accuracy of LSC predictions at data points and their neighboring area. The second is the LSC prediction error at data points that may be called as internal error (see "Appendix B"). Equation (B3) is applied for computing the internal errors for each of the data points. Using Eq. (B1), LSC prediction error has also been computed on a regular grid of 1653 interpolation points with $5' \times 5'$ spacing located in between $29.15°N \leq \phi \leq 31.50°N$ and $89.10°W \leq \lambda \leq 93.75°$W. Noise of the observation in Eq. (1) can be estimated by Eq. (B4) in "Appendix B". The estimated noise for each of the observation points gives the

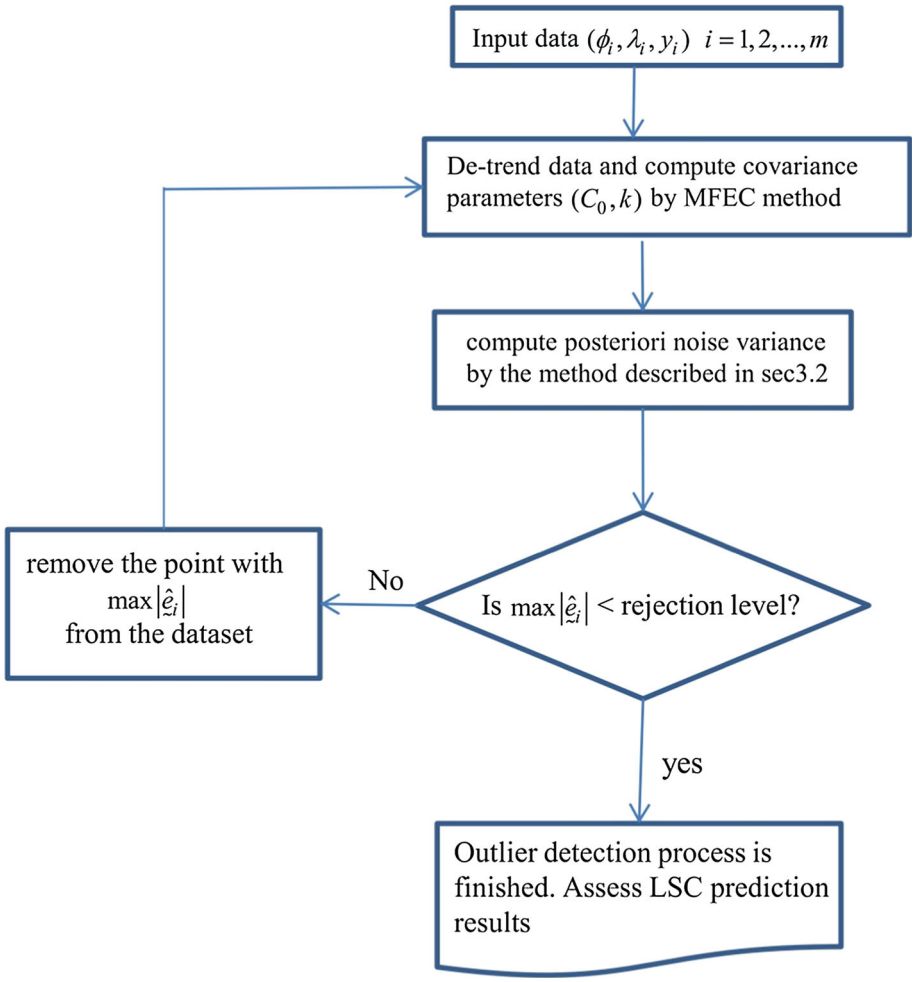


Fig. 3 Flowchart of the algorithm used for detection and removing outliers from dataset using homogeneous noise model

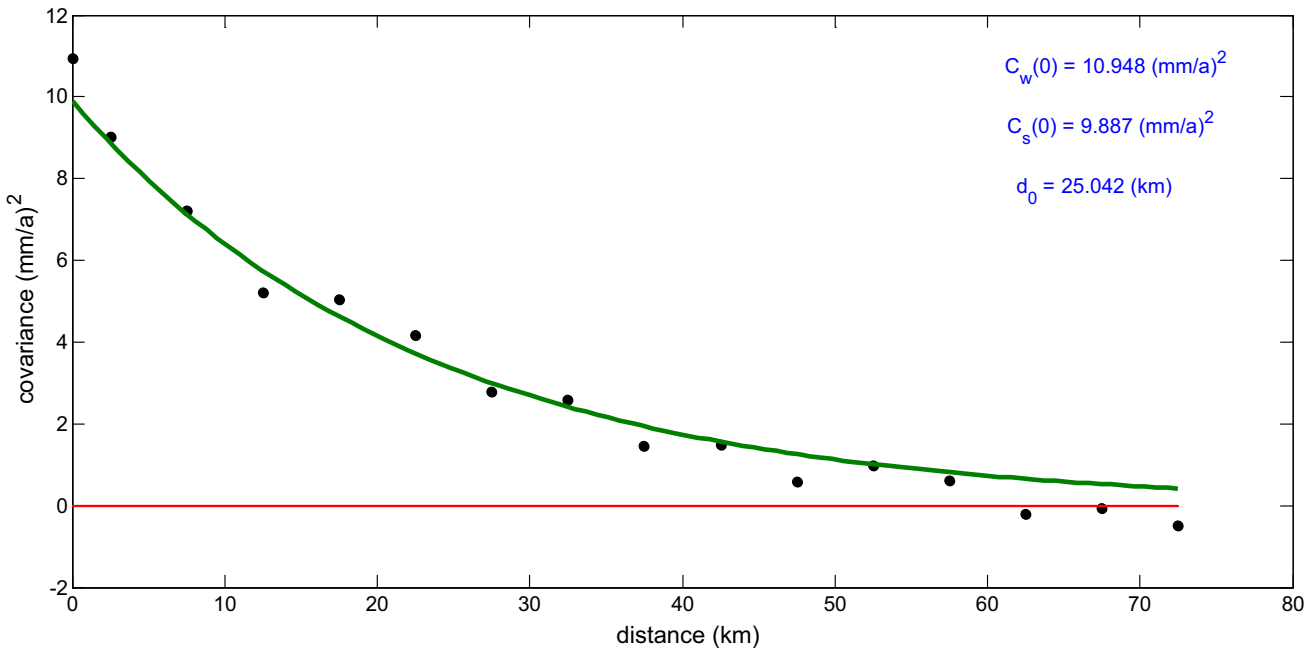


Fig. 4 Empirical covariance of residuals and the fitted exponential covariance model after removing outliers

Table 4 Noise variance factors computed using MFEC method and the proposed method and the respective quadratic forms of CVEs after removing 26 outliers

Estimated by the MFEC method		Estimated by the proposed method	
$\hat{\underset{\sim}{\sigma}}^2$	$\hat{\underset{\sim}{\Omega}}$	$\hat{\underset{\sim}{\sigma}}^2$	$\hat{\underset{\sim}{\Omega}}$
1.06	1579.5	0.97	1654

Table 5 Changes of standardized CVEs, covariance parameters and estimated noise variance after removing outliers

Status	Max $\lvert\hat{\underset{\sim}{e}}_i\rvert$	RMS $(\hat{\underset{\sim}{e}}_i)$	$\hat{\underset{\sim}{\sigma}}$	$C_w(0)$ (mm/a)^2	$C_s(0)$ (mm/a)^2	d_0 (km)
Before removing outliers	17.92	0.97	2.394	15.071	9.679	24.521
After removing outliers	4.12	0.95	0.985	10.948	9.887	25.042

difference between the observed value and the LSC prediction. The lower noise means better adaptation of predictions on the observations at the point locations. In Table 6 we have compared the quality of LSC prediction before and after removing outliers using statistics for each of these error types together with the estimated noise of observations.

The comparison of the results for different errors (Table 6) shows that after removing the detected outliers, the CVEs and internal errors are significantly decreased, while the RMS of LSC prediction errors on a regular grid of prediction points is not considerably affected. We see that RMS$(\hat{\underset{\sim}{\varepsilon}}_i)$ and RMS$(\sigma_{\hat{y}_i})$ have been reduced by about 51% and 39%, respectively. This is at the cost of losing 26 data points detected as the outliers. Nevertheless, RMS$(\sigma_{\hat{y}_0})$ is decreased about 4% which is not considerable. The reason is related to sparse spatial distribution of data points for this case study. RMS$(\hat{\underset{\sim}{\varepsilon}}_i)$ and RMS$(\sigma_{\hat{y}_i})$ can be treated as overall measures of the accuracy and precision of LSC predictions at the data points, respectively. However, their simultaneous reduction is the result of improvement of our dataset quality. As the measures of precision and accuracy, they can be compared with each other. Their difference before and after elimination of outliers is 1.53 and 0.61 (mm/a), respectively. This can be interpreted as improvement of the reliability of LSC prediction at the data points and around their neighborhood area which is the natural consequence of omitting outliers. RMS$(\hat{\underset{\sim}{\varepsilon}}_i)$ can also be considered as a measure of accuracy of observed data points from another perspective. Thus, one can compare it with estimated standard deviation of the noise of observations. Given the assumption that $\mathbf{Q_{nn}} = \mathbf{I}_m$, one can directly compare $\hat{\underset{\sim}{\sigma}}$ from Table 5 with RMS$(\hat{\underset{\sim}{\varepsilon}}_i)$ in Table 6. Before removing outliers, $\hat{\underset{\sim}{\sigma}}$ is computed as 2.39 while RMS$(\hat{\underset{\sim}{\varepsilon}}_i)$ is 2.59. After that the outliers are eliminated $\hat{\underset{\sim}{\sigma}}$ is 0.99 which is still close to 1.26 derived for RMS$(\hat{\underset{\sim}{\varepsilon}}_i)$. This comparison shows that both accuracy and precision of our dataset are significantly improved by removing outliers. The last row of Table 6 draws our attention to significant reduction in the estimated noise of observations in pursuit of omitting blunders. RMS of the noise is reduced by 67% and the noise range from maximum to minimum values is extremely decreased which concludes that implementation of the method brings a better adaptation of LSC prediction with the observations. This may be considered as the remarkable achievement of detection and removing outliers that directly affects the LSC predictions not only at data point locations but also all over of the interpolation area. According to slight reduction in LSC prediction errors on regular grid of interpolation points in Table 6, one concludes that filtering data from gross errors which improves accuracy and precision of observed data points do not have an effect on the quality of prediction at distant (with respect to the correlation distance) points. In order to assess the accuracy of prediction at those points with substantial distance to data points, we can use lower cutoff CVEs for which the neighboring points are also removed in CVE processes. By this way, during the computation of CVE for a data point not only the same point is omitted, but also all of the data points located at a certain distance to that point are also removed. Therefore, we can have a better judgment about the quality of prediction for those points located distantly to our data points. Lower cutoff CVEs can be used to measure the accuracy of prediction especially for datasets with sparse distribution of data points like our case study. From another point of view, in case of heterogeneous distribution of data point locations, one can use the lower cutoff cross-validation technique in order to simulate homogeneous resolution of the prediction everywhere in the area. Table 7 displays statistics of computed lower cutoff CVEs for some different radii of vicinity before and after elimination of the outliers.

These results indicate that removing outliers can improve LSC prediction even for those locations with no data points at their surrounding area in a certain distance. This type of error may be treated as a measure of accuracy in those parts of the prediction area which are far from the observation points. By increasing the radius of vicinity, the amounts of changes in RMS of lower cutoff CVEs before and after elimination of blunders are decreased. Considering different lower cutoff

Table 6 Comparison of different types of LSC error changes due to removing outliers detected by the proposed technique (unit: mm/a)

Type of error	Status	No. points	Min	Max	Mean	RMS
CVE ($\hat{\underset{\sim}{\varepsilon}}_i$)	Before removing outliers	1683	− 46.95	10.27	0.00	2.59
	After removing outliers	1657	− 5.72	5.35	0.00	1.26
LSC internal error ($\sigma_{\hat{y}_i}$)	Before removing outliers	1683	0.65	1.66	1.05	1.06
	After removing outliers	1657	0.32	0.90	0.64	0.65
LSC prediction error on a regular grid of points ($\sigma_{\hat{y}_0}$)	Before removing outliers	1653	0.88	3.59	2.57	2.65
	After removing outliers	1653	0.56	3.62	2.47	2.55
Estimated noise of observations ($\hat{\underset{\sim}{n}}_i$)	Before removing outliers	1683	− 39.22	8.37	0.00	2.12
	After removing outliers	1657	− 3.30	2.81	0.00	0.70

Table 7 Comparison of lower cutoff CVEs for different radii of vicinity before and after removing outliers detected by the proposed technique (unit: mm/a)

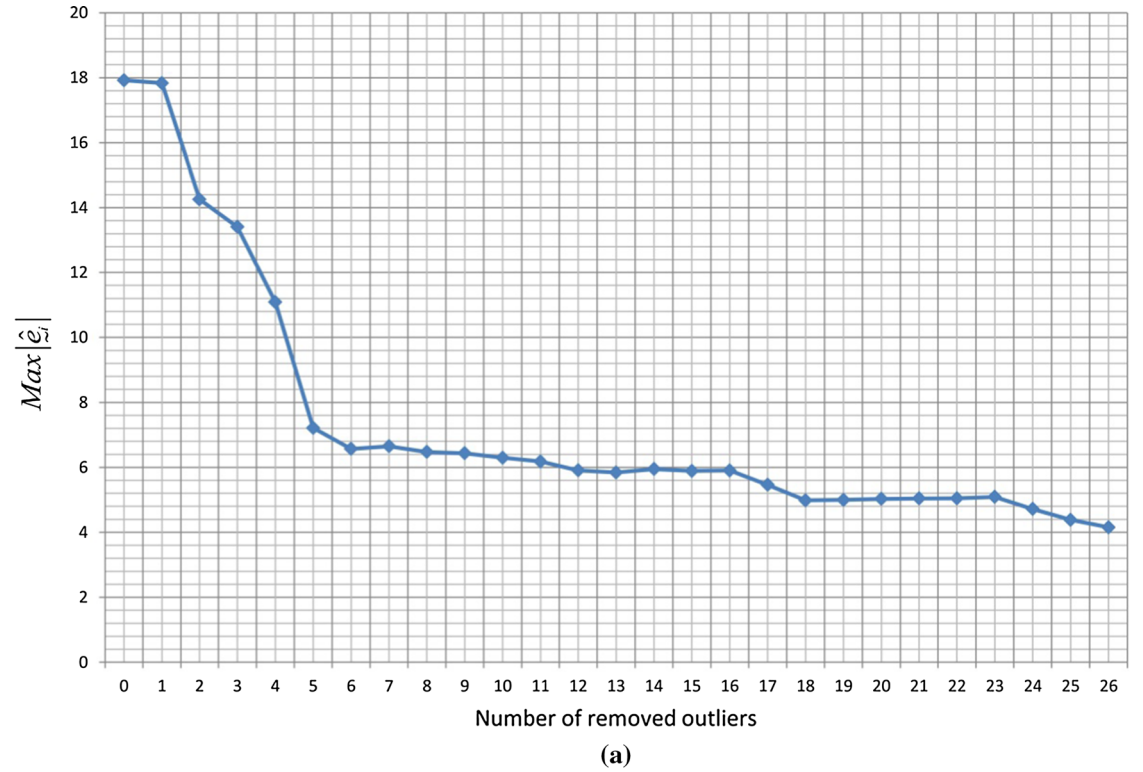
Radius of vicinity (km)	Status	No. points	Min	Max	*Mean*	Mean of absolute value	RMS
5	Before removing outliers	1683	− 43.58	14.47	0.08	2.38	3.45
	After removing outliers	1657	− 14.64	18.99	0.07	1.92	2.78
15	Before removing outliers	1683	− 42.26	12.19	0.09	3.06	4.19
	After removing outliers	1657	− 20.17	12.86	0.06	2.66	3.69
25	Before removing outliers	1683	− 40.43	14.21	0.02	3.39	4.53
	After removing outliers	1657	− 20.99	17.42	− 0.03	3.01	4.04

CVEs in Table 7, we see that RMS of the error is reduced by 19.5, 11.9 and 10.8% for 5, 15 and 25 km cutoff values, respectively. Comparing lower cutoff CVEs in Table 7 with prediction errors at grid points in Table 6 in terms of RMS shows that after removing outliers the 5 km lower cutoff CVEs (2.78 mm/a) as a measure of accuracy corresponds better with prediction errors at grid points (2.55 mm/a) which is usually treated as the measure of precision for prediction over the area. These two values may also be expected to be close to estimated noise standard deviation of data points which is actually computed as 0.985 (mm/a), but some factors can probably disturb that. One of the impressive factors is the spatial distribution of data points. For this case study, the data points are heterogeneously distributed over the proposed prediction area which is covered by the regular grid of interpolation points. This is probably the main factor causing a significant difference between the prediction accuracy (RMS of 5 km lower cutoff CVEs) and estimated standard deviation of the noise of observations. As another result of our analysis, we can also conclude that due to removing outliers while the accuracy is improved about 19.5% (from the first row of Table 7) the precision changed about 4% (from the third row of Table 6) for this case study. Figure 5 illustrates the variation of $\max \left|\hat{\underset{\sim}{e}}_i\right|$ and estimated noise variance factor $\hat{\underset{\sim}{\sigma}}^2$ through the detection and removal of outlying data. It is seen that by removing each of the detected outliers the value of $\hat{\underset{\sim}{\sigma}}^2$ is decreased.

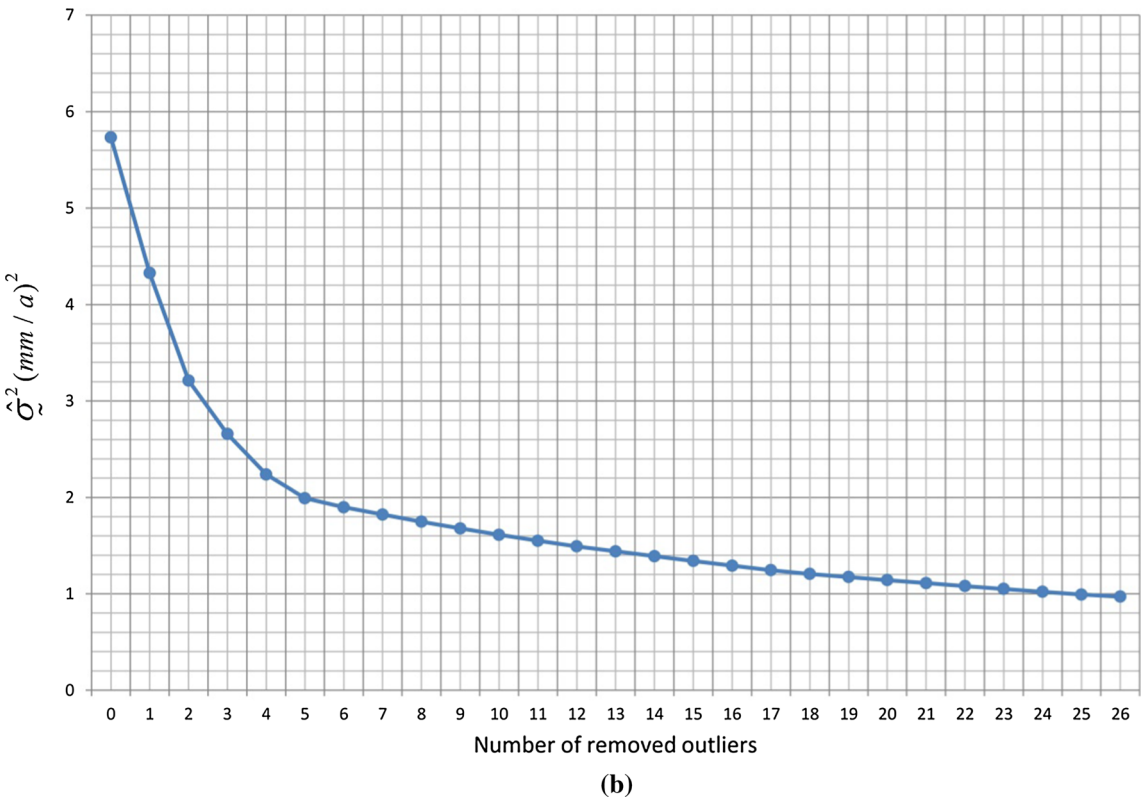
In Fig. 5, after rapid fall around 5 outliers, we see a systematic decrease for these two quantities until $\max \left|\hat{\underset{\sim}{e}}_i\right|$ reaches the rejection level. Slight reduction rate of $\max \left|\hat{\underset{\sim}{e}}_i\right|$ and $\hat{\underset{\sim}{\sigma}}^2$ from outlier 6 to outlier 26 reminds us that all these detected observations may not essentially hold blunders in them. Therefore, we think about a heterogeneous noise model for the observations. In fact, we assume that observations can be divided into at least two groups with different noise variances. If we have two groups of observations where the variance of the noise for one of them is larger than that from the other, and we estimate a homogeneous noise model for them. Then, we should expect to detect some of the observations from the group with larger noise as outliers while they are not real outliers, and the reason of larger value for $\left|\hat{\underset{\sim}{e}}_i\right|$ is that noise variance of these observations is larger than the overall noise variance of total observations.

5.2 Interpolation using a heterogeneous noise model

Here we divide the dataset into distinct groups and assume different NVCs for each of them. In order to compute $\mathbf{C_{ww}}$, Eq. (40) is employed. Assuming p groups of data points in which the number of elements for the kth group is n_k, $\mathbf{Q}_k$ is an $m \times m$ matrix with only n_k nonzero elements. The nonzero elements are equal to 1. They are the diagonal elements that



(a)



(b)

Fig. 5 The impact of removing outliers on **a** maximum absolute value of realized standardized CVEs and **b** realized estimated noise variance by the proposed method

start from $n_1 + n_2 + \cdots + n_{k-1}$ and end at $n_{k+1} + \cdots + n_p$. $\mathbf{C_{nn}}$ is then modeled as :

$$\mathbf{C_{nn}} = \sum_{k=1}^{p} \sigma_k^2 \mathbf{Q}_k = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_{n_2} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_p^2 \mathbf{I}_{n_p} \end{bmatrix} \quad (45)$$

To organize the data points into distinct groups, a priori information about the measurement errors is required. Such information is not available for the dataset of this research. To come up with a systematic procedure for organizing the data points within separate groups, this study proposes $\left|\hat{\underset{\sim}{e}}_i\right|$ as the required measure. If we have at least two groups of observations with different noise variance components and assume a single noise variance for the total observations, then by employing one estimated noise variance to all of observations leads to larger $\left|\hat{\underset{\sim}{e}}_i\right|$ for some of those observations that belong to the group with larger noise variance. In other words, the number of observations with large values of $\left|\hat{\underset{\sim}{e}}_i\right|$ in the group with larger noise variance would be significantly more than the other group. According to this assumption, we removed only the first 5 observation which certainly hold blunders (Fig. 5) and then we computed $\left|\hat{\underset{\sim}{e}}_i\right|$ for rest of the observations and separated those with $\left|\hat{\underset{\sim}{e}}_i\right| > 1.96$ which may be called as the noisy observations. In Fig. 6, we have marked those noisy observations after removing 5 outliers. According to unbalanced spatial distribution of the noisy observation points in the interpolation area, we conclude that there is a probability of having some different populations of observations with distinct noise variances. In Fig. 6 we see an aggregation of noisy data points in the southeastern part of the area, while the spatial distribution of these points in the southwestern part is approximately balanced. On the other hand, at the northern part of the study area we have smaller number of noisy observations. Accordingly, we decided to depart the data into three groups assuming to belong to three statistical populations with different noise variances.

Figure 7 displays the data groups that were distinguished via considering the spatial distribution of noisy data points on the leveling lines. In this figure red, blue and green points belong to those groups with larger, medium and smaller number of noisy data points, respectively.

Table 8 gives information about the number of total data points and noisy data points which stand in each of the distinguished data groups. We assume, $\hat{e}_i \sim N(0, 1)$, where for such a distribution only 5% of standardized CVEs lie out of ± 1.96 limits. As we do not expect to see an inordinate difference between NVCs of data points, we had to take more data points for the first group in comparison to the others.

After grouping data points, the NVCs can be computed using the REML method. But, this will not be end of the job; we should continue detection and removing outliers to make sure about accuracy of the final results. We do this by using NVCs estimated by the REML method each time after removing one outlier. In Fig. 8 we have presented a flowchart which has been employed for estimating the NVCs and removing outliers from the dataset.

The REML method was applied for estimating the noise variance components. The Fisher scoring technique was also used for numerical computations. For details, the interested reader is referred to Jarmołowski and Bakuła (2014) and Jarmołowski (2015). Table 9 illustrates the estimated noise standard deviations for each of the identified groups employing the REML method before and after detection and removing outliers. The used algorithm for elimination of outliers is the same as that performed by the flowchart in Fig. 8.

Estimated values of $\hat{\underset{\sim}{\sigma}}$ for the identified groups are completely different from each other. This difference is considerably increased after removing outliers. By taking a look at Table 8, we see that ratio of the number of noisy data points to total points in group I is significantly larger than that of the other groups. Therefore, we expect to see a larger $\hat{\underset{\sim}{\sigma}}$ for this group which holds more noisy data points which is due to the estimation of noise variances by the REML method. Removing gross errors also has an impressive impact on estimated noise variances. They are reduced about 29.5%, 46 and 62% for the three groups, respectively. Table 10 shows the changes of standardized CVEs and covariance parameters due to removing outliers by the proposed method which is depicted in Fig. 8.

From Table 10, it is seen that by removing outliers, $C_w(0)$ is reduced and $C_s(0)$ is not significantly changed. The correlation distance has slightly increased about 0.66 km which is a consequence of removing outliers. In comparison with Table 5, we see that RMS $(\hat{\underset{\sim}{e}}_i)$ has reached to 0.99 and 1.00 for the two cases which is related to employing the heterogeneous noise model that better fits to our dataset. Table 11 illustrates estimated values for CVEs, LSC internal errors and LSC prediction errors on a regular grid of interpolation points after employing a heterogeneous noise model for data points and removing outliers using the proposed methods. In comparison with those errors, after removing outliers for the homogeneous noise model in Table 6, all statistical measures in Table 11 reduce. The RMS of these three types of errors decreases about 4, 14 and 2%, respectively. LSC internal error reduces more than the two other errors because of the significant reduction in estimated NVCs for the groups II and III in comparison with noise variance of the homogeneous model.

Table 12 illustrates the lower cutoff CVEs after implementation of the proposed method for heterogeneous noise

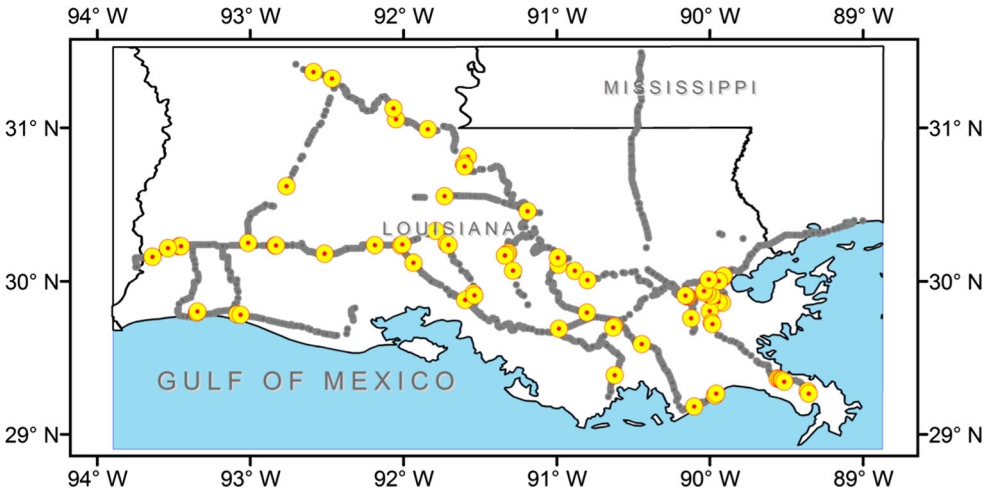


Fig. 6 Spatial distribution of data points with $\left|\hat{e}_{\sim i}\right| > 1.96$ assuming a homogeneous noise model

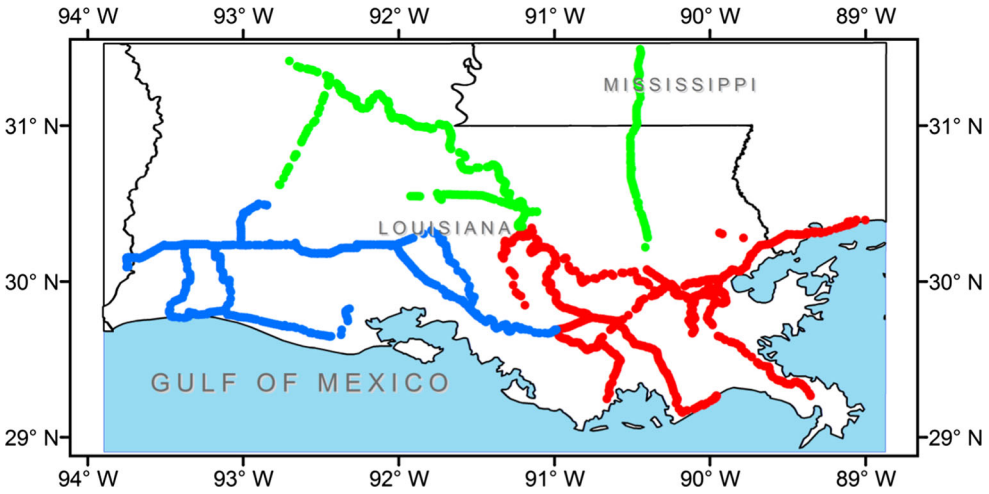


Fig. 7 Separation of data points into three groups: group I (red), group II (blue) and group III (green)

Table 8 Information about the number of data points in groups separated based on the number of noisy points distributed on the leveling lines

Data group	No data points	No noisy data points	Percent of noisy data points
Group I	754	47	6.23
Group II	530	23	4.34
Group III	394	13	3.3

Table 9 Noise standard deviations for a heterogeneous noise model estimated by the REML method, unit: mm/a

Data group	No data points before removing outliers	No data points after removing outliers	Estimated noise standard deviation $\hat{\sigma}$ before removing outliers	Estimated noise standard deviation $\hat{\sigma}$ after removing outliers
Group I	754	743	1.46	1.03
Group II	530	520	1.41	0.76
Group III	394	386	1.06	0.40

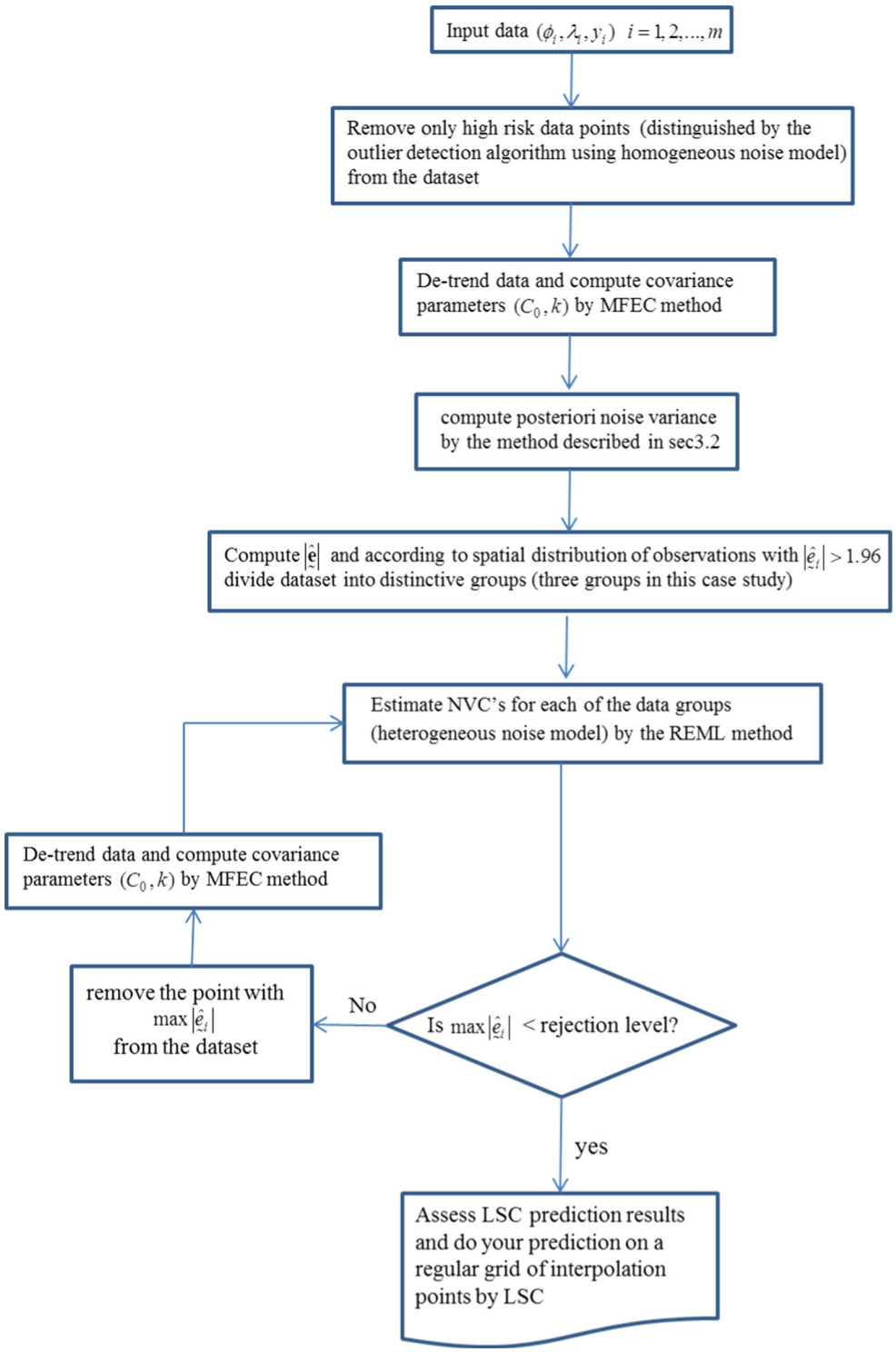


Fig. 8 Flowchart of computing NVCs together with detection and removing outliers from the dataset

Table 10 Standardized CVEs and covariance parameters before and after removing outliers using the proposed method for employing a heterogeneous noise model

Status	No data points	Max$\lvert \hat{\underset{\sim}{e}}_i \rvert$	RMS ($\hat{\underset{\sim}{e}}_i$)	$C_w(0)$ (mm/a)^2	$C_s(0)$ (mm/a)^2	d_0 (km)
Before removing outliers	1678	7.79	0.99	11.85	10.08	24.673
After removing outliers	1649	4.15	1.00	10.91	10.18	25.334

Table 11 Computed statistics for CVEs, LSC prediction errors and LSC internal errors after using heterogeneous noise model and removing outliers, units: mm/a

Type of error	No. points	Min	Max	Mean	RMS
CVE ($\hat{\underset{\sim}{\varepsilon}}_i$)	1649	− 5.56	4.70	0.00	1.21
LSC internal error ($\sigma_{\hat{y}_i}$)	1649	0.24	0.94	0.54	0.56
LSC prediction error on a regular grid of points ($\sigma_{\hat{y}_0}$)	1653	0.41	3.53	2.39	2.50

model. In comparison with Table 7 it results that this method may not be very effective on improving the overall accuracy in the prediction area for our case study. The results are a little better than the homogeneous noise model.

Our experience for this case study showed that in cases which data classification for grouping observations with different NVCs is unsuitable, lower cutoff CVEs compared with the homogeneous noise model, not only do not decrease but also increase for many of the observations. Therefore, these types of errors can be applied to check the correctness of our decision regarding the classification of observations into groups with distinct noise variances. Comparing the overall results illustrated in Tables 11 and 12 with the corresponding items for the homogeneous noise model in Tables 6 and 7 (after removing outliers), one may deduce that employing even a suitable heterogeneous noise model may not significantly improve the quality of prediction by LSC. Responding to this idea we compared the estimated noise of observations $\hat{\mathbf{n}}$ and LSC internal errors separately for each of the data groups for homogeneous and heterogeneous noise models in Table 13. The estimated noise of observations for the first group is not changed, while for the second and third groups it is obviously decreased. The RMS of the noise is reduced by 25 and 56% for the second and third groups, respectively. Maximum and minimum values are also significantly deceased. Reduction in the estimated noise of observations for these two groups indicates that we have been able to get a better adaptation between the LSC predictions and the observed values by applying the heterogeneous noise model for those areas covered by the observations of groups II and III. We also see that the RMS of internal errors is decreased by 15 and 51%, respectively, for these two groups. This implies that one can achieve a better precision of LSC prediction at observation points for these two groups. As an overall result, it concludes that implementation of a suitable heterogeneous noise model improves the regional precision of LSC prediction around the observation points.

Figure 9 depicts the interpolated rate of vertical displacements in the computation area employing a heterogeneous noise model and using estimated covariance parameters with estimated noise standard deviations listed in Tables 9 and 10 after removing outliers by running the flowchart depicted in Fig. 8.

It is obvious that rate of subsidence is increasing from north to south. In the southern region, one can recognize high rate local subsidence areas. Some of these are located at the neighborhood of the Mississippi river. Others are scattered through the southern part of this area from east to the west. According to Dokka (2011) subsidence in this area is due to the compaction of sediments from Mississippi river, oil and gas extraction, drainage projects and sea-level rise. Figure 10 illustrates the LSC prediction error which is computed on a regular grid of points with spacing $5' \times 5'$. A buffer area in green color, where the prediction error is less than 1.46 mm/a, is remarkable. In most parts of the interpolation area that is placed in between leveling lines, the error is estimated to be less than 2.5 mm/a, which could be accepted as the overall LSC precision in the region. Since the LSC prediction errors around the observation points of groups II and III are reduced with a heterogeneous noise model, the results achieved from the heterogeneous model are more precise than the homogeneous model for the areas covered by these data points. The precision is also strongly impressed by the amount of aggregation of data points. In those areas which are covered by the dark green color, we have a higher density of data points compared to the areas covered by dark brown and the following colors that suffer from the lack of data points.

Figure 11 displays the effect of employing a heterogeneous noise model versus the homogeneous one on the prediction results. Difference of the referred predictions is used for this purpose. The differences in predictions vary from − 1.51 to 1.22 mm/a. Observed differences are notable in many parts of the prediction area especially in areas covered by the leveling lines of groups II and III of observations.

Table 12 Lower cutoff CVEs for different radii of vicinity after employing a heterogeneous noise model and removing outliers detected by the proposed technique (unit: mm/a)

Radius of vicinity (km)	Min	Max	Mean	Mean of absolute value	RMS
5	− 14.64	18.75	0.06	1.89	2.75
15	− 20.01	12.57	0.03	2.64	3.68
25	− 21.12	12.97	− 0.05	3.00	4.09

Table 13 Comparison of estimated noise of observations and LSC internal errors for homogeneous and heterogeneous noise models estimated by the proposed methods (unit: mm/a)

Estimated quantity	Status	Data group	No. points	Max	Min	RMS
Estimated noise of observations ($\hat{n}_i$)	Homogeneous noise model	Group I	743	2.72	− 2.85	0.79
		Group II	523	2.81	− 3.30	0.70
		Group III	391	2.39	− 2.39	0.52
	Heterogeneous noise model	Group I	743	2.72	− 2.85	0.79
		Group II	520	1.78	− 2.17	0.52
		Group III	386	0.59	− 0.56	0.23
LSC internal error ($\sigma_{\hat{y}_i}$)	Homogeneous noise model	Group I	743	0.90	0.32	0.64
		Group II	523	0.85	0.43	0.65
		Group III	391	0.89	0.45	0.66
	Heterogeneous noise model	Group I	743	0.94	0.34	0.67
		Group II	520	0.70	0.37	0.56
		Group III	386	0.40	0.24	0.32

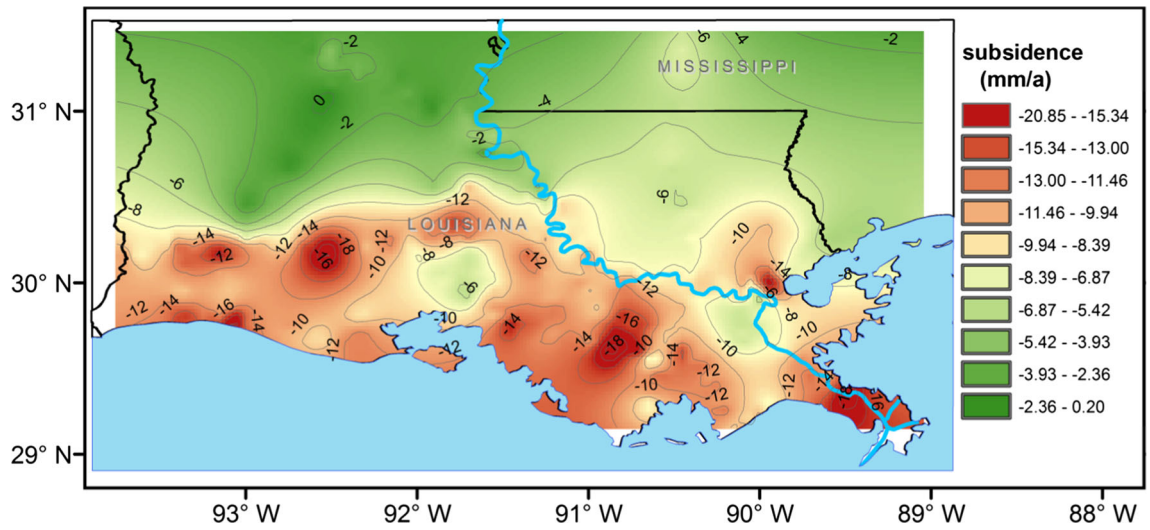


Fig. 9 Rate of vertical displacement in northern coast of the Gulf of Mexico interpolated by LSC after removing outliers and estimation of NVCs using the proposed method depicted in Fig. 8

This figure supports the influence of the heterogeneous noise model in LSC predictions for the case study of this research.

6 Conclusions

Leave-one-out cross-validation errors plays a key role in assessment of the quality of prediction by LSC. We presented two formulae for estimating the CVEs for LSC in vector form and the corresponding covariance matrix (Eqs. 13, 14). Application of these formulae extremely decreases the computation time of the CVEs and their variance–covariance information. This is especially noticeable when a massive dataset is concerned. We use the CVEs for statistical testing in LSC. We propose a global test on the quadratic form of the CVEs in order to check the reliability of the mathematical model together with the covariance parameters and observational noises. This hypothesis test, however, fails if blunders

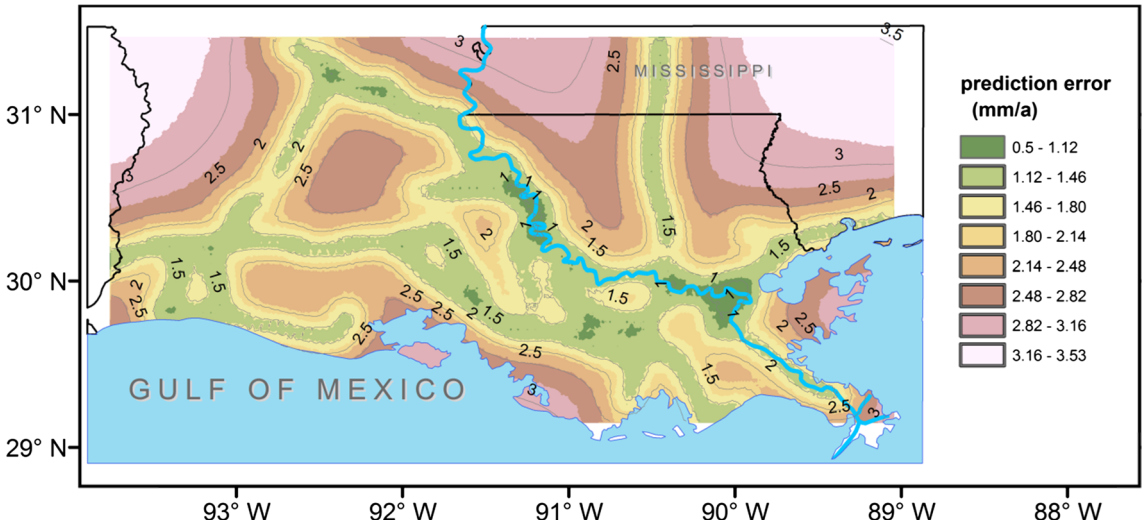


Fig. 10 LSC prediction error after removing outliers and estimation of NVCs

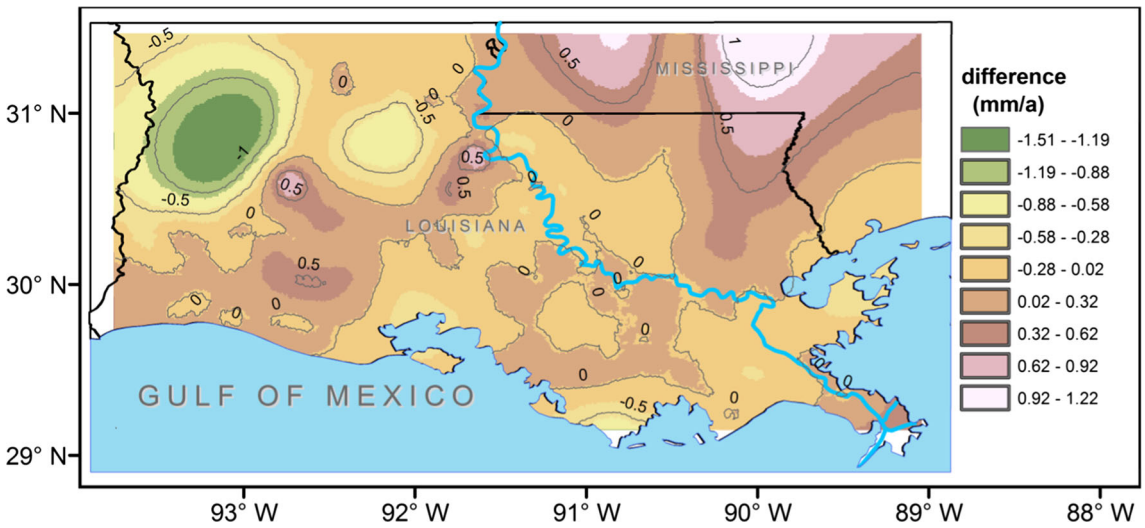


Fig. 11 Difference between predictions of the rate of vertical displacements using homogeneous and heterogeneous noise models

are present in the dataset. In order to detect blunders, statistical tests of LSA, i.e., Pope and Baarda's hypothesis tests, are applied in LSC using the CVEs. The presented statistics may be compared with the criterion proposed by Tscherning (1991a, b). However, in contrast to the Tscherning's scheme, instead of the neighboring points, we use all of the data points for computing the CVEs, where the denominator of the statistic is different in comparison to that used by Tscherning. As a side result, one can deduce that the proposed outlier detection tests are an improved version of the Tscherning's scheme for detecting the gross errors using LSC. We also proposed an iterative method for estimating a posteriori noise variance factor of the observations. By employing this method, we have computed the realized value of a posteriori noise variance for the rates of vertical displacement in the northern coast of the Gulf of Mexico. In addition, we state that the realized noise variance estimated by the MFEC method in this case study shows only about 6% difference in respect to the value estimated by the method proposed in Sect. 3.2. The results indicate that the realized value of the estimated quadratic form of the CVEs exactly matches on the degree of freedom of the LSC problem using realized noise variance factor estimated by the proposed method in Sect. 3.2 for the case study of this research.

Using standardized CVEs computed with accurately estimated noise variance, the dataset is tested for detection and removing outliers. According to the obtained results in Sect. 5, removing outliers by the proposed method illustrated by Fig. 3 significantly improves precision of the data used for the LSC prediction due to reduction in the noise variance

of observations. As a measure of precision, the estimated noise standard deviation of the dataset is reduced by 59%. There is no doubt that the accuracy of the entire dataset is also improved via the elimination of the gross errors. It has a significant influence on the estimated trend parameters which results in the reduction in the estimated residuals. In Table 5 we see that the estimated variance for the residuals is decreased about 27%. As an overall result, one concludes that the method improves the reliability of the data used for prediction by LSC.

In order to investigate the influence of filtering out blunders from the observations on precision and accuracy of LSC predictions, we computed different types of LSC errors. The CVEs and internal errors were used for assessing the accuracy and precision of the LSC prediction at the data points and their neighborhood area. However, by removing the detected outliers, a considerable reduction in the various statistics regarding the LSC internal errors is observed. As an overall measure, the RMS of CVEs and internal errors decreases 51 and 39%, respectively. Despite this, the RMS of the LSC prediction error on the regular grid of interpolation points slightly reduces with an amount of about 4%. This is a consequence of the sparse distribution of the data point locations in this case study. Note that the LSC prediction error is less decreased at the points further away from the data points compared with the near points. Anyway, we can expect more reduction in the RMS of the LSC prediction error for those datasets with a regular spatial distribution of observation points. Lower cutoff CVEs can be treated as reliable measures of accuracy for interpolation points which are distantly located from data points. Through this type of errors, we can have a fair judgment on the improvement of LSC accuracy after removing outliers. The results illustrated in Table 7 also indicate that by removing the gross errors causes a reduced RMS for lower cutoff CVEs by 19, 12 and 11% for different radii of vicinity. As an overall conclusion, we showed that the precision of LSC prediction for interpolation points distantly located from the data points is not significantly affected by removing outliers where the accuracy of the LSC prediction is more influenced. The other achievement of eliminating outliers is the considerable reduction in the estimated noise of observations. In Table 6 we showed that due to the implementation of the proposed method for the case study, the RMS of the estimated noise of observations is reduced from 2.12 to 0.70 (mm/a). This leads to a better adaptation of LSC predictions to data points. According to the better precision of data points, the smoothing property of LSC is somehow regularized in a way to give predictions which are closer to the observed values.

For the case study of this research in which a priori information on the noise variance of data points was not available, we showed that spatial distribution of the observations with large standardized CVEs (i.e., $\left|\hat{\underset{\sim}{e}}_i\right| > 1.96$) that may be called as noisy observations, can be an appropriate criterion for deciding over how to divide a dataset into subsets with expected different noise variances. We could distinguish three groups of observations via checking the number of noisy data points compared to the total number of data points in each of the groups. Implementation of the REML method by the fisher scoring technique together with elimination of outliers revealed distinct values of the NVCs for each of the groups. In comparison with the homogeneous noise model, we found 8 more outliers that lead to a better precision and accuracy of the data points. However, the RMS of CVEs and the LSC prediction error on a regular grid of interpolation points as the overall measures are decreased only 4 and 2% which is not considerable. In order to reveal the advantage of employing a heterogeneous noise model for the case study, the noise of observations and internal errors was estimated separately for each of the data groups by employing the heterogeneous noise model and compared with those of the homogeneous model. The results indicate significant reduction in the noise for two of the groups with lower level of NVCs. Therefore, one concludes that the application of a suitable heterogeneous noise model for a dataset which obviously improves precision of classified data points leads to a better quality of LSC predictions via the reduction in estimated noise of observations that implies a better adaptation of LSC predictions to observations. This is remarkable only for those observation points that lie in groups with lower NVCs contrasted with the unique noise variance of the homogeneous model. As an overall conclusion, in cases of unbalanced spatial distribution of noisy observations, employing an appropriate heterogeneous noise model based on a logical separation of data points would be an effective approach to achieve more reliable LSC predictions in comparison with homogeneous noise models. Finally, according to the discussed subjects, detection and removing outliers based on the statistical tests described in Sect. 3 is strongly recommended for any LSC prediction. Monitoring changes of various assessment measures, before and after application of any method discussed in this paper is also recommended.

Acknowledgements The US National Oceanic and Atmospheric Administration and US National Geodetic Survey are appreciated for providing access to the observational data for this research. We would like to thank the editors and reviewers for many constructive and insightful comments that lead to major improvements of the manuscript. Dr. Soheil Vasheghani is acknowledged for proofreading the English of the manuscript.

Appendix A

A lemma in linear algebra

Notation For an arbitrary matrix $\mathbf{D} = \{d_{ij}\}$, $\mathbf{d}_{i,-i}$ is the ith row of $\mathbf{D}$ whose ith element is removed and $\boldsymbol{D}_{-i,-i}$ is the same matrix whose ith row and column are removed.

Lemma *If* $\mathbf{A}$ *represents an arbitrary symmetric positive definite matrix and* $\mathbf{B} = \mathbf{A}^{-1}$*, then*

$$\mathbf{a}_{i,-i}\mathbf{A}^{-1}_{-i,-i} = -\frac{1}{b_{ii}}\mathbf{b}_{i,-i} \tag{A1}$$

Proof we define the vector $\mathbf{d}^{(i)}$ by

$$\mathbf{d}^{(i)} = \mathbf{b}_i\mathbf{A} \tag{A2}$$

where $\mathbf{b}_i$ is the ith row of $\mathbf{B}$, the kth element of $\mathbf{d}^{(i)}$ is simply derived

$$d_k^{(i)} = \sum_j b_{ij}a_{jk} = \delta_{ik} \tag{A3}$$

where δ_{ik} is the Kronecker delta. Considering the arbitrary vector $\mathbf{e}^{(i)}$ that is defined by

$$\mathbf{e}^{(i)} = \mathbf{b}_{i,-i}\mathbf{A}_{-i,-i} \tag{A4}$$

and using Eq. (A3), one can conclude that:

$$\begin{aligned} e_k^{(i)} &= \sum_{j\neq i} b_{ij}a_{jk} = d_k^{(i)} - b_{ii}a_{ik} \\ &= -b_{ii}a_{ik} \quad \forall k \neq i \end{aligned} \tag{A5}$$

Finally, the following relations are deduced from Eq. (A5)

$$\begin{aligned} a_{ik} &= \frac{-1}{b_{ii}}e_k^{(i)} \\ &= \frac{-1}{b_{ii}}\sum_{j\neq i} b_{ij}a_{jk} \quad \forall k \neq i \end{aligned} \tag{A6}$$

Therefore,

$$\mathbf{a}_{i,-i} = -\frac{1}{b_{ii}}\mathbf{b}_{i,-i}\mathbf{A}_{-i,-i} \tag{A7}$$

It has to be mentioned here that any principal submatrix of a positive definite matrix is also positive definite (Harville 1997, p. 214). Therefore, for the positive definite matrix $\mathbf{A}$, $\mathbf{A}_{-i,-i}$ is always invertible. □

Appendix B

LSC prediction errors and noise estimation

LSC prediction error at an unobserved point p_0 is computed by (Moritz 1972, p. 47; Mikhail and Ackermann 1976, p. 422)

$$\begin{aligned} \sigma^2_{\hat{y}_0} &= c_{s_0s_0} - \mathbf{c}_{s_0\mathbf{s}}\mathbf{C}^{-1}_{\mathbf{ww}}\mathbf{c}^T_{s_0\mathbf{s}} \\ &\quad + \left(\mathbf{c}_{s_0\mathbf{s}}\mathbf{C}^{-1}_{\mathbf{ww}}\mathbf{A} - \mathbf{a}_0\right)\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}\left(\mathbf{c}_{s_0\mathbf{s}}\mathbf{C}^{-1}_{\mathbf{ww}}\mathbf{A} - \mathbf{a}_0\right)^T \end{aligned} \tag{B1}$$

where $\hat{y}_0$ is prediction of y at p_0 and $c_{s_0s_0}$ is the signal variance, $\mathbf{c}_{s_0\mathbf{s}}$ is the cross-covariance vector of the predicted point and the vector of data points, $\mathbf{a}_0$ is the vector of trend for predicted point, and $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$ denotes the covariance matrix of estimated trend parameters which is computed by the following formula

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \left(\mathbf{A}^T\mathbf{C}^{-1}_{\mathbf{ww}}\mathbf{A}\right)^{-1} \tag{B2}$$

LSC internal error (adopted from Darbeheshti and Featherstone 2009) is LSC prediction error at an observed point p_i

$$\begin{aligned} \sigma^2_{\hat{y}_i} &= c_{s_is_i} - \mathbf{c}_{s_i\mathbf{s}}\mathbf{C}^{-1}_{\mathbf{ww}}\mathbf{c}^T_{s_i\mathbf{s}} \\ &\quad + \left(\mathbf{c}_{s_i\mathbf{s}}\mathbf{C}^{-1}_{\mathbf{ww}}\mathbf{A} - \mathbf{a}_i\right)\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}\left(\mathbf{c}_{s_i\mathbf{s}}\mathbf{C}^{-1}_{\mathbf{ww}}\mathbf{A} - \mathbf{a}_i\right)^T \end{aligned} \tag{B3}$$

where $\hat{y}_i$ is prediction of y at p_i and $c_{s_is_i}$ is the signal variance, $\mathbf{c}_{s_i\mathbf{s}}$ is the cross-covariance vector of the predicted point and the vector of data points, $\mathbf{a}_i$ is the ith row of $\mathbf{A}$.

Noise of the observations in Eq. (1) is always unknown. It can be estimated by the following formula (Moritz 1980, p. 119)

$$\hat{\mathbf{n}} = \mathbf{C}_{\mathbf{nn}}\mathbf{C}^{-1}_{\mathbf{ww}}\left(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\right) \tag{B4}$$

References

Amiri-Simkooei A (2007) Least-squares variance component estimation: theory and GPS applications. Doctoral dissertation, TU Delft, Delft

Arabelos DN, Forsberg R, Tscherning CC (2007) On the a priori estimation of collocation error covariance functions: a feasibility study. Geophys J Int 170:527–533

Baarda W (1968) A testing procedure for use in geodetic networks. In: Geodesy, New series, vol 2. issue 5, Netherlands Gedetic Commission, Delft

Burden RL, Faires JD (2011) Numerical analysis, 9th edn. Brooks/Cole, Pacific Grove

Darbeheshti N, Featherstone WE (2009) Non-stationary covariance function modelling in 2D least-squares collocation. J Geod 83(6):495–508

Dokka RK (2011) The role of deep processes in late 20th century subsidence of New Orleans and coastal areas of southern Louisiana

and Mississippi. J Geophys Res Solid Earth 116:B06403. https://doi.org/10.1029/2010jb008008
El-Fiky G, Kato T, Fuji Y (1997) Distribution of vertical crustal movement rates in the Tohoku district, Japan, predicted by least-squares collocation. J Geod 71(7):432–442
Eshagh M, Sjöberg LE (2011) Determination of gravity anomaly at sea level from inversion of satellite gravity gradiometric data. J Geodyn 51(5):366–377
Featherstone WE, Sproule DM (2006) Fitting AusGeoid98 to the Australian height datum using GPS-levelling and least squares collocation: application of a cross-validation technique. Surv Rev 38(301):573–582
Grafarend EW (1976) Geodetic applications of stochastic processes. Phys Earth Planet Inter 12(3):151–179
Grodecki J (1999) Generalized maximum-likelihood estimation of variance components with inverted gamma prior. J Geod 73(7):367–374
Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, New York
Jarmołowski W (2013) A priori noise and regularization in least squares collocation of gravity anomalies. Geod Cartogr 62(2):199–216
Jarmołowski W (2015) Least squares collocation with uncorrelated heterogeneous noise estimated by restricted maximum likelihood. J Geod 89(6):577–589
Jarmołowski W, Bakuła M (2014) Precise estimation of covariance parameters in least-squares collocation by restricted maximum likelihood. Stud Geophys Geod 58(2):171–189
Kitanidis PK (1983) Statistical estimation of polynomial generalized covariance functions and hydrologic applications. Water Resour Res 19(4):909–921
Koch KR (1977) Least squares adjustment and collocation. Bull Geod 51(2):127–135
Koch KR (1986) Maximum likelihood estimate of variance components. Bull Geod 60(4):329–338
Koch KR (1999) Parameter estimation and hypothesis testing in linear models, 2nd edn. Springer, Berlin
Koch KR (2007) Introduction to Bayesian statistics, 2nd edn. Springer, New York
Koch KR, Kusche J (2002) Regularization of geopotential determination from satellite data by variance components. J Geod 76(5):259–268
Krakiwsky EJ, Biacs ZF (1990) Least squares collocation and statistical testing. Bull Geod 64(1):73–87
Krarup T (1969) A contribution to the mathematical foundation of physical geodesy, pub. 44. Dan Geod Inst, Copenhagen
Kusche J, Klees R (2002) Regularization of gravity field estimation from satellite gravity gradients. J Geod 76(6–7):359–368
Mikhail EM, Ackermann F (1976) Observations and least squares. Harper and Row, New York
Moritz H (1962) Interpolation and prediction of gravity and their accuracy, rep. 24. Inst Geod Phot Cart, Ohio State University, Columbus
Moritz H (1972) Advanced least-squares methods, vol 175. Department of Geodetic Science, Ohio State University, Columbus
Moritz H (1980) Advanced physical geodesy. Herbert Wichmann Verlag, Karlsruhe
Pope AJ (1976) The statistics of residuals and the detection of outliers. NOAA technical report NOS 65 NGS 1
Rummel R, Schwarz KP, Gerstl M (1979) Least squares collocation and regularization. Bull Geod 53(4):343–361
Sadiq M, Tscherning CC, Ahmad Z (2009) An estimation of the height system bias parameter N0 using least squares collocation from observed gravity and GPS-levelling data. Stud Geophys Geod 53(3):375–388
Schaffrin B (2001) Softly unbiased prediction. Part 2: the random effects model. Boll Geod Sci Affini 60(1):49–62
Shinkle KD, Dokka RK (2004) Rates of vertical displacement at benchmarks in the lower Mississippi Valley and the northern Gulf Coast, US Department of Commerce NOAA technical report NOS/NGS 50
Snow KB (2012) Topics in total least-squares adjustment within the errors-in-variables model: singular cofactor matrices and prior information. Doctoral dissertation, The Ohio State University
Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, New York
Teunissen PJG (2000) Testing theory an introduction. Series on mathematical geodesy and positioning. Delft University Press, Delft
Tscherning CC (1991a) Strategy for gross-error detection in satellite altimeter data applied in the Baltic-sea area for enhanced geoid and gravity determination. Determination of the geoid. Springer, New York, pp 95–107
Tscherning CC (1991b) The use of optimal estimation for gross-error detection in databases of spatially correlated data. Bull d'Inf 68:79–89
Vaníček P, Krakiwsky EJ (1986) Geodesy: the concepts. North Holland, Amsterdam
Vestøl O (2006) Determination of postglacial land uplift in Fennoscandia from leveling, tide-gauges and continuous GPS stations using least squares collocation. J Geod 80(5):248–258
Wei M (1987) Statistical problems in collocation. Manuscr Geod 12:282–289
Yang Y, Zeng A, Zhang J (2009) Adaptive collocation with application in height system transformation. J Geod 83(5):403–410