J Geod (2005) 79: 341–350
DOI 10.1007/s00190-005-0471-0

ORIGINAL ARTICLE

C. Kotsakis

A type of biased estimators for linear models with uniformly biased data

Received: 11 November 2004 / Accepted: 17 May 2005 / Published online: 14 July 2005
© Springer-Verlag 2005

Abstract The objective of this paper is the comparison of various types of estimators that can be used in linear models with uniformly biased data. This particular case refers to adjustment problems where the available measurements are affected by a common, unknown and uniform offset. The classic least-squares (LS) unbiased estimators for this type of models are reviewed in detail, and some additional remarks on their properties and performance are given. Furthermore, a family of biased estimators for linear models with uniformly biased data is introduced, which has the potential to provide better performance (in terms of mean squared estimation error) than the ordinary LS unbiased solutions. A number of different regularization viewpoints that can be equivalently associated with these biased estimators are presented, along with a discussion on various selection strategies that can be employed for the choice of the regularization parameter that enters into the biased estimation algorithm.

Keywords Linear model · Least squares estimation · Uniformly biased data · Biased estimation · Regularization

1 Introduction

A problem often encountered in the analysis of geodetic measurements is the presence of unmodeled systematic effects and biases in their values. In fact, most geodetic observations are carried out under complex physical conditions, which may not correspond exactly to the mathematical models that we often employ for their analysis. Moreover, any model is only an image of physical reality that encompasses some level of abstraction and simplification, whereas real-world measurements can capture additional details that are not necessarily included in the original model formulation.

With the intention of studying some of the problems associated with the optimal inversion of biased data sets, let us start with the usual linear(-ized) system of observation equations that is commonly employed in geodesy for the estimation of an unknown parameter vector $\mathbf{x}$ from a set of noisy measurements $\mathbf{y}$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}\,, \tag{1}$$

where $\mathbf{A}$ is a matrix of known coefficients, and $\mathbf{v}$ is a vector of zero-mean random errors with a covariance (CV) matrix $\mathbf{C}$. Obviously, under the presence of unknown systematic effects in the input data, the previous model needs to be generally modified as follows:

$$\mathbf{y}' = \mathbf{A}\mathbf{x} + \mathbf{b} + \mathbf{v}\,, \tag{2}$$

where $\mathbf{b}$ denotes the external systematic disturbances. The observation vector is now denoted by $\mathbf{y}'$, in order to distinguish it from the vector of non-biased data $\mathbf{y}$ that is used in Eq. (1). Although the previous model assumes only *additive* biases, it can still cover most geodetic applications where external disturbances exist in the input data. Three important application areas should be mentioned in connection with the general model of Eq. (2), namely the analysis of the linearization error in non-linear models, the analysis of the aliasing error when a continuous (e.g., potential) field is a approximated by a finite-dimensional parametric model using a set of discrete observations, and the data outlier detection in linear models (Kusche 2004).

The sole inclusion of the bias term in Eq. (2) is not sufficient to generate a feasible inversion scheme for the optimal estimation of the model parameters $\mathbf{x}$ from the biased measurements $\mathbf{y}'$. Indeed, the least-squares (LS) principle, $\mathbf{v}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{v} = (\mathbf{y}' - \mathbf{A}\mathbf{x} - \mathbf{b})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\mathbf{x} - \mathbf{b}) = \min$, yields a singular system of normal equations, with an infinite number of solutions for the parameter vector $\mathbf{x}$ and the bias term $\mathbf{b}$. Essentially, the data $\mathbf{y}'$ do not really 'know' yet how to distinguish between the random disturbances ($\mathbf{v}$) and the systematic disturbances ($\mathbf{b}$).

C. Kotsakis
Department of Geodesy and Surveying,
Aristotle University of Thessaloniki,
University Box 440,
Thessaloniki 54124, Greece
Tel.: +30-2310-996150
Fax: +30-2310-995948
E-mail: kotsaki@topo.auth.gr

The singularity of the system of observation equations in Eq. (2) signifies the need for further modeling of the bias term **b**. Bias modeling and elimination has been a topic of continuous research interest in geodesy by both theoreticians and practitioners. An extensive discussion, related mostly to the mathematical details of bias treatment in geodetic data analysis, can be found in Kukuča (1987). A comparison of different approaches for dealing with systematic effects that arise from the integration of heterogeneous geodetic data sets is given in Schaffrin and Baki-Iz (2001). A theoretical analysis of the nuisance parameter elimination problem within the standard Gauss-Markov linear model, along with some practical aspects for GPS data processing, have been presented in Schaffrin and Grafarend (1986). Other studies that have appeared in the geodetic literature with a focus on the problem of bias modeling and elimination, include Tscherning and Knudsen (1986); Kubáčková and Kubáček (1993); Gaspar et al. (1994); Satirapod et al. (2003); Lerch (1991); Jia et al. (2000) and Aduol (1987), among others.

In this paper, we compare a number of alternative estimation schemes that can be implemented in linear models with *uniformly biased* data. This type of problem refers to cases where all measurements are affected by a common, unknown and uniform offset. Consequently, the bias vector is assumed to take the parametric form $\mathbf{b} = \beta\mathbf{s}$, where $\mathbf{s}$ is a vector of ones and β is a scalar parameter that accounts for all "zero-order" systematic effects in the input data. In this way, the model of Eq. (2) becomes

$$\mathbf{y}' = \mathbf{A}\mathbf{x} + \beta\mathbf{s} + \mathbf{v}\,. \tag{3}$$

A review of two well-known types of optimal unbiased estimators that can be associated with the above model is given in Sect. 2, along with some additional remarks on their properties and performance. In particular, we examine the LS inversion of Eq. (3) using a fixed-effects or a mixed-effects model, depending whether the bias parameter β is viewed as a deterministic or a stochastic quantity. In Sect. 3, a family of biased estimators for the model of Eq. (3) is introduced, which is capable of providing better accuracy (in terms of mean squared estimation error) than the ordinary LS unbiased solution. In Sect. 4, we present alternative regularization viewpoints that can be equivalently associated with the aforementioned biased estimators, and in Sect. 5 we discuss several criteria that can be used in practice for the choice of the regularization parameter that appears in the biased estimation formulae for $\mathbf{x}$ and β. Finally, some concluding remarks and suggestions for future work are given in Sect. 6.

2 Least-squares (unbiased) estimators for the general model of equation (3)

The scope of this section is to review the two main types of optimal unbiased estimators that can be associated with a linear model that uses uniformly biased data. Specifically, the LS inversion for the linear model of Eq. (3) is presented from two distinct viewpoints, depending on the interpretation (deterministic or stochastic) that we choose to assign to the bias parameter β. The derived estimators can be considered as special cases of other, more general LS-based estimators, which have been developed and studied thoroughly in the literature for the case of the extended linear model $\mathbf{y}' = \mathbf{A}\mathbf{x} + \mathbf{G}\mathbf{z} + \mathbf{v}$ (see, e.g., Koch 1999; Grafarend and Schaffrin 1993; Dermanis 1979, 1991).

2.1 Fixed-effects linear model with uniformly biased data

In this approach, the data bias β corresponds to a fixed unknown parameter that needs to be estimated, along with the other model parameters $\mathbf{x}$, through an integrated procedure. This type of "fixed-effects" model is summarized in Box 1.

Box 1 Fixed-effects linear model with uniformly biased data

$\mathbf{y}' = \mathbf{A}\mathbf{x} + \beta\mathbf{s} + \mathbf{v}$	$\mathbf{s} = [1 \ldots 1]^{\mathrm{T}}$
$E\{\mathbf{v}\} = \mathbf{0}$	$\mathbf{x}$: deterministic model parameters
$E\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\} = \mathbf{C}$	β : deterministic data bias

Applying the LS optimal inversion principle

$$\begin{aligned}\mathbf{v}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{v} &= (\mathbf{y}' - \mathbf{A}\mathbf{x} - \beta\mathbf{s})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\mathbf{x} - \beta\mathbf{s}) \\ &= \text{minimum}\,,\end{aligned} \tag{4}$$

yields the following system of normal equations:

$$\begin{bmatrix} \mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A} & \mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} \\ \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A} & \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}^{\mathrm{FE}} \\ \hat{\beta}^{\mathrm{FE}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}' \\ \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}' \end{bmatrix}\,. \tag{5}$$

The superscript FE indicates that the corresponding estimates refer to the LS inversion of the *fixed-effects* model. If we further assume that the partitioned design matrix $[\mathbf{A} \quad \mathbf{s}]$ has full column rank, then the above system has a unique solution that can be expressed as

$$\hat{\mathbf{x}}^{\mathrm{FE}} = \hat{\mathbf{x}}^{\mathrm{o}} - \dot{k}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})\xi \tag{6}$$

$$\hat{\beta}^{\mathrm{FE}} = \dot{k}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})\,, \tag{7}$$

where the auxiliary quantities $\hat{\mathbf{x}}^{\mathrm{o}}$, ξ and $\dot{k}$ are defined by the equations

$$\hat{\mathbf{x}}^{\mathrm{o}} = (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}' \tag{8}$$

$$\xi = (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} \tag{9}$$

$$\dot{k} = \frac{1}{\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} - \xi^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})\xi}\,. \tag{10}$$

The above estimation algorithm provides a special case of a well-known procedure in geodetic data analysis, which is generally known as *nuisance parameter elimination* (e.g., Teunissen 2000, Chap. 6).

Remark 1 The vector $\hat{\mathbf{x}}^{\mathrm{o}}$ corresponds to the LS solution that we would obtain if we ignored the bias presence in the input

data $\mathbf{y}'$. It gives a biased estimate of the true unknown parameter $\mathbf{x}$, since

$$\begin{aligned} E\{\hat{\mathbf{x}}^{\mathrm{o}}\} &= (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}E\{\mathbf{y}'\} \\ &= (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{A}\mathbf{x}+\beta\mathbf{s}) \\ &= \mathbf{x}+\beta\xi \\ &\neq \mathbf{x}\,. \end{aligned} \tag{11}$$

Remark 2 The vector ξ identifies a characteristic quantity for the fixed-effects model shown in Box 1. It can be directly computed from the known matrices $\mathbf{A}$ and $\mathbf{C}$, without the knowledge of the actual data $\mathbf{y}'$. It gives a normalized measure of the distortion that a uniform data bias would cause on the LS solution, if β is left out of the model formulation.

Remark 3 It is often claimed that systematic biases can be neglected in the geodetic practice, if their magnitude is below the noise level of the input data. In such situations, it is assumed that the systematic effects are entirely absorbed by the noise component $\mathbf{v}$, without the need to modify the functional and/or the stochastic part of the model. Nevertheless, it is interesting to point out that the difference between the LS estimates for $\mathbf{x}$ that are obtained from: (1) non-biased data, and (2) uniformly biased data with the same noise level as in (1) but ignoring the bias presence in the model formulation, is not necessarily affected by the bias-to-noise ratio. Indeed, if we consider the case where $\mathbf{C}=\sigma^2\mathbf{I}$ and substitute Eq. (3) into Eq. (8), we get the following relationship

$$\underbrace{\hat{\mathbf{x}}^{\mathrm{o}}}_{\substack{\text{LS solution from uniformly biased data}\\ \text{(ignoring the bias presence)}}} = \underbrace{(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\overbrace{(\mathbf{A}\mathbf{x}+\mathbf{v})}^{\text{non-biased data}}}_{\text{LS solution from non-biased data}} + \beta(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{s}\,. \tag{12}$$

From Eq. (12), it is seen that the closeness of $\hat{\mathbf{x}}^{\mathrm{o}}$ to the LS solution that is obtained from strictly non-biased data does not depend on the bias-to-noise ratio β/σ^2.

Remark 4 The scalar quantity $\dot{k}$ in Eq. (10) always has a positive value. This is easily verified by analyzing the denominator of Eq. (10) as follows:

$$\begin{aligned} &\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} - \xi^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})\xi \\ &\overset{\text{Eq. (9)}}{=} \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} - \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A}) \\ &\qquad \times(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} \\ &= \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} - \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} \\ &= \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\left[\mathbf{I}-\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\right]\mathbf{s} \\ &= \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\left[\mathbf{C}-\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\right]\mathbf{C}^{-1}\mathbf{s} \\ &= \mathbf{a}^{\mathrm{T}}\mathbf{R}\mathbf{a}\,. \end{aligned} \tag{13}$$

Since the data CV matrix $\mathbf{C}$ is a symmetric positive-definite matrix, the term $\mathbf{R}=\mathbf{C}-\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}$ also corresponds to a positive-definite symmetric matrix for any full column rank matrix $\mathbf{A}$ (Harville 1997). As a result, the quadratic form $\mathbf{a}^{\mathrm{T}}\mathbf{R}\mathbf{a}$ always attains positive values for any non-zero vector $\mathbf{a}$, including the case where $\mathbf{a}=\mathbf{C}^{-1}\mathbf{s}$.

The mean squared estimation error for the parameter vector $\mathbf{x}$ is defined in terms of the mean squared error (MSE) matrix

$$\mathrm{MSE}(\hat{\mathbf{x}}^{\mathrm{FE}}) = E\left\{(\hat{\mathbf{x}}^{\mathrm{FE}}-\mathbf{x})(\hat{\mathbf{x}}^{\mathrm{FE}}-\mathbf{x})^{\mathrm{T}}\right\}\,, \tag{14}$$

whereas the mean squared estimation error for the bias parameter β corresponds to the scalar quantity

$$\mathrm{MSE}(\hat{\beta}^{\mathrm{FE}}) = E\left\{(\hat{\beta}^{\mathrm{FE}}-\beta)^2\right\}\,. \tag{15}$$

Using standard matrix calculus, it is easily proven that

$$\mathrm{MSE}(\hat{\mathbf{x}}^{\mathrm{FE}}) = (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1} + \dot{k}\xi\xi^{\mathrm{T}}\,, \tag{16}$$

and

$$\mathrm{MSE}(\hat{\beta}^{\mathrm{FE}}) = \dot{k}\,. \tag{17}$$

Some additional comments regarding the MSE performance of the unbiased LS estimators $\hat{\mathbf{x}}^{\mathrm{FE}}$ and $\hat{\beta}^{\mathrm{FE}}$ are given in Sect. 2.3.

2.2 Mixed-effects linear model with uniformly biased data

In this approach, the data bias parameter β is modeled as a random variable, in contrast to the other model parameters $\mathbf{x}$ which refer to fixed (deterministic) quantities. The formulation of such a "mixed-effects" model is analytically given in Box 2. Note that the stochastic bias parameter β is assumed to have a zero mean and also to be uncorrelated with the random noise in the input measurements.

Box 2 Mixed-effects linear model with a common stochastic bias in the data

$\mathbf{y}' = \mathbf{A}\mathbf{x}+\beta\mathbf{s}+\mathbf{v}$	$\mathbf{s}=[1 \quad \ldots \quad 1]^{\mathrm{T}}$
$E\{\mathbf{v}\}=\mathbf{0}$	$\mathbf{x}$: deterministic model parameters
$E\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\}=\mathbf{C}$	β : stochastic data bias
$E\{\beta\}=0,\ E\{\beta^2\}=\sigma_\beta^2,$	$E\{\beta\mathbf{v}\}=\mathbf{0}$

The LS optimal inversion of the mixed-effects model can be performed in two different, yet equivalent, ways. In particular, we can either follow a two-step procedure as described in Dermanis and Rummel (2000, pp. 52–53), or we can adopt a single generalized LS criterion of the form

$$\begin{aligned} \mathbf{v}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{v} + \frac{\beta^2}{\sigma_\beta^2} &= (\mathbf{y}'-\mathbf{A}\mathbf{x}-\beta\mathbf{s})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}'-\mathbf{A}\mathbf{x}-\beta\mathbf{s}) + \frac{\beta^2}{\sigma_\beta^2} \\ &= \text{minimum}\,. \end{aligned} \tag{18}$$

Assuming that the design matrix $\mathbf{A}$ has full column rank and that the 'noise + bias' total CV matrix $(\mathbf{C}+\sigma_\beta^2\mathbf{s}\mathbf{s}^{\mathrm{T}})$ is invertible, the LS solution of the mixed-effects model can be expressed as

$$\hat{\mathbf{x}}^{\mathrm{ME}} = \hat{\mathbf{x}}^{\mathrm{o}} - \ddot{k}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}'-\mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})\xi \tag{19}$$

$$\hat{\beta}^{\mathrm{ME}} = \ddot{k}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}'-\mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})\,, \tag{20}$$

where the quantities $\hat{\mathbf{x}}^{\mathrm{o}}$ and ξ have been defined previously in Eqs. (8) and (9). The scalar term $\ddot{k}$ differs from the quantity $\dot{k}$ that was used in the corresponding LS solution for the fixed-effects model, and it is given by the formula

$$\ddot{k} = \frac{1}{\frac{1}{\sigma_\beta^2} + \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} - \xi^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})\xi} \,. \tag{21}$$

Note that Eq. (19) provides an unbiased estimate for the deterministic model parameter ($E\{\hat{\mathbf{x}}^{\mathrm{ME}}\} = \mathbf{x}$), whereas Eq. (20) provides an unbiased prediction for the stochastic bias β ($E\{\hat{\beta}^{\mathrm{ME}}\} = E\{\beta\}$).

Using standard rules of matrix calculus, it can be shown that

$$\mathrm{MSE}(\hat{\mathbf{x}}^{\mathrm{ME}}) = (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1} + \ddot{k}\xi\xi^{\mathrm{T}} \,, \tag{22}$$

and

$$\mathrm{MSE}(\hat{\beta}^{\mathrm{ME}}) = \ddot{k} \,. \tag{23}$$

Some remarks about the MSE performance of the estimators $\hat{\mathbf{x}}^{\mathrm{ME}}$ and $\hat{\beta}^{\mathrm{ME}}$ are given in the following section.

2.3 Remarks

(1) If we take into account Eqs. (10) and (21), the following equation is obtained

$$\ddot{k} = \dot{k}\frac{\sigma_\beta^2}{\sigma_\beta^2 + \dot{k}} = \frac{\dot{k}}{1 + \frac{\dot{k}}{\sigma_\beta^2}} \,. \tag{24}$$

Using Eq. (24), the relationship between the LS solutions for the fixed-effects and the mixed-effects linear models can now be described by

$$\hat{\mathbf{x}}^{\mathrm{ME}} = \frac{1}{1 + \frac{\dot{k}}{\sigma_\beta^2}}\hat{\mathbf{x}}^{\mathrm{FE}} + \frac{1}{1 + \frac{\sigma_\beta^2}{\dot{k}}}\hat{\mathbf{x}}^{\mathrm{o}} \tag{25}$$

$$\hat{\beta}^{\mathrm{ME}} = \frac{1}{1 + \frac{\dot{k}}{\sigma_\beta^2}}\hat{\beta}^{\mathrm{FE}} \,. \tag{26}$$

It is seen that the two models give identical results when the variance of the stochastic bias β is equal to infinity ($\sigma_\beta^2 = \infty \rightarrow \ddot{k} = \dot{k}, \hat{\mathbf{x}}^{\mathrm{ME}} = \hat{\mathbf{x}}^{\mathrm{FE}}, \hat{\beta}^{\mathrm{ME}} = \hat{\beta}^{\mathrm{FE}}$). This equivalency represents a well-known result, which asserts that a mixed-effects model becomes identical (in terms of its LS inversion results) with a fixed-effects model when the statistical uncertainty of its stochastic parameters is set equal to infinity (Leibelt 1967; Dermanis 1976).

(2) The stochastic treatment of an unknown data bias can be equivalently viewed as a LS inversion of a fixed-effects model accompanied by an appropriate *downweighting* of the input data. Indeed, the estimate $\hat{\mathbf{x}}^{\mathrm{ME}}$ in Eq. (19) can be alternatively obtained by using the following LS principle for the optimal inversion of Eq. (1)

$$\mathbf{v}^{\mathrm{T}}\tilde{\mathbf{C}}^{-1}\mathbf{v} = \text{ minimum} \,, \tag{27}$$

where $\tilde{\mathbf{C}}$ is a downweighted modification of the CV matrix for the pure random noise in the input data

$$\tilde{\mathbf{C}} = \mathbf{C} + \sigma_\beta^2\mathbf{s}\mathbf{s}^{\mathrm{T}} \,. \tag{28}$$

The use of *downweighting* as a method to account for unmodeled systematic effects in the analysis of geodetic measurements has been already discussed in Lerch (1991). In his study, it was suggested that the analysis of satellite-tracking data with a constant unknown bias should be made by downscaling the diagonal elements of their CV matrix with an appropriate factor corresponding to the number of tracking points within each satellite orbital pass. A more recent study on data downweighting as an alternative technique for the optimal combination of heterogeneous geodetic observations with partial systematic inconsistencies can be found in Schaffrin and Baki-Iz 2001.

(3) The statistical accuracy of the LS estimate for $\mathbf{x}$ that is obtained from a biased data set $\mathbf{y}'(= \mathbf{A}\mathbf{x}+\beta\mathbf{s}+\mathbf{v})$ is *lower* than the statistical accuracy of the corresponding LS estimate obtained from a non-biased data set $\mathbf{y}(= \mathbf{A}\mathbf{x} + \mathbf{v})$. In the latter case, the MSE matrix of the LS estimate for $\mathbf{x}$ is equal to $(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}$. Since both $\dot{k}$ and $\ddot{k}$ have positive values, the diagonal elements of $\mathrm{MSE}(\hat{\mathbf{x}}^{\mathrm{FE}})$ and $\mathrm{MSE}(\hat{\mathbf{x}}^{\mathrm{ME}})$ are always larger than the corresponding elements of $(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}$; see Eqs. (16) and (22).

(4) The accuracy degradation for the LS estimate of $\mathbf{x}$ is smaller in the case of the mixed-effects model, since it holds that $\ddot{k} < \dot{k}$; see Eq. (24). As a result, we have that $\mathrm{traceMSE}(\hat{\mathbf{x}}^{\mathrm{ME}}) \leq \mathrm{traceMSE}(\hat{\mathbf{x}}^{\mathrm{FE}})$. Given that the fixed-effects model is equivalent to the mixed-effects model for $\sigma_\beta^2 = \infty$, the accuracy improvement for $\hat{\mathbf{x}}^{\mathrm{ME}}$ (compared to $\hat{\mathbf{x}}^{\mathrm{FE}}$) should be attributed to the lower uncertainty for the data bias that is associated with the mixed-effects model approach ($\sigma_\beta^2 < \infty$).

(5) Another interesting point is that the degradation in the LS estimation accuracy for $\mathbf{x}$, in the case of the fixed-effects model, is completely independent of the magnitude of the data bias; i.e., the MSE matrix of $\hat{\mathbf{x}}^{\mathrm{FE}}$ does not depend on β. This means that, regardless of the actual magnitude of the constant bias that has affected the input measurements, the reduction in the statistical accuracy of the LS solution $\hat{\mathbf{x}}^{\mathrm{FE}}$ will always be the same.

3 Biased estimation in fixed-effects linear models with uniformly biased data

3.1 Motivation

The mixed-effects model appears to be the preferred approach for the LS inversion of uniformly biased data since it leads to optimal parameter estimates with better accuracy than the fixed-effects model. The key factor that controls the MSE improvement for $\hat{\mathbf{x}}^{\mathrm{ME}}$ and $\hat{\beta}^{\mathrm{ME}}$ is the variance of the bias effect in the input data. It is clear from Eq. (21) or Eq. (24) that the factor $\ddot{k}$ becomes smaller (relative to $\dot{k}$) as the variance

σ_β^2 decreases, thus resulting in MSE reduction for both $\hat{\mathbf{x}}^{\mathrm{ME}}$ and $\hat{\beta}^{\mathrm{ME}}$. The limit case $\sigma_\beta^2 = 0$ implies that we essentially have complete knowledge of the bias effect, which in principle makes possible the a-priori correction of the input data, without the need to include any additional nuisance parameter into the LS adjustment process.

Nevertheless, the adoption of a stochastic interpretation for the data bias β creates theoretical and practical concerns for most geodetic applications. Some of these concerns can be summarized in terms of the following three questions:

1. Is it always possible (from a physical point of view) to accept a stochastic interpretation for systematic effects in the data?
2. Even if we are willing to accept a stochastic treatment for the bias parameter β, how can we obtain a value for the bias variance σ_β^2 in order to implement in practice the LS inversion of the mixed-effects model?
3. What complications arise when the mixed-effects model in Box 2 incorporates a non-zero mean for the stochastic bias parameter? Is it not more realistic to consider the case $E\{\beta\} \neq 0$ in practice?

The answer to the first question depends more on the modeling preferences that we are inclined to follow, and much less on the existence of some hidden physical mechanism that underlies the true behaviour of data biases. The dilemma of using a deterministic or a stochastic representation for (some or all of) the unknown quantities that are involved in the LS estimation process has stimulated interesting theoretical arguments among scientists (Moritz 1980; Dermanis and Sansò 1993; Scales and Snieder 1998), without however seriously hampering the geodetic practice over the years. After all, it does not really matter whether nature admits truly random signals or not, unless you are doing quantum mechanics (Moritz 1997).

The second question signifies a much more important issue from a practical viewpoint. A possible solution to this problem is to follow a variance component estimation (VCE) approach within the mixed-effects model of Box 2, where the unknown variance σ_β^2 of the stochastic bias β is estimated a-posteriori from the available (biased) data with the help of some VCE technique. This particular topic has been discussed in Kusche (2003) and Schaffrin and Baki-Iz (2001).

The third question is also an important one, since the assumption of zero-mean for the stochastic bias parameter provides a rather strong (and possibly unrealistic) restriction. The LS inversion of the mixed-effects model, in the general case $E\{\beta\} = \mu_\beta \neq 0$, results in optimal estimators that depend on the expected value μ_β; see, e.g., Dermanis (1991). Such a result makes the optimal estimation process non-feasible in practice, since the value μ_β is generally unknown.

In order to overcome the concerns raised with the above questions, and also to provide an alternative to the "stochastic regularization" that is embedded in the mixed-effects model of Box 2, a biased estimation algorithm for the inversion of uniformly biased data is proposed and discussed in the sequel.

3.2 Formulation of the biased estimators

Our aim is to exploit the advantage of the estimators in Eqs. (19) and (20), which improve the estimation accuracy for the parameter vector $\mathbf{x}$ and the bias parameter β with respect to the LS (unbiased) solution of the fixed-effects model in Box 1. At the same time, we want to abandon the stochastic interpretation for the data bias that is implied by the mixed-effects model approach. Hence, we apply the following modifications to the estimators of Eqs. (19) and (20):

- remove the stochastic interpretation from the bias parameter β;
- replace the bias variance σ_β^2 with a positive scalar term λ, which should now be seen as an arbitrary regularization factor.

The resulting estimators are now associated with the fixed-effects model of Box 1 and they are given by the equations

$$\hat{\mathbf{x}}_b^{\mathrm{FE}} = \hat{\mathbf{x}}^{\mathrm{o}} - \frac{\dot{k}}{1+\frac{\dot{k}}{\lambda}} \mathbf{s}^{\mathrm{T}} \mathbf{C}^{-1} (\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}}) \xi \tag{29}$$

$$\hat{\beta}_b^{\mathrm{FE}} = \frac{\dot{k}}{1+\frac{\dot{k}}{\lambda}} \mathbf{s}^{\mathrm{T}} \mathbf{C}^{-1} (\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}}) \tag{30}$$

where the subscript b is used to distinguish the above solution from the LS solution $\hat{\mathbf{x}}^{\mathrm{FE}}$ and $\hat{\beta}^{\mathrm{FE}}$ that was discussed in Sect. 2.1. In general, the previous estimators are biased:

$$\mathrm{BIAS}(\hat{\mathbf{x}}_b^{\mathrm{FE}}) = E\{\hat{\mathbf{x}}_b^{\mathrm{FE}}\} - \mathbf{x} = \frac{\dot{k}}{\lambda + \dot{k}} \beta \xi \tag{31}$$

$$\mathrm{BIAS}(\hat{\beta}_b^{\mathrm{FE}}) = E\{\hat{\beta}_b^{\mathrm{FE}}\} - \beta = -\frac{\dot{k}}{\lambda + \dot{k}} \beta \,. \tag{32}$$

The estimation bias in both cases is controlled by the regularization parameter λ and it decreases as the value of λ increases. When the regularization parameter takes an infinitely large value, the estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ converge to the (unbiased) LS solution $\hat{\mathbf{x}}^{\mathrm{FE}}$ and $\hat{\beta}^{\mathrm{FE}}$.

The accuracy of the biased estimators in Eqs. (29) and (30) can be described in terms of their corresponding MSE. Specifically, the MSE matrix of $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ is given by

$$\mathrm{MSE}(\hat{\mathbf{x}}_b^{\mathrm{FE}}) = (\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A})^{-1} + \dddot{k}\, \xi \xi^{\mathrm{T}} , \tag{33}$$

whereas the MSE for $\hat{\beta}_b^{\mathrm{FE}}$ is

$$\mathrm{MSE}(\hat{\beta}_b^{\mathrm{FE}}) = \dddot{k} \,, \tag{34}$$

with the scalar factor $\dddot{k}$ given by the formula

Table 1 Comparison between various estimators for the parameter vector $\mathbf{x}$, using (i) non-biased data and (ii) uniformly biased data with different modeling schemes for the data bias effect

General model $\mathbf{y}' = \mathbf{A}\mathbf{x} + \beta\mathbf{s} + \mathbf{v}$	Model parameters $\mathbf{x}$		
	$\hat{\mathbf{x}}$	MSE ($\hat{\mathbf{x}}$)	BIAS ($\hat{\mathbf{x}}$)
Fixed-effects model (Box 1) with non-biased data (β=0) LS/BLUE solution	$\hat{\mathbf{x}}^{\mathrm{o}}[= (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}']$	$(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}$	$\mathbf{0}$
Fixed-effects model (Box 1) with biased data ($\beta \neq 0$, fixed & unknown) LS/BLUE solution	$\hat{\mathbf{x}}^{\mathrm{o}} - \dot{k}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})\xi$	$(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1} + \dot{k}\xi\xi^{\mathrm{T}}$	$\mathbf{0}$
Mixed-effects model (Box 2) with biased data (β is zero-mean RV) LS/BLUE solution	$\hat{\mathbf{x}}^{\mathrm{o}} - \frac{\dot{k}}{1+\frac{\dot{k}}{\sigma_\beta^2}}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})\xi$	$(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1} + \ddot{k}\xi\xi^{\mathrm{T}}$	$\mathbf{0}$
Fixed-effects model (Box 1) with biased data ($\beta \neq 0$, fixed & unknown) Biased solution	$\hat{\mathbf{x}}^{\mathrm{o}} - \frac{\dot{k}}{1+\frac{\dot{k}}{\lambda}}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})\xi$	$(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1} + \dddot{k}\,\xi\xi^{\mathrm{T}}$	$\frac{\dot{k}}{\dot{k}+\lambda}\beta\xi$

Table 2 Comparison between various estimators for the bias parameter β, using (1) non-biased data and (2) uniformly biased data with different modeling schemes for the data bias effect

General model $\mathbf{y}' = \mathbf{A}\mathbf{x} + \beta\mathbf{s} + \mathbf{v}$	Data bias parameter β		
	$\hat{\beta}$	MSE ($\hat{\beta}$)	BIAS ($\hat{\beta}$)
Fixed-effects model (Box 1) with non-biased data ($\beta = 0$) LS/BLUE solution	–	–	–
Fixed-effects model (Box 1) with biased data ($\beta \neq 0$, fixed & unknown) LS/BLUE solution	$\dot{k}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})$	$\dot{k}$	0
Mixed-effects model (Box 2) with biased data (β is zero-mean RV) LS/BLUE solution	$\frac{\dot{k}}{1+\frac{\dot{k}}{\sigma_\beta^2}}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})$	$\ddot{k}$	0
Fixed-effects model (Box 1) with biased data ($\beta \neq 0$, fixed & unknown) Biased solution	$\frac{\dot{k}}{1+\frac{\dot{k}}{\lambda}}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})$	$\dddot{k}$	$-\frac{\dot{k}}{\dot{k}+\lambda}\beta$

$$\dddot{k} = \frac{\dot{k}\lambda^2 + \dot{k}^2\beta^2}{(\dot{k}+\lambda)^2}\,. \tag{35}$$

The derivation of the previous MSE expressions is straightforward and it can easily be verified using standard properties from matrix calculus.

It is interesting to observe that, although $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ are numerically equal to the unbiased LS estimates $\hat{\mathbf{x}}^{\mathrm{ME}}$ and $\hat{\beta}^{\mathrm{ME}}$ obtained from the mixed-effects model *when* $\sigma_\beta^2 = \lambda$, their respective MSE measures are not necessarily equal in this case. Indeed, if we form the ratio $\ddot{k}/\dddot{k}$ under the restriction $\sigma_\beta^2 = \lambda$ and take into account Eqs. (24) and (35), we have the relationship

$$\frac{\ddot{k}}{\dddot{k}} = \frac{\dot{k} + \sigma_\beta^2}{\frac{\beta^2}{\sigma_\beta^2}\dot{k} + \sigma_\beta^2}, \tag{36}$$

which reveals that $\ddot{k} \neq \dddot{k}$ and thus $\mathrm{MSE}(\hat{\mathbf{x}}_b^{\mathrm{FE}}) \neq \mathrm{MSE}(\hat{\mathbf{x}}^{\mathrm{ME}})$, $\mathrm{MSE}(\hat{\beta}_b^{\mathrm{FE}}) \neq \mathrm{MSE}(\hat{\beta}^{\mathrm{ME}})$ when $\sigma_\beta^2 = \lambda$, unless $\beta^2 = \sigma_\beta^2$. An analogous example for the case of generalized ridge regression is discussed in Xu and Rummel (1994).

In Tables 1 and 2, a summary of the main characteristics for the three types of estimators in linear models with uniformly biased data is given. In the next section, a number of alternative regularization schemes that can be associated with the biased estimators of Eqs. (29) and (30) are presented. The problem of the optimal choice for the value of the regularization parameter λ, as well as a comparison of the MSE performance for the three types of estimators ($\hat{\mathbf{x}}^{\mathrm{FE}}$, $\hat{\beta}^{\mathrm{FE}}$), ($\hat{\mathbf{x}}^{\mathrm{ME}}$, $\hat{\beta}^{\mathrm{ME}}$) and ($\hat{\mathbf{x}}_b^{\mathrm{FE}}$, $\hat{\beta}_b^{\mathrm{FE}}$), are presented in Sect. 5.

4 Alternative views for the biased estimators in the fixed-effects linear model with uniformly biased data

4.1 Partial ridge regression

The biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ can be derived by applying a *Tikhonov-Phillips* regularization procedure (e.g., Moritz 1980; Bouman and Koop 1997) within the fixed-effects model of Box 1. In this case, the underlying "stabilizer" is applied only to the bias parameter β, without taking into account the rest of the model parameters $\mathbf{x}$. The analytical expression of the regularized inversion principle is

$$\mathbf{v}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{v} + \left(\frac{1}{\lambda}\right)\beta^2 = (\mathbf{y}' - \mathbf{A}\mathbf{x} - \beta\mathbf{s})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\mathbf{x} - \beta\mathbf{s}) + \left(\frac{1}{\lambda}\right)\beta^2 = \text{minimum}\,, \tag{37}$$

where λ corresponds to a regularization factor whose role is to provide a more stable solution by smoothing out the effect of a uniform bias β in the input data. It can easily be verified that the minimization of the above function leads to the particular biased estimators given in Eqs. (29) and (30).

In the statistical literature, such an approach is usually referred to as *partial* or *partitioned ridge regression* (Brown 1977; Farebrother 1978). It differs from the usual ridge regression method (Hoerl and Kennard 1970; Marquardt 1970; Xu and Rummel 1994) as it incorporates only a subset of the model parameters into the optimization principle.

4.2 LS inversion using a constraint for the bias size

An alternative viewpoint for the biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ is obtained by formulating the inversion of the fixed-effects model as a LS estimation problem subject to a constraint for the size of the bias parameter β. Such an approach is based on the argument that the range of possible values for the data bias is certainly bounded, since the true value of β cannot exceed some physically reasonable limits. This kind of "prior information" can be incorporated into the data analysis procedure with the help of an *inequality constraint*, which restricts the range of admissible values for β.

In our case, the estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ provide the solution to the following constrained optimization problem

$$\begin{aligned}\mathbf{v}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{v} &= (\mathbf{y}' - \mathbf{A}\mathbf{x} - \beta\mathbf{s})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\mathbf{x} - \beta\mathbf{s}) \\ &= \text{minimum} \end{aligned} \tag{38}$$

subject to

$$\beta^2 \leq R^2 , \tag{39}$$

where R^2 denotes an upper bound for the squared value of the data bias. The solution to the above problem is obtained via the method of Lagrange multipliers and it leads to the biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ given in Eqs. (29) and (30); see Björck (1996, pp. 205–206) for more mathematical details. In this particular case, it can be shown that the following condition should be satisfied by the optimal estimate of the bias parameter β

$$(\hat{\beta}_b^{\mathrm{FE}})^2 = R^2 \tag{40a}$$

or more analytically

$$\left(\frac{\dot{k}}{1+\frac{\dot{k}}{\lambda}}\right)^2 \left[\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})(\mathbf{y}' - \mathbf{A}\hat{\mathbf{x}}^{\mathrm{o}})^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s}\right] = R^2 . \tag{40b}$$

From Eq. (40b), the value of the regularization factor λ can be determined as a function of the actual data $\mathbf{y}'$ and the given bound R^2 for the bias size. Further details can be found in Draper and Smith (1998, pp. 392–394).

4.3 The "phony data" viewpoint

A third interpretation for the linear estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ is possible by adopting a pseudo-Bayesian approach for the inversion of the fixed-effects model in Box 1. According to this approach, we introduce prior information for the data bias β by adding new "data" in our analysis, in terms of the following observation equation

$$\beta^{\mathrm{obs}} = \beta + \mathrm{v}_\beta \tag{41}$$

where β^{obs} denotes an "observed" value for the bias parameter, and v_β is its associated error with the following stochastic description

$$E\{\mathrm{v}_\beta\} = 0 ; \tag{42}$$

$$E\{\mathrm{v}_\beta^2\} = \sigma_{\mathrm{v}_\beta}^2 . \tag{43}$$

It can easily be shown that the joint inversion of the fixed-effects model in Box 1 along with Eq. (41), under the LS optimal principle, leads to the estimates $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$, when

$$\sigma_{\mathrm{v}_\beta}^2 = \lambda ; \tag{44}$$

$$\beta^{\mathrm{obs}} = 0 . \tag{45}$$

The above approach, which augments the original data vector $\mathbf{y}'$ with additional "phony" (zero) observed values for some (or all) of the model parameters, provides a well-known alternative viewpoint for ridge regression problems (Draper and Smith 1998, pp. 394–395). More details for more general cases of this approach can be found in Schaffrin (1983).

4.4 Shrinking and shifting the unbiased LS solution

Further insight into the behaviour of the biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ can be obtained by examining their relationship with the corresponding LS unbiased estimates $\hat{\mathbf{x}}^{\mathrm{FE}}$ and $\hat{\beta}^{\mathrm{FE}}$.

Taking into account Eqs. (7) and (30), it is easily verified that

$$\hat{\beta}_b^{\mathrm{FE}} = \frac{\lambda}{\lambda + \dot{k}}\hat{\beta}^{\mathrm{FE}} \tag{46}$$

which shows that the biased estimate for the bias parameter β is obtained by shrinking the corresponding LS unbiased estimate. The amount of shrinkage depends on the regularization parameter λ, and it increases according to the ratio $\dot{k}/\lambda$. Note that the factor $\dot{k}$ is equal to the variance of the LS unbiased estimate $\hat{\beta}^{\mathrm{FE}}$; see Eq. (17).

Using Eqs. (6) and (29), it can also be shown that the estimator $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ satisfies the following equations:

$$\hat{\mathbf{x}}_b^{\mathrm{FE}} = \hat{\mathbf{x}}^{\mathrm{FE}} + \frac{\dot{k}}{\lambda + \dot{k}}\hat{\beta}^{\mathrm{FE}}\xi ; \tag{47}$$

$$\hat{\mathbf{x}}_b^{\mathrm{FE}} = \hat{\mathbf{x}}^{\mathrm{FE}} + \frac{\dot{k}}{\lambda}\hat{\beta}_b^{\mathrm{FE}}\xi . \tag{48}$$

According to Eqs. (47) and (48), the biased estimate $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ is a shifted version of the LS unbiased estimate $\hat{\mathbf{x}}^{\mathrm{FE}}$. The amount of the shift depends not only on the ratio $\dot{k}/\lambda$, but also on the estimated value for the bias parameter β and the auxiliary quantity ξ.

An alternative expression for $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ as a function of $\hat{\mathbf{x}}^{\mathrm{FE}}$ is

$$\hat{\mathbf{x}}_b^{\mathrm{FE}} = \frac{\lambda}{\lambda + \dot{k}} \hat{\mathbf{x}}^{\mathrm{FE}} + \frac{\dot{k}}{\lambda + \dot{k}} \hat{\mathbf{x}}^{\mathrm{o}} \tag{49}$$

which shows that the biased estimator $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ is a weighted average of (1) the unbiased LS solution $\hat{\mathbf{x}}^{\mathrm{FE}}$ and (2) the LS solution $\hat{\mathbf{x}}^{\mathrm{o}}$ that we obtain when we neglect the bias presence in the input data. Under this viewpoint, the regularization parameter λ plays the role of a relative weighting factor between the two estimators $\hat{\mathbf{x}}^{\mathrm{FE}}$ and $\hat{\mathbf{x}}^{\mathrm{o}}$.

5 Choice of the regularization parameter

The most critical aspect in every regularization technique is the choice of the regularization parameter(s) that enter into the corresponding estimation formulae. This problem has been studied extensively in the geodetic literature, mainly within the framework of the Tikhonov-Phillips regularization procedure (e.g., Schwarz 1979; Rummel et al. 1979; Bouman and Koop 1997; Cai et al. 2004). In this section, we examine briefly the problem of choosing the value of the regularization parameter λ for the particular case of the biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$.

5.1 Criterion I: MSE reduction

The inclusion of an additional unknown parameter β for the inversion of a linear model from uniformly biased data leads to estimates, for the original model parameters $\mathbf{x}$, which are less accurate (in terms of MSE) than those obtained from non-biased data. This result applies to all three types of estimators that have been studied in this paper, namely $\hat{\mathbf{x}}^{\mathrm{FE}}$, $\hat{\mathbf{x}}^{\mathrm{ME}}$ and $\hat{\mathbf{x}}_b^{\mathrm{FE}}$. The MSE matrix for each of these estimators always has larger diagonal elements than the MSE matrix of the LS estimator with non-biased data (see Table 1). The accuracy degradation in each case is solely controlled by a single positive factor ($\dot{k}$ for $\hat{\mathbf{x}}^{\mathrm{FE}}$, $\ddot{k}$ for $\hat{\mathbf{x}}^{\mathrm{ME}}$ and $\dddot{k}$ for $\hat{\mathbf{x}}_b^{\mathrm{FE}}$). In particular, the larger the values of these factors, the greater is the degradation in the estimation accuracy for the model parameters $\mathbf{x}$. Note that the scalar factors $\dot{k}$, $\ddot{k}$ and $\dddot{k}$ also correspond to the MSE for the respective estimates $\hat{\beta}^{\mathrm{FE}}$, $\hat{\beta}^{\mathrm{ME}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ of the bias parameter β (see Table 2). As a result, a reasonable criterion for choosing an optimal value for the regularization parameter λ, in the case of the biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$, is the minimization of the factor $\dddot{k}$.

The analytical expression of $\dddot{k}$ as a function of λ has already been given in Eq. (35). Its minimization ensures that (i) the estimation accuracy for the model parameters $\mathbf{x}$ has the smallest possible degradation due to the use of uniformly biased data, and (ii) the MSE of the bias estimate $\hat{\beta}_b^{\mathrm{FE}}$ is minimum. Using Eq. (35), it is easily established that the optimal value of λ which minimizes $\dddot{k}$ is

$$\lambda^{\mathrm{opt}} = \beta^2 , \tag{50}$$

where β is the true value of the data bias. In this case, the MSE performance of the biased solution ($\hat{\mathbf{x}}_b^{\mathrm{FE}}$, $\hat{\beta}_b^{\mathrm{FE}}$) is always better than the MSE performance of the LS unbiased solution ($\hat{\mathbf{x}}^{\mathrm{FE}}$, $\hat{\beta}^{\mathrm{FE}}$) since we have

$$\frac{\dddot{k}}{\dot{k}} = \frac{\beta^2}{\beta^2 + \dot{k}} < 1 . \tag{51}$$

The optimal choice for λ according to Eq. (50) is of theoretical value only, since the true bias β is obviously not known in practice. A more practical approach can be followed by first requiring that the factor $\dddot{k}$ is smaller than $\dot{k}$, and then finding the values of λ that satisfy this condition. In this way, we can ensure a priori that the biased solution ($\hat{\mathbf{x}}_b^{\mathrm{FE}}$, $\hat{\beta}_b^{\mathrm{FE}}$) offers better MSE performance than the LS unbiased solution ($\hat{\mathbf{x}}^{\mathrm{FE}}$, $\hat{\beta}^{\mathrm{FE}}$).

Again using Eq. (35), it is easily verified that

$$\frac{\dddot{k}}{\dot{k}} < 1 \Leftrightarrow \lambda > \frac{\beta^2 - \dot{k}}{2} . \tag{52}$$

The above inequality offers a range of values within which we should select the regularization parameter in order to guarantee better MSE performance than the LS unbiased solution. Based on Eq. (52), a reasonable choice for λ is

$$\lambda = \frac{(\beta^2)_{\max} - \dot{k}}{2} \tag{53}$$

where $(\beta^2)_{\max}$ denotes a maximum bound that we are willing to accept for the squared value of the data bias.

5.2 Criterion II: bounded bias

An alternative methodology for the choice of the regularization parameter λ is obtained by following the viewpoint of Sect. 4.2. According to this approach, the biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ provide the solution to a constrained LS adjustment problem which is formulated in terms of Eqs. (38) and (39). The regularization parameter can be determined in this case by solving Eq. (40b). This yields a second-order polynomial in λ, which has the analytic form

$$[R^2 - (\hat{\beta}^{\mathrm{FE}})^2]\lambda^2 + 2\dot{k}R^2\lambda + (\dot{k})^2 R^2 = 0 , \tag{54}$$

where $\hat{\beta}^{\mathrm{FE}}$ is the unbiased LS estimate for the data bias parameter and R^2 is the imposed upper bound for the squared true value of β.

Equation (54) has only real roots since its discriminant Δ is always positive, i.e.,

$$\Delta = 4R^2(\dot{k}\hat{\beta}^{\mathrm{FE}})^2 . \tag{55}$$

Assuming that $\dot{k}$ and R^2 are both non-zero, there are three different cases that can be distinguished for the possible solutions of Eq. (54), depending on the relationship between R^2 and $(\hat{\beta}^{\mathrm{FE}})^2$. In particular, the following alternative scenarios can be identified:

Case 1: if $R^2 > (\hat{\beta}^{\mathrm{FE}})^2$, then Eq. (54) has two negative real roots;

Case 2: if $R^2 < (\hat{\beta}^{\mathrm{FE}})^2$, then Eq. (54) has one positive and one negative real root;

Case 3: if $R^2 = (\hat{\beta}^{\mathrm{FE}})^2$, then Eq. (54) has only one negative real root, $\lambda = -\frac{\hat{k}}{2}$.

From the above cases, only the second leads to an acceptable (positive) value for the regularization parameter λ. In fact, the condition in *Case 2* provides a justification for the practical use of the biased estimators $\hat{\mathbf{x}}_b^{\mathrm{FE}}$ and $\hat{\beta}_b^{\mathrm{FE}}$ in cases where a realistic bound for the squared data bias is *smaller* than the squared value of the unbiased LS estimate of β. Such a result is compatible with the well-known behaviour of ordinary (unbiased) LS estimators, which tend to over-estimate the size of the unknown parameters in a linear model.

Note that, in contrast to the methodology presented in Sect. 5.1, the value of the regularization parameter λ that is obtained from the solution of Eq. (54) depends directly on the available data, since the term $\hat{\beta}^{\mathrm{FE}}$ is a linear function of the measurements $\mathbf{y}'$.

6 Summary and conclusions

The aim of this paper was to study and compare various estimation schemes that can be used in linear models with uniformly biased data. A type of biased estimators for such models has been introduced, which has the potential to provide better MSE performance than the ordinary LS (unbiased) solution. Different regularization viewpoints that can be associated with these biased estimators were also discussed, along with a number of selection strategies that can be followed for the choice of the regularization parameter that enters into the biased estimation algorithm. Although it has not been treated explicitly in this paper, we should note that results similar to the ones presented here are also obtained in the case where only a part of the input data is affected by a constant unknown bias.

Acknowledgements The author would like to thank Drs. A. Dermanis, J. Kusche, K-R Koch and B. Schaffrin for the constructive criticism and the useful comments that they provided during the review of this paper.

References

Aduol FWO (1987) Detection of outliers in geodetic networks using principal component analysis and bias parameter estimation. Tech Rep 2, Institute of Geodesy, Univ of Stuttgart, Stuttgart

Björck A (1996) Numerical methods for least squares problems. Society for Industrial and Applied Mathematics (SIAM), Philadelphia

Bouman J, Koop R (1997) Quality differences between Tikhonov regularization and generalized biased estimation in gradiometric analysis. DEOS Progr Lett 97(1):42–48

Brown PJ (1977) Centering and scaling in ridge regression. Technometrics 19(1):35–36

Cai J, Grafarend EW, Schaffrin B (2004) The A-optimal regularization parameter in uniform Tykhonov-Phillips regularization – α-weighted BLE. IAG Symposia, vol 127. Springer, Berlin Heidelberg New York, pp 309–324

Dermanis A (1976) Probabilistic and deterministic aspects of linear estimation in geodesy. Rep 244, Dept of Geodetic Sci, Ohio State Univ, Columbus

Dermanis A (1979) Adjustment of geodetic networks in the presence of signals. ZfV 104(11):505–518

Dermanis A (1991) A unified approach to linear estimation and prediction. Unpublished report (http://der.topo.auth.gr/). Presented at the 20th general assembly of the international union of geodesy and geophysics, Vienna

Dermanis A, Sansò F (1993) A dialogue on geodetic theory today. IAG IV Bull 1:6–9

Dermanis A, Rummel R (2000) Data analysis methods in geodesy. In: Dermanis A, Grun A, Sansò F (eds) Geomatic methods for the analysis of data in earth sciences. Lecture notes in earth sciences series, vol 95. Springer, Berlin Heidelberg New York, pp 17–92

Draper NR, Smith H (1998) Applied regression analysis, 3rd edn. Wiley, New York

Farebrother RW (1978) Partitioned ridge regression. Technometrics 20(2):121–122

Gaspar P, Ogor F, Le Traon P-Y, Zanife O-Z (1994) Estimating the sea state bias of the TOPEX and POSEIDON altimeters from crossover differences. J Geophys Res 99(C12):24981–24994

Grafarend EW, Schaffrin B (1993) Adjustment computations in linear models (in German). Bibliograph Inst, Manheim, Germany

Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, Berlin Heidelberg New York

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12(1):55–67

Jia M, Tsakiri M, Stewart M (2000) Mitigating multipath errors using semi-parametric models for high precision static positioning. IAG Symposia, vol 121. Springer, Berlin Heidelberg New York, pp 393–398

Koch K-R (1999) Parameter estimation and hypothesis testing in linear models, 2nd edn. Springer, Berlin Heidelberg New York

Kubáčková L, Kubáček L (1993) A group of gravimeters, stochastical problems and their solution. IAG Symposia, vol 112. Springer, Berlin Heidelberg New York, pp 275–278

Kukuča J (1987) Systematic influences and their elimination. In: Kubáčková L, Kubáček L, Kukuča J (eds) Probability and statistics in geodesy and geophysics. Elsevier, Amsterdam, pp 257–289

Kusche J (2003) A Monte-Carlo technique for weight estimation in satellite geodesy. J Geod 76:641–652

Kusche J (2004) Personal communication

Leibelt PR (1967) An introduction to optimal estimation. Addison-Wesley, Reading

Lerch FJ (1991) Optimum data weighting and error calibration for estimation of gravitational parameters. Bull Géod 65:44–52

Marquardt DW (1970) Generalized inverses, ridge regression, biased linear estimation and non-linear estimation. Technometrics 12(3):591–612

Moritz H (1980) Advanced physical geodesy. Herbert Wichmann Verlag, Karlsruhe

Moritz H (1997) Gauss, Gödel, Heisenberg. In: Natke HG, Ben-Haim Y (eds) Uncertainty: models and measures. Akademie Verlag, Berlin, pp 101–108

Moritz H, Sansò F (1980) A dialogue on collocation. Boll Geod Sci Affn 39(1):49–51

Rummel R, Schwarz K-P, Gerstl M (1979) Least squares collocation and regularization. Bull Géod 53:343–361

Satirapod C, Wang J, Rizos C (2003) Comparing different global positioning system data processing techniques for modeling residual systematic errors. J Surv Eng 129(4):129–135

Scales JA, Snieder R (1998) What is noise? Geophysics 63(4): 1122–1124

Schaffrin B (1983) Model choice and adjustment techniques in the presence of prior information. Rep 351, Department of Geodetic Science, Ohio State University, Columbus

Schaffrin B, Grafarend EW (1986) Generating classes of equivalent linear models by nuisance parameter elimination. Manuscr Geod 11:262–271

Schaffrin B, Baki-Iz H (2001) Integrating heterogeneous data sets with partial inconsistencies. IAG Symposia, vol 123, Springer, Berlin Heidelberg New York, pp 49–54

Schwarz K-P (1979) Geodetic improperly posed problems and their regularization. Boll Geod Sci Affn 38(3):389–416

Teunissen PJG (2000) Adjustment theory: an introduction. Series on mathematical geodesy and positioning. Delft University Press, Delft

Tscherning CC, Knudsen P (1986) Determination of bias parameters for satellite altimeter by least-squares collocation. In: Proceedings of the 1st Hotine-Marussi symposium on mathematical geodesy, Rome, June 3–6, 1985, pp 833–851

Xu P, Rummel R (1994) Generalized ridge regression with applications in determination of potential fields. Manuscr Geod 20:8–20