Iterative generalized cross-validation for fusing heteroscedastic data of inverse ill-posed problems

Peiliang Xu

Disaster Prevention Research Institute, Kyoto University, Uji, Kyoto 611-0011, Japan. E-mail: pxu@rcep.dpri.kyoto-u.ac.jp

SUMMARY
The method of generalized cross-validation (GCV) has been widely used to determine the regularization parameter, because the criterion minimizes the average predicted residuals of measured data and depends solely on data. The data-driven advantage is valid only if the variance–covariance matrix of the data can be represented as the product of a given positive definite matrix and a scalar unknown noise variance. In practice, important geophysical inverse ill-posed problems have often been solved by combining different types of data. The stochastic model of measurements in this case contains a number of different unknown variance components. Although the weighting factors, or equivalently the variance components, have been shown to significantly affect joint inversion results of geophysical ill-posed problems, they have been either assumed to be known or empirically chosen. No solid statistical foundation is available yet to correctly determine the weighting factors of different types of data in joint geophysical inversion. We extend the GCV method to accommodate both the regularization parameter and the variance components. The extended version of GCV essentially consists of two steps, one to estimate the variance components by fixing the regularization parameter and the other to determine the regularization parameter by using the GCV method and by fixing the variance components. We simulate two examples: a purely mathematical integral equation of the first kind modified from the first example of Phillips (1962) and a typical geophysical example of downward continuation to recover the gravity anomalies on the surface of the Earth from satellite measurements. Based on the two simulated examples, we extensively compare the iterative GCV method with existing methods, which have shown that the method works well to correctly recover the unknown variance components and determine the regularization parameter. In other words, our method lets data speak for themselves, decide the correct weighting factors of different types of geophysical data, and determine the regularization parameter. In addition, we derive an unbiased estimator of the noise variance by correcting the biases of the regularized residuals. A simplified formula to save the time of computation is also given. The two new estimators of the noise variance are compared with six existing methods through numerical simulations. The simulation results have shown that the two new estimators perform as well as Wahba’s estimator for highly ill-posed problems and outperform any existing methods for moderately ill-posed problems.

Key words: Inverse theory; Tomography; Satellite geodesy.

1 INTRODUCTION
An inverse problem is to reconstruct a continuous function (or an unknown vector) from indirect measurements. A linearized inverse problem can almost always be written as follows:

\[ y = A\beta + \epsilon, \]

where \( y \) is a vector of noisy measurements, \( A \) is the design matrix, \( \beta \) is a vector of unknown parameters to be estimated and \( \epsilon \) is the error vector of measurements \( y \). In general, \( \epsilon \) is assumed to be normally distributed, or more precisely, \( \epsilon \sim N(0, W^{-1}\sigma^2) \), where \( W \) is a given positive definite matrix and \( \sigma^2 \) is an unknown positive scalar. Very often, \( W \) is an identity matrix, as can be seen, in particular, in statistical literature. \( \sigma^2 \) is also called the noise variance and/or the variance of unit weight. If the linear model is well posed or well conditioned, the problem (1) can be readily solved by using the least-squares (LS) method. However, as an important class of models, when (1) is referred to as an inverse ill-posed problem, the LS solution has been known to be too noisy and physically meaningless, since the signal of interest can be completely inundated into the noise of the solution. As an ill-posed or ill-conditioned problem, \( A \) is almost always assumed, explicitly or implicitly, to be
near but not exact multicollinearity. We will follow this assumption in the remainder of this paper. Applications of (1) as an ill-posed problem can now be found in almost every field of science and engineering, for example, in mathematical and statistical science (see, e.g. Tikhonov & Arsenin 1977; Vinod & Ullah 1981), in astronomy and physical science (see, e.g. Schaffrin et al. 1977; Tarantola 1987; Xu & Rummel 1994; Kotsakis 2007; Schaffrin 2008), and in image reconstruction and computerized tomography (see, e.g. Deans 1983; Natterer 1986), just to mention a few such examples only.

Regularization is critical to obtain a meaningful solution to the linear inverse ill-posed problem (1). The regularized solution to (1) can be written as follows:

$$\hat{\beta}_r = \left( A^T W A + \kappa R \right)^{-1} A^T W y,$$

(2)

where $\kappa$ is positive and has been called as the regularization and/or ridge parameter and $R$ is a given positive (semi-)definite matrix. The quality of solution (2) essentially depends on the choice of $\kappa$, which may be motivated and selected by emphasizing a different attribute of $\hat{\beta}_r$ (see, e.g. Phillips 1962; Hoerl & Kennard 1970; Tikhonov & Arsenin 1977; Golub et al. 1979; Vinod & Ullah 1981; Morozov 1984; Xu & Rummel 1994; Xu 1998). If multiple regularization parameters are employed, the solution (2) becomes

$$\hat{\beta}_r = \left( A^T W A + \sum_{i=1}^{q} \kappa_i R_i \right)^{-1} A^T W y$$

(3)

(see, e.g. Barry 1986; Gu & Wahba 1991; Xu & Rummel 1992, 1994; Xu et al. 2006a), where $\kappa_i$ are non-negative regularization parameters and $R_i$ are given positive (semi-)definite matrices. In a recent simulation study of determining the Earth’s gravity field from satellite data measurements, Xu et al. (2006a) showed that the multiple parameter regularization with statistically empirical $R_i$ only marginally improved the solution (2) in terms of mean squared error (MSE). Instead of using multiple regularization parameters as represented by (3), Kusche & Klees (2002) introduced a new unknown parameter for $R$ in (2) and simultaneously estimated it, together with the regularization parameter $\kappa$.

Generalized cross-validation (GCV) has been widely used to determine the regularization parameter $\kappa$. Cross-validation was originally developed for model selection (see, e.g. Allen 1974; Stone 1974). The basic idea is to build a model on part of data and use the remaining data for confirmation or validation of the model. Since splitting the data into two subsets will result in a huge number of combinations (even for a moderate number of data), common practice is to use all but one of the data for building a particular model and then use the only left-out data for validating the model. The model that minimizes the predicted errors over all the data is selected as the best model (Allen 1974; Stone 1974). Instead of directly minimizing the predicted errors over the data, Golub et al. (1979) proposed minimizing the average predicted residuals. By further imposing the invariance of rotation and by assuming $W = I$, they developed the well-known GCV method for choosing the regularization parameter. The corresponding GCV objective function is given as follows:

$$V(\kappa) = \frac{(1/n) y^T (I - H(\kappa))^2 y}{(1/n) tr[(I - H(\kappa))^{-1}]}$$

(4a)

where $n$ is the number of measurements $y$, $tr(\cdot)$ stands for the trace of a square matrix, and

$$H(\kappa) = A (A^T A + \kappa R)^{-1} A^T$$

(4b)

which has been better known as the hat matrix in the case of $\kappa = 0$. The most appealing advantages of the GCV method are twofold: (i) that (4a) is independent of any unknown model parameters $\beta$ and the noise variance $\sigma^2$ and (ii) that one can find the optimal $\kappa$ by minimizing (4a) as soon as the data are available.

The GCV method was then extended by Silverman (1985) to account for a given diagonal weight matrix and by Diggle & Hutchinson (1989) to account for a full weight matrix $W$. Actually, Diggle & Hutchinson (1989) went further to assume that the weight matrix $W$ contains one unknown correlation coefficient $\rho$ and then used the penalized maximum log-likelihood to estimate $\rho$. By assuming that the weight matrix $W$ can be represented by a small number of unknown parameters $\tau$, Wang (1998) proposed using the GCV method and generalized maximum likelihood (GML) to simultaneously estimate $\kappa$ and $\tau$.

In this paper, we assume that the data of inverse ill-posed problems are heteroscedastic and the variance–covariance matrix of $y$ or $\epsilon$ can be represented by a number of unknown variance components $\Sigma_y = \sum_{i=1}^{m} U_i \sigma_i^2$,

(5)

where $U_i$ are given positive (semi-)definite matrices, and $\sigma_i^2 (i = 1, 2, \ldots, m)$ are the unknown variance components to be estimated. The stochastic model (5), together with the functional model (1), represents a standard linear ill-posed problem for all joint geophysical inversion (see, e.g. Vozoff & Jupp 1975; Delouis et al. 2002; Feigl et al. 2002; Kozlovskaya et al. 2007; Amoruso et al. 2008; Maceira & Ammon 2009). If $\Sigma_y = \text{diag}(W \sigma_i^2)$, then (5) simply means that different types of data are combined to jointly solve a geophysical ill-posed problem. In case that $\Sigma_y$ is not block-diagonal, spatial and/or temporal correlation of data of the same type can also be readily taken into account. For example, Delouis et al. (2002) combined InSAR, GPS, teleseismic and strong-motion data to jointly invert for the spatial and temporal distribution of slip. They concluded that the excellent fitting of one type of data did not indicate that the resolved solution was good. Feigl et al. (2002) used GPS, ERS-1, RADARSAT and SPOT data to jointly estimate the slip distribution for the Izmit earthquake of 1999 August 17. Their results showed that different inversions of seismic moment with different types of data could be different by 40 per cent. The importance of correlation among the same type of data in geophysical inversion was demonstrated by Yagi & Fukahata (2008). Although the weighting factors, or equivalently, the variance components in the stochastic model (5), have been shown to significantly affect results of geophysical inversion (see, e.g. Delouis et al. 2002; Feigl et al. 2002), they have been either assumed known or empirically chosen in joint geophysical inversion (see, e.g. Vozoff & Jupp 1975; Delouis et al. 2002; Feigl et al. 2002; Gallardo & Mejia 2007; Khan et al. 2007; Konca et al. 2007; Kozlovskaya et al. 2007; Amoruso et al. 2008; Maceira & Ammon 2009).

To the best knowledge of this author, no solid statistical foundation is documented in geophysical literature to let data decide the correct weighting factors of different types of data in joint geophysical inversion of ill-posed problems.

Schwintzer (1990) and Koch & Kusche (2002) treated the regularization parameter $\kappa$ as if it were the ratio of two variance components and used variance component estimation techniques to simultaneously estimate $\sigma_i^2$ and $\kappa$. Recently, Xu et al. (2006b) proposed an iterative MSE-based method to simultaneously estimate $\sigma_i^2$ and $\kappa$, together with the model parameters $\beta$. The estimates of $\sigma_i^2$ were proved to be unbiased up to the second order approximation.
Motivated by the success of Xu et al. (2006b), we will extend the GCV method to include the noise model (5) in this paper. Diggle & Hutchinson (1989) and Wang (1998) proposed a few methods for simultaneously estimating $\kappa$ and $\tau$. In order to compare these methods with the iterative GCV method, we will extend these methods such that they can be used to simultaneously estimate the variance components $\sigma_i^2$ and the regularization parameter $\kappa$ in Section 2. As a special case, we will also summarize the results of estimating the noise variance in this section. We will then develop the iterative GCV method in Section 3. Experiments and simulation results will be reported in Section 4.

2 EXTENDING EXISTING METHODS FOR SIMULTANEOUSLY ESTIMATING THE REGULARIZATION PARAMETER AND VARIANCE COMPONENTS

2.1 Extending existing methods

By assuming that the weight matrix is unknown up to a few unknown parameters, namely, $\Sigma_l = \sigma_l^2 W_l^{-1}(x)$, Diggle & Hutchinson (1989) and Wang (1998) proposed GCV, generalized maximum likelihood (GML), and penalized maximum likelihood (PML) to simultaneously estimate $\kappa$ and $\tau$. Although the original GCV criterion has been appealing for containing the only unknown parameter of regularization, these methods should essentially imply their usefulness in estimating unknown parameters of the stochastic model of the measurements $y$. Thus we will straightforwardly extend these three methods in order to simultaneously estimate the regularization parameter and variance components.

For the assumed stochastic model (5) of the measurements $y$, the corresponding GCV criterion (4a) can be rewritten as follows:

$$V(\kappa, \sigma) = (1/n)r^T(\kappa, \sigma) \Sigma^{-1} r(\kappa, \sigma) / \left[(1/n)r(I - H(\kappa, \sigma))\right]^T,$$  \hspace{1cm} (6a)

where

$$H(\kappa, \sigma) = A(A^T \Sigma^{-1} A + \kappa R)^{-1} A^T \Sigma^{-1},$$

$$r(\kappa, \sigma) = (I - H(\kappa, \sigma))y,$$

$$\sigma = (\sigma_1^2, \sigma_2^2, ..., \sigma_m^2)^T.$$  \hspace{1cm} (6b)

At first glance, the weighted GCV objective function (6a) contains all the $m$ unknown parameters of variance components and one regularization parameter $\kappa$. As a result, one might immediately proceed and use optimization algorithms to solve for all the $(m + 1)$ unknown parameters. After carefully examining (6b), we see that the total number of independent variables in (6b) is equal to $m$ but not $(m + 1)$. In fact, (5) can be rewritten as follows:

$$\Sigma_l = \sigma_l^2 \left( \sum_{i=1}^n U_i \sigma_i^2 / \sigma_i^2 + \sum_{i=l+1}^n U_i \sigma_i^2 / \sigma_i^2 \right).$$  \hspace{1cm} (7)

By inserting (7) into (6b), we can immediately see that $H(\kappa, \sigma)$ of (6b) is essentially the matrix function of $(m - 1)$ independent variables $\sigma_i^2 / \sigma_i^2$ ($i \neq l$) plus the new variable $\kappa \sigma_l^2$ only. Since $r(\kappa, \sigma)$ is connected to the unknown parameters through $H(\kappa, \sigma)$ of (6b), we finally prove that (6a) contains only $m$ meaningful, independent variables.

Without loss of generality, we fix $l$ to 1, collect all the independent variables $\sigma_i^2 / \sigma_1^2$ ($i \geq 2$) into an unknown vector $\tau$, and still denote $\kappa \sigma_1^2$ by $\kappa$. In fact, the elements of $\tau$ uniquely determine the weighting factors of different types of data in joint geophysical inversion, though they are unknown here and to be computed from the collected data. We use $U(\tau)$ to denote the whole matrix on the right-hand side of (7), namely,

$$U(\tau) = U_1 + \sum_{i=2}^n U_i \tau_{i-1}. \hspace{1cm} (8)$$

In order to derive the GCV function with the full matrix of $U(\tau)$, we can first transform the correlated measurements $y$ into independent measurements and then insert all the related quantities and the transformation into (4a). Thus the final weighted GCV for simultaneously estimating the regularization parameter and variance components becomes

$$V(\kappa, \tau) = \frac{(1/n)r^T(\kappa, \tau) U^{-1}(\tau) r(\kappa, \tau)}{[(1/n)r(I - H(\kappa, \tau))]^T},$$  \hspace{1cm} (9a)

where

$$H(\kappa, \tau) = A(A^T U^{-1}(\tau) A + \kappa R)^{-1} A^T U^{-1}(\tau),$$  \hspace{1cm} (9b)

$$r(\kappa, \tau) = (I - H(\kappa, \tau))y.$$  \hspace{1cm} (9c)

After the minimization problem (9a) is solved, we still need the estimate of $\sigma_1^2$ in (7) in order to obtain the final results of the estimated variance components, which will be discussed in the next subsection.

Before extending the GML method for simultaneously estimating the regularization parameter and variance components, we note that the original GML method is derived on the assumption of $W = I$. The GML method was first proposed for choosing the regularization parameter $\kappa$ (Wahba 1985). Given the normal density function of the data $y$ and by treating the regularizer $R_\kappa$ as it were the normal prior weight matrix of $B$ with prior zero mean, one can then derive the marginal distribution of $y$ given $\sigma^2$ and $\kappa$. Further solving the maximum log-likelihood of the marginal distribution for $\sigma^2$, one then obtains the reduced log-likelihood which contains the regularization parameter $\kappa$ only. For more details of the derivation, the reader is referred to Galatsanos & Katsaggelos (1992) or Wahba (1985). Actually, this reduced log-likelihood is nothing but the objective function of the GML method, which is given as follows:

$$M(\kappa) = \frac{y^T(I - H(\kappa))y}{[det(I - H(\kappa))]^{(n-t)}}.$$  \hspace{1cm} (10)

(see, e.g. Wahba 1985; Gu & Wahba 1991), where $det[I - H(\kappa)]$ stands for the product of all the $n_1$ non-zero eigenvalues of $[I - H(\kappa)]$, $H(\kappa)$ has been defined by (4b). If $A$ is of full rank of columns and $R$ is only positive semi-definite, then $n_1$ is equal to $(n - t + r_k)$, where $r_k$ stands for the rank of $R$. In particular, if $r_k = t$, the power in the denominator of (10) becomes $1/n$, which can also be found in Galatsanos & Katsaggelos (1992). In this special case, it is rather trivial to prove algebraically that all the $n$ eigenvalues of $[I - H(\kappa)]$ are positive, with the $(n-t)$ eigenvalues being exactly equal to unity and the other $t$ eigenvalues between zero and one. In the case of $\kappa = 0$, these latter $t$ eigenvalues become zero exactly.

If the measurements are of the stochastic model (5), Wang (1998) extended the Wahba’s GML criterion (10) to the case of correlated noise, which will be called the weighted GML method in this paper and is listed, in our notation, in the following.

$$M(\kappa, \tau) = \frac{y^T U^{-1}(\tau) (I - H(\kappa, \tau)) y}{[det[U^{-1}(\tau) (I - H(\kappa, \tau))]]^{(n-t)}},$$  \hspace{1cm} (11)

where $det(\cdot)$ follows the definition by Wahba (1985). As explained in the above, for any positive $\kappa$, the total number of nonzero eigenvalues of the matrix $(I - H(\kappa, \tau))$ is equal to $(n - t + r_k)$. Thus it

© 2009 The Author, GJI, 179, 182–200
Journal compilation © 2009 RAS
would be wise to use all the non-zero eigenvalues in (11); otherwise, one would have to choose \((n - t)\) values from the \((n - t + r_\kappa)\) non-zero eigenvalues.

Now we will derive a new weighted GML criterion. It is well known that any correlated noise can be whitened by diagonalizing the correlation information/matrix. As a result, we can readily obtain the corresponding weighted GML objective function through transformation. In other words, we can apply any decomposition techniques to \(\Sigma_y\) of (5) in order to whiten the correlated measurements \(y\), then insert the transformation into the original GML method (10), and finally derive the weighted GML criterion. In the case of \(U(\tau)\) of (8), we can easily write, with almost no effort of derivation, a new weighted GML objective function as follows:

\[
M(\kappa, \tau, \sigma_\kappa^2) = \frac{y^T U^{-1}(\tau) (I - H(\kappa, \tau)) y}{\sigma_\kappa^2 [\det(U^{-1}(\tau) - H(\kappa, \tau))]^{1/n_1}},
\]

(12)

where \(H(\kappa, \tau)\) and \(n_1\) have been defined before. It is obvious from (12) that the appearance of \(\sigma_\kappa^2\) in \(M(\kappa, \tau, \sigma_\kappa^2)\) does not make sense at all in terms of optimization. In other words, we should only limit ourselves to the estimation of \(\kappa\) and \(\tau\) only, and independently estimate \(\sigma_\kappa^2\), as is exactly the case of the weighted GCV in the above. The final weighted GML criterion is given by

\[
M(\kappa, \tau) = \frac{y^T U^{-1}(\tau) (I - H(\kappa, \tau)) y}{[\det(U^{-1}(\tau) - H(\kappa, \tau))]^{1/n_1}},
\]

(13)

Except for the difference in whether all or parts of the nonzero eigenvalues are chosen, (13) and (11) are also different in that \(U^{-1}(\tau)\) does not appear in the denominator of (13). Given the complete knowledge on the correlation information, (13) and (11) are essentially equivalent in determining the regularization parameter \(\kappa\), because the correlation matrix is independent of \(\kappa\). However, as far as \(\kappa\) and \(\tau\) are concerned, (13) and (11) are essentially different, because \(U^{-1}(\tau)\) imposes extra constraint on the unknown stochastic parameters \(\tau\) in the case of (11).

Finally, we will extend the third (and last) class of methods, namely, PML, for simultaneously estimating the regularization parameter and the variance components. As might be obvious by its name, this class of methods is to minimize the log-likelihood or the sum of weighted residuals of \(y\), with an extra term of penalty to favour a particular model among those that are of about the same performance in the log-likelihood. More specifically, the PML method proposed by Diggle & Hutchinson (1989) consists of three terms of penalty. The first term is to penalize the residuals, the second term to penalize the variance–covariance matrix of measurements and the last term to account for the hat or influence matrix \(H(\kappa, \tau)\). Three different combinations of penalizing these terms were reported in Diggle & Hutchinson (1989). Since the first PML is actually equivalent to the weighted GCV, and since the second PML method proposed by Diggle & Hutchinson (1989) was shown to have the worst performance by themselves, we will limit ourselves here to their third PML criterion, which is to minimize

\[
L(\kappa, \tau) = n \ln[\tau^T (\kappa, \tau) U^{-1}(\tau) \tau(\kappa, \tau)] + \ln |U(\tau)| + \ln n \cdot \tau r[H(\kappa, \tau)].
\]

(14)

The function (14) will be called the weighted PML in this paper (see also Wang 1998). In fact, the first two terms on the right-hand side of (14) correspond exactly to the log-likelihood of \(y\) in the case of \(\kappa = 0\), and the last term of penalty tends to choose the model that would produce the maximum variance improvement of the adjusted/corrected measurements against the original data \(y\).

### 2.2 Estimation of the noise variance

Obviously, an appealing advantage of the above weighted GCV, GML and PML methods is that they all require no information on the noise variance and can be used to simultaneously determine the regularization parameter \(\kappa\) and the relative variance factors (and/or correlation coefficients) \(\tau\). In other words, these methods do not directly estimate the noise variance.

Estimating the noise variance is an indispensable component of solving an inverse ill-posed geophysical problem. The reasons are twofold: (i) evaluating a regularized solution statistically cannot be complete without the noise variance and (ii) a good estimate of the noise variance reveals the actual accuracy of the measurements. In the case of the variance component model (7), we need an estimate of the noise variance \(\sigma_\kappa^2\) in order to transform \(\tau\) into the variance component estimates.

If the model (1) is well posed, it is well known that the optimal estimate of the noise variance can be readily derived from the residuals and its degrees of freedom (see, e.g. Searle 1971). This latter number is actually equal to the trace of \([I - A(A^T \tau A)^{-1}A^T \tau W]\) and often also called the number of redundant measurements. In the case of inverse ill-posed models, the first estimate of the noise variance from the regularized residuals was proposed by Wahba (1983). By treating the trace of \([I - H(\kappa)]\) as if it were the degrees of freedom, Wahba (1983) proposed to estimate \(\sigma_\kappa^2\) by

\[
\sigma_{\hat{\kappa}}^2 = \frac{\|I - H(\kappa)\|^2}{tr[I - H(\kappa)]},
\]

(15)

where \(\kappa\) is chosen by GCV. Accordingly, \(tr[I - H(\kappa)]\) was called the equivalent degrees of freedom by Wahba (1983) (see also Hall & Titterington 1987; Galatsanos & Katsaggelos 1992). From the point of view of the average squared errors, Nychka (1990) modified the estimator (15) by applying a scaling factor slightly larger than unity to the matrix \(H(\kappa)\) in the denominator of (15). Nychka’s estimator of the noise variance then becomes

\[
\sigma_{\hat{\kappa}}^2 = \frac{\|I - H(\kappa)\|^2}{tr[I - \delta H(\kappa)]},
\]

(16)

where \(\delta = 2 - 1/k\), and

\[
K = \frac{2h}{2h - 1 + 4h},
\]

with \(h\) being equal to two in the case of minimum energy spline smoothing.

Based on the same error structure of \(I \sigma^2\) as in Wahba (1983), Buckley et al. (1988) investigated a class of noise variance estimators of the following type:

\[
\hat{\sigma}_w^2 = \frac{y^T D y}{tr[D]},
\]

where \(D\) is any symmetric non-negative–definite matrix. When the above formula is applied to the regularized residuals \([I - H(\kappa)]y\), they immediately obtained the estimator of the noise variance

\[
\hat{\sigma}_w^2 = \frac{\|I - H(\kappa)\|^2}{tr[I - H(\kappa)])^2]},
\]

(17)

An alternative estimator of \(\sigma^2\) was also proposed in the literature (see, e.g. Carter & Eagleson 1992) and is simply given by

\[
\hat{\sigma}_d^2 = \frac{y^T(I - H(\kappa))y}{tr[I - H(\kappa)]}.
\]

(18)

Carter & Eagleson (1992) made a comparison of the three estimators, namely, \(\hat{\sigma}_{\hat{\kappa}}^2\), \(\hat{\sigma}_w^2\) and \(\hat{\sigma}_d^2\). They mentioned that \(\hat{\sigma}_{\hat{\kappa}}^2\) had a similar bias and variance property to \(\hat{\sigma}_{\hat{\kappa}}^2\), and conducted small numerical...
simulations with a few tens of measurements to compare $\hat{\sigma}_n^2$ and $\hat{\sigma}_k^2$, with the recommendation that $\hat{\sigma}_k^2$ should be used to estimate the noise variance $\sigma^2$.

Motivated by the cross-validation function and, again, based on the same error structure of $10^\sigma$ as in Wahba (1983), Galatsanos & Katsaggelos (1992) proposed an alternative estimator for the noise variance, which is given by

$$\hat{\sigma}_n^2 = \frac{\|H^{1/2}(\kappa)[I - H(\kappa)]y\|^2}{tr[H(\kappa)[I - H(\kappa)]]},$$

(19)

where $\kappa$ is chosen by GCV. Although an ML estimator of $\sigma^2$ was also presented by Galatsanos & Katsaggelos (1992), it was shown to have the worst performance in comparison with (15) and (19). Thus this ML estimator of the noise variance is omitted here.

The estimator (15) was then extended by Silverman (1985) to the case where the measurements were assumed to be independent but of different accuracy, namely, $\sigma_i^2 = \sigma^2/w_i$ with $w_i$ being given and all being positive. A further extension can readily be made by replacing the diagonal weight matrix with the full weight matrix $W$, which is simply given by

$$\hat{\sigma}_n^2 = \frac{r^T(k, W)W r(k, W)}{tr[I - H(k, W)]},$$

(20)

where

$$r(k, W) = [I - H(k, W)]y,$$

$$H(k, W) = A(A^TWA + \kappa I)^{-1}A^T W.$$

If we only have the stochastic model (5), then (15) (or (20)), (16), (17), (18) and (19) can also be rewritten, respectively, as follows:

$$\hat{\sigma}_n^2 = \frac{r^T(k, \tau)U^{-1}(\tau)r(k, \tau)}{tr[I - H(k, \tau)]},$$

(21)

$$\hat{\sigma}_n^2 = \frac{r^T(k, \tau)U^{-1}(\tau)r(k, \tau)}{tr[I - H(k, \tau)]},$$

(22)

$$\hat{\sigma}_n^2 = \frac{r^T(k, \tau)U^{-1}(\tau)r(k, \tau)}{tr[I - H(k, \tau)]},$$

(23)

$$\hat{\sigma}_n^2 = \frac{y^T U^{-1}(\tau)[I - H(k, \tau)]y}{tr[I - H(k, \tau)]},$$

(24)

and

$$\hat{\sigma}_n^2 = \frac{r^T(k, \tau)U^{-1}(\tau)H(k, \tau)r(k, \tau)}{tr[H(k, \tau)[I - H(k, \tau)]]},$$

(25)

where $\kappa$ and $\tau$ have been obtained and known.

In this most general case of a full weight matrix with a number of unknown coefficients, Wang (1998) proposed a GML estimate of the noise variance

$$\hat{\sigma}_n^2 = \frac{y^T U^{-1}(\tau)[I - H(k, \tau)]y}{n - t}.$$

(26)

In the case of ridge regression with the weight matrix $W$, Xu et al. (2006b) proposed first correcting the biases in the regularized residuals and then using the bias-corrected residuals to estimate the noise variance. As a result, they obtained the following estimator:

$$\hat{\sigma}_n^2 = \frac{\hat{\beta}_v W \hat{\beta}_v}{n - t + \kappa^4 tr[(A^T W A + \kappa I)^{-4}]}.$$

(27)

where

$$\hat{\beta}_v = y - A\hat{\beta}_e - \kappa A(A^T W A + \kappa I)^{-1}\hat{\beta}_e,$$

$$\hat{\beta}_e = (A^T W A + \kappa I)^{-1}A^T W y.$$

### 3 Iterative Generalized Cross-Validation

Under the framework of ridge regression, Xu et al. (2006b) proposed a two-step, iterative procedure to determine the regularization parameter and the variance components as formulated by (5). The regularization parameter $\kappa$ is determined by minimizing the MSE of the regularized solution of $\hat{\beta}$, assuming that the variance components are given. The other step is to estimate the variance components $\sigma$ and $\beta$. Since the regularized residuals are biased, they first use the regularized solution of $\hat{\beta}$ to estimate the biases of the residuals, remove the estimated biases from the regularized residuals and then apply the techniques of variance component estimation to derive the estimates of $\sigma$. The simulated results confirmed the almost unbiasedness of the estimated variance components. We should like to note that the MSE criterion has been thought to be not practically applicable by some researchers, this is not necessarily true. In fact, one can simply start with a small regularization parameter to compute the first regularized solution for $\hat{\beta}$, replace the true but unknown $\beta$ in the MSE criterion with the regularized solution, and finally continue to iteratively find the MSE-based regularization parameter. In other words, the iterative MSE criterion does not require the true values of $\sigma$ and $\beta$.

In this section, we will extend the work of Xu et al. (2006b) in two aspects: (i) instead of using the MSE criterion to determine the regularization parameter, we will use the GCV method to find the optimal $\kappa$. As a result of using the GCV approach, we will only need the knowledge of $\tau$ but not the unknown model parameters of $\hat{\beta}$ nor the value of any particular $\sigma_i^2$ and (ii) unlike Xu et al. (2006b) who assumed that $\mathbf{R} = \mathbf{I}$, we will directly deal with a general positive (semi-)definite matrix $\mathbf{R}$ in this paper.

As the first step or element of the iterative GCV method, we assume that the variance components $\sigma$ are available. We can then compute the vector $\tau$ from $\sigma$. In other words, we can assume that the weight matrix $\mathbf{W}$ is known at this moment. As a result of this assumption, we can straightforwardly use the weighted GCV method, as formulated in Section 2.1, to determine the regularization parameter $\kappa$. For convenience of reference, we write this version of the weighted GCV method as follows:

$$V(\kappa, \mathbf{W}_e) = \frac{1}{(1/n)tr(\tau H(k, \mathbf{W}_e)\mathbf{W}_e)} - \frac{(1/n)tr[I - H(k, \mathbf{W}_e)]^2}{(1/n)tr[I - H(k, \mathbf{W}_e)]^2},$$

(28a)

where

$$H(k, \mathbf{W}_e) = A(A^T \mathbf{W}_e A + \kappa \mathbf{R})^{-1} A^T \mathbf{W}_e,$$

(28b)

$$\mathbf{r}(\kappa, \mathbf{W}_e) = (I - H(k, \mathbf{W}_e))y.$$

(28c)

$$\mathbf{W}_e = U^{-1}(\tau).$$

(28d)

Obviously, this step of finding the minimizer of (28a) is essentially nothing but the conventional approach of GCV, and thus requires no further explanation.

To proceed to the second step or element of the iterative GCV method, we assume that an intermediate regularization parameter $\kappa$
has been obtained or given. We will now focus on the estimation of variance components $\sigma$. Although any technique of variance component estimation can be used here, we will analyse, without loss of generality, the technique of estimating $\sigma$ by using the quadratic forms of residuals, which was first developed by Helmert (1907) in geodesy and is actually also the basis of minimum norm quadratic unbiased estimation (MINQUE) of variance components developed by Rao (1970) (see also Rao & Kleffe 1988).

Given the regularized solution $\hat{\beta}_k$ and denoting the regularized residuals of the measurements $y$ by $r_k$, namely,

$$r_k = y - A\hat{\beta}_k,$$

we have the regularized solution of MINQUE to estimate the variance components of (5),

$$S_k \hat{\sigma}_k = q_k,$$

where the elements of the matrix $S_k$ and the vector $q_k$ are respectively given by

$$s_{ij}^k = tr[M_i U_j M_j], \quad (i, j = 1, 2, \ldots, m)$$

and

$$q_{ik}^k = r_i^k W_0 U_i r_k, \quad (i = 1, 2, \ldots, m).$$

Substituting $\hat{\beta}_k = N_k^{-1} A^T W_0 y,$

$$\hat{\beta}_k = N_k^{-1} A^T W_0 y,$$

we have the unbiased estimation (MINQUE) of variance components developed by Rao (1970) (see also Rao & Kleffe 1988).

Substituting $\hat{\beta}_k$ with its regularized solution gives

$$\hat{\beta}_k = \kappa A C \hat{\beta}_k.$$

Thus the estimate of variance components with a given $\kappa$ is given by

$$\hat{\sigma} = S^{-1} q,$$

where $S$ is equal to $tr(U_i M_i U_j M_j)$, and $q = \{ r_i^k W_0 U_i r_k, r_i^k W_0 U_i r_k, \ldots, r_i^k W_0 U_i r_k \}$. It is easy to prove that the regularized residuals $r_k$ are unbiased as an effect of regularization, and the bias of $r_k$ is simply listed below $b_k = \kappa A C \hat{\beta}_k.$

$$b_k = \kappa A C \hat{\beta}_k.$$

where

$$C = N_k^{-1} R.$$ Substituting $\beta$ with its regularized solution $\hat{\beta}_k$, we can readily estimate $b_k$, namely,

$$\hat{b}_k = \kappa A C \hat{\beta}_k.$$

Subtracting the bias estimate $\hat{b}_k$ from the regularized residuals $r_k$ of (29), we finally obtain the bias-corrected residuals, which are denoted by $\tilde{r}_k$ and read

$$\tilde{r}_k = y - A\hat{\beta}_k - \kappa A C \hat{\beta}_k.$$}

For convenience of deriving formulae for variance component estimation, we can equivalently rewrite $\tilde{r}_k$ as follows:

$$\tilde{r}_k = \kappa^2 A^2 \hat{\beta}_k + Q_e,$$

where

$$Q = I - A N_k^{-1} A^T W_t - \kappa A N_k^{-1} A^T W_t.$$ By neglecting the second- and higher-order biases of $r_k$, or equivalently to neglect the biases of $r_k$, we can then go forward to derive the formulae for estimating the variance components $\sigma$. More specifically, we first construct the quadratic form $F_i^T W_0 U_i W_0 \tilde{r}_k,$

derive its expectation, neglect the second order bias terms with $\beta$, and finally obtain the following equations:

$$E \{ F_i^T W_0 U_i W_0 \tilde{r}_k \} = \sum_{j=1}^m tr(U_i M_i U_j M_j^T) \sigma_j^2, \quad (i = 1, 2, \ldots, m),$$

where

$$M_i = W_0 [I - H(k, W_0) - \kappa A N_k^{-1} A^T W_0].$$

Now the problem is how we can use the theoretical results (34) to estimate the variance components, or equivalently, the weighting factors in the case of different types of data. In fact, given a regularization parameter $\kappa$ and a realization of measurements $y$, all the quantities in (34), except for the variance components, are all either known or directly computable from $y$. By removing the expectation operator from (34) and writing eq. (34) in matrix form, we obtain the linear system of equations

$$\tilde{\Sigma} = \tilde{q}.$$

where the $ij$th element of the matrix $\tilde{\Sigma}$ is equal to $tr(U_i M_i U_j M_j^T)$, and

$$\tilde{q} = \{ r_i^k W_0 U_i r_k, r_i^k W_0 U_i r_k, \ldots, r_i^k W_0 U_i r_k \}.$$ Thus the estimate of variance components is given by

$$\hat{\sigma} = S^{-1} \tilde{q}.$$ if $S$ is regular.

As a special example of (5), we assume a block-diagonal variance–covariance matrix with a number of unknown variance components for $y$

$$\Sigma = \begin{bmatrix} W_1^{-1} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & W_2^{-1} \sigma_2^2 & & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & W_m^{-1} \sigma_m^2 \end{bmatrix},$$

where $W_i$ are the weight matrices of the corresponding subsets of measurements of different types. The block-diagonal model often takes place in practice, in particular, if measurements of different types are collected. The variance components of (37) are often estimated, independently and componentwise. It is well known that if the model is well conditioned, the variance components are more often iteratively solved by

$$\hat{\sigma}_i^2 = (y_i - A_{i,LS}^T W_i(y_i - A_{i,LS})/f_i,$$\n
$$f_i = n_i - \frac{tr(A_i (A_i^T \Sigma_0^{-1} A_i)^{-1} A_i^T W_i)}{\sigma_0^2},$$

which has been shown to be almost unbiased and non-negative (see, e.g. Horn et al. 1975; Rao & Kleffe 1988). Here $\Sigma_0$ is equal to $\Sigma$ with $\sigma_i^2$ replaced by their initial or prior values $\sigma_0^2$, $A_{i,LS}$ is the least squares estimate of $\beta$, $f_i$ is the equivalent degrees of freedom for the subset of measurements $y_i$, $n_i$ is the number of elements of $y_i$, and $A_i$ is the submatrix of $A$ corresponding to the measurements $y_i$. However, if (1) is an inverse ill-posed problem, (38a) will not be able to produce a meaningful solution.

The regularized, bias-corrected version of (38a) can be obtained by applying the procedure as described in the derivation of (36). After a lengthy technical derivation, which is omitted here, we finally obtain

$$\hat{\sigma}_i^2 = \frac{F_i^T W_i \tilde{r}_i / (f_i)}{f_i (\kappa)},$$

$$f_i = n_i - \frac{tr(A_i (A_i^T \Sigma_0^{-1} A_i)^{-1} A_i^T W_i)}{\sigma_0^2},$$

which has been shown to be almost unbiased and non-negative (see, e.g. Horn et al. 1975; Rao & Kleffe 1988). Here $\Sigma_0$ is equal to $\Sigma$ with $\sigma_i^2$ replaced by their initial or prior values $\sigma_0^2$, $A_{i,LS}$ is the least squares estimate of $\beta$, $f_i$ is the equivalent degrees of freedom for the subset of measurements $y_i$, $n_i$ is the number of elements of $y_i$, and $A_i$ is the submatrix of $A$ corresponding to the measurements $y_i$. However, if (1) is an inverse ill-posed problem, (38a) will not be able to produce a meaningful solution.

The regularized, bias-corrected version of (38a) can be obtained by applying the procedure as described in the derivation of (36). After a lengthy technical derivation, which is omitted here, we finally obtain

$$\hat{\sigma}_i^2 = \frac{F_i^T W_i \tilde{r}_i / (f_i (\kappa))}{f_i (\kappa)}.$$
where
\[ f_i(\kappa) = n_i - tr \left( \left[ N_i^{-1} N_i \right] \right) - \kappa tr \left( C N_i^{-1} N_i \right) \]
\[ - \kappa^2 tr \left( C^2 N_i^{-1} N_i \right) - \kappa^3 tr \left( C^3 N_i^{-1} N_i \right), \]
with each \( N_i \) being derived from the corresponding \( A_i \) and the
weight matrix \( W_i \) given by (28d), namely, \( N_i = A_i^T [U^{-1}(\tau)]_{ii} A_i \)
and \([U^{-1}(\tau)]_{ii}\) is the \( i \)th diagonal block matrix of \( U^{-1}(\tau) \), \( F_i \) is the
\( i \)th subvector of \( F \), and corresponds to the submatrix \( A_i \).

In particular, if there is only one unknown variance component,
namely, the noise variance or the variance of unit weight, (39)
becomes
\[ \hat{\sigma}_{W}^2 = \frac{\tilde{f}^T_i W_i \tilde{f}_i}{n - t}. \]

4 EXPERIMENTS AND RESULTS

The experiments in this section are designed and used for three
purposes: (i) to demonstrate and evaluate the iterative GCV method;
(ii) to compare the performances of eight estimators of the noise
variance, namely, \( \hat{\sigma}_{W}^2 \) of (21) as a generalization of Wahba’s (1983)
estimator of \( \sigma^2 \), \( \hat{\sigma}_{W}^2 \) of (22), \( \hat{\sigma}_{W}^2 \) of (23), \( \hat{\sigma}_{W}^2 \) of (24), \( \hat{\sigma}_{W}^2 \) of
(25) proposed by Galatsanos & Katsaggelos (1992) as the second
generalization of Wahba’s (1983) estimator, \( \hat{\sigma}_{W}^2 \) of (26) proposed
by Wang (1998) and Galatsanos & Katsaggelos (1992) as an ML or
GML estimator of \( \sigma^2 \), our estimator \( \hat{\sigma}_{W}^2 \) given in (40), and \( \hat{\sigma}_{W}^2 \)
as a simplified formula of (40) and (iii) to compare our estimator
of variance components with those derived from the estimation
of the unknown parameters of the weight matrix, as proposed
by Diggle & Hutchinson (1989) and Wang (1998), keeping in mind
that an accurate estimate of the variance components is essential
to obtain the correct weighting factors of different types of data
in joint geophysical inversion and further to warrant the determination
of a proper data-driven regularization parameter.

The experiments to be reported in this section are based on two
synthetic examples: one is purely mathematical and the other is
(geodetically) physical. The first example is an integral equation of
the first kind modified from the first example of Phillips (1962).
The second example is to determine the gravity anomalies on the
surface of the Earth from satellite measurements, which is known
as a downward continuation problem and is a typical example of
inverse ill-posed problems in geophysics. The two examples are
also purposely designed such that the design matrix \( A \) is extremely
ill-conditioned for the first example (the smallest eigenvalue being
practically equal to zero with MATLAB on a DELL Dimension
8300 Desktop) and only moderately ill-conditioned for the second
element, with the condition number and the smallest eigenvalue
being around \( 10^8 \) and \( 10^4 \), respectively. Roughly speaking, an ill-
posed problem may be said to be moderately ill-conditioned, if
the corresponding normal matrix could be inverted in the classical
sense; otherwise, it could be said to be extremely ill-conditioned.

As the first example, the modified integral equation of the first
kind without random noise is given by
\[ y(z) = \int_{-3}^{3} K(x - z) f(x) dx, \]
where
\[ y(z) = 6.0 + 3.0 \cos(\pi z/3), \]
\[ K(z) = 1.0 + \cos(\pi z/3), \]
\[ -6 \leq z \leq 6, \]
and the exact solution \( f(x) \) of (42a) is known as
\[ f(x) = 1.0 + \cos(\pi x/3). \]
As in Phillips (1962), we use the composite Simpson’s rule of
quadrature to discretize the integral equation (42a). The correspond-
ing matrix \( R \) can be found in Phillips (1962) and Twomey (1963).
The number of the points \( f(x_i) \) is set to 60. In the following simu-
lations, we will generate Gaussian random errors and add them to
the true measurements of \( y(z) \).

The second example is to estimate the gravity values on the
surface of the Earth from gravity tensors collected at the altitude of
a gravity satellite. The discretized linear observation equations are
\[ 4\pi R/10T_{p} = \sum_{k} N_k(r, \psi) \Delta g_k + \epsilon, \]
(see, e.g. Xu 1998), where \( T_{rr} \) is the second derivative of the anomalous potential \( T \) with respect to \( r \) and is observable in satellite gradiometry, \( \Delta g_{ij} \) are the block-mean gravity anomalies to be estimated, and \( R \) is the mean radius of the Earth. \( S'(r, \psi) = R^2 S''(r, \psi) \), where \( S''(r, \psi) \) is the second derivative of the generalized Stokes' kernel with respect to \( r \). More details on the above equation can be found in Xu (1998). If the reader is interested in satellite gradiometry, please consult Rummel (1986). As our second synthetic example, we only simulate 100 unknown gravity anomalies and 400 measurements \( T_{rr} \).

4.1 Comparisons of the estimators of the noise variance

As the first part of the experiments with the first example, we will compare the performances of the eight estimators of the noise variance by simulation, assuming that the weight matrix is an identity matrix. To make sure that the simulation results are convincing and not significantly affected due to the lack of measurements, we set the number of measurements \( y(z_i) \) to 20 000. The positions of the measurements are, however, randomly generated. We then generate three sets of Gaussian random errors, which correspond to the standard deviations of 1, 5 and 10 per cent of the maximum absolute values of \( y(z_i) \), or more specifically, 0.09, 0.45 and 0.90, respectively. As a result, we obtain three sets of randomly simulated measurements \( y(z_i) \).

Since the GCV method has been observed to produce too small a regularization parameter from time to time (see, e.g. Diggle \& Hutchinson 1989; Thompson et al. 1991), in order to minimize the effect of a particular choice of regularization parameter on the comparison of the eight estimators of the noise variance, we compare the estimators of the noise variance over an interval of \( \kappa \), or more specifically, \( \kappa \in [10^{-7}, 0.05] \). By choosing \( R \) to warrant the smoothness of the second derivatives of the function \( f(x) \) to be estimated, as given by Phillips (1962), we compute all the eight estimators of the noise variance, namely, \( \hat{\sigma}^2_{SW} \) of (21), \( \hat{\sigma}^2_{NW} \) of (22), \( \hat{\sigma}^2_{BW} \) of (23), \( \hat{\sigma}^2_{BK} \) of (24), \( \hat{\sigma}^2_{AW} \) of (25), \( \hat{\sigma}^2_{WX} \) of (26), \( \hat{\sigma}^2_{PW} \) of (40) and \( \hat{\sigma}^2_{KW} \) of (41), as a function of the regularization parameter \( \kappa \), and plot their estimated errors of the noise variance in Fig. 1. Since the true values of the noise variances are known in the simulations, we can compute the estimated error by \( (\hat{\sigma}^2 - \sigma^2) \), where \( \sigma^2 \) and \( \hat{\sigma}^2 \) are the estimated and true values of the noise variance, respectively. It can be clearly seen from Fig. 1 that five of the estimators, namely, \( \hat{\sigma}^2_{SW} \), \( \hat{\sigma}^2_{NW} \), \( \hat{\sigma}^2_{BW} \) and \( \hat{\sigma}^2_{PX} \), accurately reproduce the three simulated noise variances. The estimated errors of these estimators are extremely stable within the scale shown in the figure. In particular, it is interesting to see that the results of \( \hat{\sigma}^2_{SW} \), \( \hat{\sigma}^2_{NW} \), \( \hat{\sigma}^2_{BW} \) and \( \hat{\sigma}^2_{PX} \) are almost identical within the scale. They produce basically the same results as \( \hat{\sigma}^2_{SW} \) for the first level of noise. For the other two levels of noise, \( \hat{\sigma}^2_{SW} \), \( \hat{\sigma}^2_{BW} \) and \( \hat{\sigma}^2_{PX} \) are better than \( \hat{\sigma}^2_{SW} \) for the second level of noise but worse for the third level of noise. All these five estimators perform much better than the remaining three estimators \( \hat{\sigma}^2_{BK} \), \( \hat{\sigma}^2_{AW} \) and \( \hat{\sigma}^2_{PW} \) [the three line almost not distinguishable in the scale shown in panel (E) of Fig. 1] and \( \hat{\sigma}^2_{KW} \). Although (25) was reported to be as good as (21) by Galatsanos \& Katsaggelos (1992), this is not the case here with any of the three levels of noise. In fact, the results from using \( \hat{\sigma}^2_{BK} \) are simply erroneous for a small regularization parameter [compare panel (E) of Fig. 1]. Even after the estimates of the three

![Figure 1](https://academic.oup.com/gji/article-fig/179/1/182/734731)
levels of noise become stable for a relatively large $\kappa$, the maximum estimated errors can still be too large by a factor of 214.03, 7.59 and 1.16 for the three levels of noise, respectively. The GML estimator $\hat{\sigma}_{GK}^2$ performs significantly better than $\hat{\sigma}_{NW}^2$. The maximum relative error of $\hat{\sigma}_{GK}^2$ occurs with the first level of noise and can be as large as 203 per cent. This error reduces to about 9 per cent for the second level of noise. However, when compared with the results of the five best performing estimators, the errors of $\hat{\sigma}_{GK}^2$ are still much larger by more than two orders of magnitude for the first level of noise in the case of a small regularization parameter. Nevertheless, the differences from the three best estimators significantly decrease with the increase of the regularization parameter. We should also like to note that although Carter & Eagleson (1992) suggested that $\hat{\sigma}_{GK}^2$ was basically as good as $\hat{\sigma}_{NW}^2$, this is clearly not true with the current experiments [compare panels (C) and (D) of Fig. 1].

Since estimators of ridge type are, at least, as popular in inverse problems as GCV smoothing regularization, we decide to compute the eight estimators of the noise variance as a function of $\kappa$ by setting $R$ to an identity matrix to imitate ridge regression (Hoerl & Kennard 1970). The estimated errors of the eight estimators of the noise variance at the three different levels are shown in Fig. 2. It can be clearly seen that except for $\hat{\sigma}_{GK}^2$ of (25), all the other seven estimators perform excellently. In particular, $\sigma_{SW}^2$ and $\sigma_{KY}^2$ are now as good as $\hat{\sigma}_{SW}^2$, $\hat{\sigma}_{SW}^2$, $\sigma_{SW}^2$, $\sigma_{KY}^2$ and $\hat{\sigma}_{KY}^2$. The results are inconsistent with those by Galatsanos & Karaggeles (1992). They tested and found that the estimators $\hat{\sigma}_{SW}^2$ and $\hat{\sigma}_{KY}^2$ had a better performance with $R$ than with an identity matrix $I$. However, in our simulations, these two estimators (25) and (26) perform better with an identity matrix $I$ than with a general smoothing matrix $R$ (though only negligible in the case of $\hat{\sigma}_{SW}^2$). In fact, the five best performing estimators basically produce the same nice results, no matter which matrix, $R$ or $I$, is used. We should like to point out that $R$ and $I$ result in significantly different function recoveries, which are not shown here.

A further test has also been conducted to investigate the effect of weights of measurements on the estimation of the noise variance. We generate different weights for the 20,000 measurements by using a random number generator with a uniform distribution. By assuming a general smoothing matrix $R$ and with the above generated different weights, we compute the eight estimators of the noise variance, again, as a function of the regularization parameter. The estimated errors are shown in Fig. 3. Comparing this figure with Fig. 1, we can clearly see that the performances of the five best estimators, namely, $\hat{\sigma}_{SW}^2$, $\hat{\sigma}_{SW}^2$, $\sigma_{SW}^2$, $\sigma_{KY}^2$ and $\sigma_{KY}^2$, remain excellent. $\sigma_{SW}^2$ and $\sigma_{KY}^2$ are basically not affected by the inclusion of the weight matrix. However, $\sigma_{GK}^2$ deteriorates significantly.

Now let us come to the second example of gravity, for which we simulate two levels of noise, with a standard deviation of $T_0$ being equal to 0.01 and 0.005 Eötvös units (1 Eötvös unit = $10^{-26}$ gal cm$^{-1}$), respectively. Taking the original observation equations (43) into account, we rescale these two levels of noise into two dimensionless variances of unit weight, which are equal to $\sigma_1^2 = 1$ and $\sigma_2^2 = 0.25$, respectively. Following the same procedure as in the first example, we compute and plot the estimated errors of the eight variance estimators for both $\sigma_1^2$ and $\sigma_2^2$ as a function of $\kappa$. Figure 1 is on the derivative-smoothing matrix $R$.

Figure 2. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 3. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the five levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 4. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 5. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 6. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 7. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 8. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 9. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 10. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 11. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 12. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 13. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 14. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 15. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 16. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$. 

Figure 17. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it is based on the identity matrix $I$, while Fig. 1 is on the derivative-smoothing matrix $R$.
Figure 3. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the three levels of noise. This figure is the same as Fig. 1, except for that it uses a random number generator to generate different weights for the measurements, while Fig. 1 assumes that all the measurements are of the same weight.

Figure 4. The estimated errors of the eight estimators of the noise variance as a function of the regularization parameter $\kappa$ for the two levels of noise of satellite gradiometric observables. They are rescaled into two dimensionless standard deviations, namely, $\sigma^2_1 = 1$ and $\sigma^2_2 = 0.5^2$. Shown in this figure are the estimated errors of $\hat{\sigma}^2_1$ (dashed line) and $\hat{\sigma}^2_2$ (dotted line), with $\mathbf{R} = \mathbf{I}$: $\hat{\sigma}^2_{2w}$ of (21) in panel (A), $\hat{\sigma}^2_{2w}$ of (22) in panel (B), $\hat{\sigma}^2_{2w}$ of (23) in panel (C), $\hat{\sigma}^2_{1w}$ of (24) in panel (D), $\hat{\sigma}^2_{2w}$ of (25) in panel (E), $\sigma_{2w}$ of (26) in panel (F), $\sigma_{2w}$ of (40) in panel (G) and $\sigma_{2w}$ of (41) in panel (H).

A is not full of rank, and if $\kappa = 0$, then $\{\mathbf{A} - \mathbf{A}(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{A}\}$ should be invariant with respect to a generalized inverse $(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}$ and is actually equal to a zero matrix. We also know that a generalized inverse of $(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}$ can be computed as $(\mathbf{A}^T\mathbf{W}\mathbf{A} + \kappa \mathbf{I})^{-1}$ if $\kappa$ is positive and sufficiently small. Since an extremely ill-posed problem could be treated as a practically rank-defect problem, this means that for a small regularization parameter $\kappa$, $\{\mathbf{A} - \mathbf{A}(\mathbf{A}^T\mathbf{W}\mathbf{A} + \kappa \mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{A}\}$ should be almost invariantly equal to zero. This explains very well why $\hat{\sigma}^2_{2w}$ and alike perform excellently in the first example. However, with the improvement of ill-posedness, or if $(\mathbf{A}^T\mathbf{W}\mathbf{A})$ is of full rank with the smallest eigenvalue significantly different from zero, then the deviation of

\( \{ A - (A^TWA + \kappa R)^{-1}A^TWA \} \) from its invariant property of zero will become significant with the increase of \( \kappa \). This nicely explains why \( \sigma_{pX}^2 \) and \( \sigma_{kS}^2 \) are much better than \( \sigma_{S}^2 \) with the gravity example, since the biases of \( \sigma_{pX}^2 \) and \( \sigma_{kS}^2 \) have been corrected up to the first order of approximation.

All these simulations have also clearly shown that the three best estimators of the noise variance, namely, \( \sigma_{S}^2 \) of (21), \( \sigma_{P}^2 \) of (40), and \( \sigma_{XS}^2 \) of (41), are very insensitive to the regularization parameter. In other words, the regularization parameter can become unstable, if one solves for it by equating the estimator of the noise variance to a given value.

### 4.2 Comparisons of the iterative GCV method with the weighted GCV, weighted GML and weighted PML approaches

Given the ill-posed model (1) and data distribution, a successful geophysical joint inversion depends essentially on two key elements: (i) correct weighting factors of different types of data, or equivalently, correct estimation of the variance components and (ii) an appropriate regularization parameter \( \kappa \). In this second part of the experiment with the integral equation (42) of the first kind, we will now compare our iterative GCV method with the weighted GCV criterion \( V(\kappa, \tau) \) of (9a), two weighted GML criteria (11) and (13), and the weighted PML criterion \( L(\kappa, \tau) \) of (14). The iterative GCV method can either be based on the bias-corrected estimation (35) or the regularized MINQUE estimation (30) of variance components. To conduct this part of the experiment, we transform the three levels of noise used in the first part of the experiment into three variance components, which are assumed to correspond to the first 6000 measurements, the second 6000 measurements and the remaining 8000 measurements, respectively. In this experiment, we do not assume correlation among the measurements. We repeat the same experiment independently for 200 times, each with a newly generated set of 20000 Gaussian random numbers. In all the computations, we implement the BFGS Quasi-Newton method (see, e.g. Fletcher 2000) to solve the non-linear optimization objective functions (9a), (11), (13) and (14). In the case of the iterative GCV method, we directly use the golden section search as implemented in MATLAB to find the optimal regularization parameter.

The 200 sets of the three variance components estimated by using the four methods (9a), (11), (13) and (14) are shown in Fig. 6, with the corresponding regularization parameters plotted in Fig. 7. Although the weighted PML criterion has the best performance of estimating the variance components for these 200 sets of independent replications, it is far from being able to produce satisfactory results. Actually, if we compare the three lines in each of panels of Fig. 6 with the true values of the three variance components \((-2.0915, -0.6936, -0.0915)\) in logarithmic scale, we can immediately conclude that none of these four methods are able to recover the three variance components, or equivalently, the weighting factors of the three data sets of different accuracy. In particular, the weighted GCV method (9a) and the second GML criterion (13) are erroneous by an order of magnitude \(10^7\). Since this order of magnitude corresponds to the tolerance error set in the computation, we expect even worse results if we lower the tolerance error. As far as the regularization parameter is concerned, the weighted GCV method produces a rather stable regularization parameter in all the 200 replications of experiment, while the two weighted GML methods often result in too small a regularization parameter. However, the second weighted GML method also creates an unacceptably large regularization parameter from time to time (compare the black-solid line of Fig. 7). On the other hand, the weighted PML method produces a highly oscillating regularization parameter, which is often too large, as can be seen in the green, dash–dotted line of Fig. 7.

We should like to note, however, that the poor estimates of variance components are inconsistent with the results reported by Wang (1998), where Wang (1998) concluded that the GML method worked well in all cases to simultaneously estimate the unknown function.
Iterative GCV for fusing heteroscedastic data

Figure 6. The estimates of the three variance components: panel (A) – the weighted GCV method (9a); panel (B) – the first GML method (11); panel (C) – the second GML method (13) and panel (D) – the weighted PML criterion (14). In each of these panels, the three variance components are shown in the blue-dotted, red-dashed and black-solid lines, respectively.

Figure 7. The estimates of the regularization parameters: the blue-dotted line – the weighted GCV method (9a); the red-dashed line – the first GML method (11); the black-solid line – the second GML method (13) and the green dash–dotted line – the weighted PML criterion (14).

and the unknown correlation parameter of the time-series example. In addition, Wang (1998) found that the GCV method worked reasonably well to estimate the unknown correlation parameter as well. To further clarify the worst performance of estimating the variance components by the weighted GCV method theoretically, let us assume, for simplicity but without loss of generality, (i) \( m = 2 \) in (5); and (ii) both \( A_1 \) and \( A_2 \) are reasonably conditioned. We further assume that \( y_1 \) and \( y_2 \) are statistically independent. As a result, by fixing \( l = 1 \), we can readily denote the inverse of the matrix \( U(\tau) \) by

\[
U^{-1}(\tau) = \begin{bmatrix} I_1 & 0 \\ 0 & \tau I_2 \end{bmatrix},
\]

where \( \tau = \sigma_1^2/\sigma_2^2 \). Then it is rather easy to prove that given a small \( \kappa \), the denominator of the weighted GCV function (9a), namely, the trace of \( \{I - H(\kappa, \tau)\} \), basically remains constant for any weight.
very large variance component will result in a very small weight or a decrease (or increase) of $\tau$. Therefore, minimizing (9a) should finally result in a very small value of $\tau$. In other words, the weighting factor between two sets of data is totally incorrect. To give the reader a better idea, we fix the first two weights to their true values and the regularization parameter $\kappa$ to 0.001, and plot the denominator of (9a) and the weighted GCV value as a function of the remaining weight in Fig. 8. This figure clearly shows that minimizing the weighted GCV function (9a) will result in a very small weight or a very large variance component $\sigma^2_3$, which is exactly what we have obtained and reported in the above. Thus we can theoretically conclude that the weighted GCV function is not suitable for estimating the unknown variance components, nor for the weighting factors of data of different types. This theoretical analysis should also apply to the first weighted GML method.

We will now investigate the performance of the iterative GCV method by using the example of the integral equation of the first kind. The 200 sets of absolute errors of the three variance components are shown in Fig. 9. It can be seen from this figure that the first variance component is most precisely estimated. In fact, the estimated errors of the variance components are roughly proportional to the variance components themselves. To remove the effect of size of variance components, we show in Fig. 10 the relative errors of the estimated variance components, together with the corresponding regularization parameter. The statistics of the 200 relative errors for each variance component are summarized in Table 1. The standard deviations of the relative estimated errors of variance components are roughly consistent with their theoretical values. Thus the iterative GCV method is indeed able to recover the variance components from measured data. Panel (D) also shows that the iterative GCV method results in a rather stable estimate of regularization parameter. In other words, the iterative GCV method can be used to simultaneously determine the correct weighting factors of the measurements and a stable regularization parameter for the inverse ill-posed problem.

Motivated by the success of Wahba’s estimate of the noise variance, we additionally compute the naive version of the MINQUE estimation of variance components without applying the bias corrections to the regularized residuals. The differences from the estimates with the bias corrections are shown in Fig. 11. It is obvious that the differences are negligible in comparison with the estimated errors of these variance components. Thus the naive MINQUE method could also be used to precisely estimate the variance components in this extremely ill-posed problem. Finally, recognizing the fact that variance component estimation was proposed to regularize ill-posed problems under the Bayesian framework by interpreting the matrix $R$ as the prior weight matrix of the unknown model parameters (see, e.g. Schwintzer 1990), we implement the algorithm to determine the regularization parameter. The average value of the regularization parameter over the 200 replications is equal to 192.48, with the minimum and maximum values being equal to 183.64 and 199.39, respectively. Obviously, the regularization parameters obtained in the experiments are far too large to be acceptable.

For the second example of gravity, we follow the same procedure as in the first example to transform the two different levels of noise into two variance components, each corresponding to 200 data. No correlation is assumed among the measurements. As a result, we construct the second (moderately) ill-posed problem of 400 data, 100 model unknowns and two variance components. As in Section 4.1, we will report the experiments with the two dimensionless variance components $\sigma^2_1$ and $\sigma^2_2$. Since the weighted GCV, weighted GML and weighted PML methods have been shown, numerically and partly theoretically, to fail to produce satisfactory results, we will focus on two versions of the iterative GCV, one with

Figure 8. The weighted GCV function (9a) and its denominator as a function of the only weight, which corresponds to the third variance component: the blue dotted line – the weighted GCV function (9a) and the red solid line – the denominator of (9a).
bias-corrected variance components and the other directly with the regularized MINQUE-estimated variance components. The algorithm terminates if the improvements of both components are smaller than $10^{-5}$.

By using the iterative GCV with bias-corrected variance components and with the 200 sets of independently simulated data, we obtain 200 sets of independent estimates of the variance components and 200 sets of regularization parameters. The absolute and relative errors of the estimates over the 200 independent replications are shown in Figs 12 and 13, respectively. Obviously, the variance components and the regularization parameter are satisfactorily estimated. One may argue that the estimates of the variance components seem to oscillate over the 200 replications. However, this is acceptably normal due to a relatively small number of data and can be well explained theoretically. In fact, it is well known that the standard deviation of the variance estimate

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2$$

is equal to $\sqrt{2/n} \sigma^2$, where $\epsilon_i \sim N(0, \sigma^2)$. If we have 200 data points, then the standard deviation for the estimate $\hat{\sigma}^2$ is 10 per cent, which explains the relative errors in Fig. 13 well.

On the other hand, if we directly use the regularized solution and the MINQUE method to estimate the variance components and then use the iterative GCV to find the regularization parameter, we can also obtain 200 sets of variance components and regularization parameters. The differences in the estimated variance components and the regularization parameters between these two versions of the iterative GCV method (with and without correcting the biases in the estimated variance components) are plotted in Fig. 14. Since the regularization parameters are all small, it is not surprising to see that the differences between the results with and without correcting the biases of the estimated variance components are very small, as was also exactly the case that we have observed in the first example. However, the differences will become significant with the increase of the regularization parameter, which can be clearly seen in the plots of the absolute (Fig. 15) and relative (Fig. 16) errors of the estimated variance components as a function of $\kappa$. In this case, the bias-corrected iterative GCV performs significantly better than the naive MINQUE estimation of variance components, if the regularization parameter is not small. In particular, the variance components are incorrectly estimated by a factor of about 6000 with $\kappa = 0.06$, as can be clearly seen from panel (C) of Fig. 16. With the increase of the regularization parameter, it would not be surprised to see that some of the variance components can be incorrectly estimated by a factor of far more than 6000.

5 CONCLUSIONS

Combining measurements of different types to solve inverse ill-posed problems is increasingly important. The stochastic model of measurements can either be represented in this case by a number of unknown variance components or by a number of unknown correlation parameters. Given an ill-posed geophysical problem and data distribution, a successful joint inversion depends on two basic elements: (i) correct weighting factors of data of different types and (ii) an appropriate regularization parameter. Thus, one will have to simultaneously estimate the unknown parameters of the stochastic model (or equivalently, the weighting factors of data of different types) and determine the optimal regularization parameter. We have
extended the GCV, GML and PML methods proposed by Diggle & Hutchinson (1989) and Wang (1998) to solve inverse ill-posed problems with a number of unknown variance components. Although these methods were shown by Diggle & Hutchinson (1989) and Wang (1998) to work reasonably well for simultaneously determining the unknown correlation parameters and the regularization parameter, our simulations reported in Section 4 have clearly demonstrated that they are not suitable for estimating the unknown variance components. We have also provided a theory to explain why the weighted GCV criterion always tends to produce very large values for some of the unknown variance components and is unable to recover the variance components correctly.

Motivated by the success of Xu et al. (2006b), we have extended the iterative MSE-based method to cross-validation. The iterative GCV method proposed here consists of two steps: (i) to estimate the unknown variance components by using the bias-corrected

Figure 10. The relative estimated errors of the three variance components and the corresponding regularization parameter by using the iterative GCV method: the relative errors (in per cent) of $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ are shown in panels (A), (B) and (C), respectively, and the corresponding regularization parameter in logarithmic scale shown in panel (D).

Figure 11. The differences of the three estimated variance components with and without applying the bias corrections to the regularized residuals: $\hat{\sigma}_1^2$ – the red dotted line; $\hat{\sigma}_2^2$ – the blue solid line and $\hat{\sigma}_3^2$ – the green dashed line.
Iterative GCV for fusing heteroscedastic data

Figure 12. The absolute estimated errors of the two dimensionless bias-corrected variance components by applying the iterative GCV method to the satellite gradiometric data. The errors of $\sigma_1^2$ and $\sigma_2^2$ are shown in the upper and lower panels, respectively.

Figure 13. The relative estimated errors of the two dimensionless bias-corrected variance components by the iterative GCV method (in per cent). The relative errors of $\sigma_1^2$ and $\sigma_2^2$ are shown in panels (A) and (B), respectively. Also shown in this figure are the regularization parameters over the 200 replications (panel C).

estimators, given a regularization parameter and (ii) to use the weighted GCV method to optimally determine the regularization parameter with the estimated variance components. The simulations have shown that the iterative GCV scheme works well to simultaneously estimate the unknown variance components and determine the regularization parameter. As a general guiding rule, one may simply estimate the variance components with a sufficiently small regularization parameter and then use them and the GCV method to find the regularization parameter without further iteration, as could be inferred from all the above simulation results. Unlike Schwintzer (1990), we do not assume any prior information on the unknown function. We do not treat the regularization parameter as if it were an extra variance component of the prior values of the unknown function either. The iterative GCV method is solely data-driven without any assumption of prior information. The simulations have shown that the regularization parameter can be overestimated by a few orders of magnitude, if it is treated as a new variance component. The simulations have also shown that without correcting the biases of the estimated variance components, some of the variance components can be readily estimated incorrectly by a factor of more than 6000.
Comparisons between iterative GCV with and without bias corrections

Figure 14. The differences of the two estimated dimensionless variance components with and without applying the bias corrections to the regularized residuals. The results on $\sigma_1^2$ and $\sigma_2^2$ are shown in the dotted and solid lines, respectively. Also shown in this figure are the differences of the regularization parameters in the dash-dotted line.

Figure 15. The estimated errors of the two dimensionless variance components as a function of the regularization parameter. The upper panel shows the results with the bias corrections to the variance components, while the lower two panels show the results by directly using the MINQUE method to estimate the two dimensionless variance components. Since the results get worse in the last part of the regularization parameter, we split the MINQUE results into two panels for a better view.

The factor should increase with the increase of the regularization parameter.

We have also discussed the estimation of the noise variance, derived a new bias-corrected estimator of the noise variance with a most general smoothing positive (semi-)definite matrix $R$ and obtained a simplified formula. The two new estimators of the noise variance have been numerically compared with six other estimators in the literature. The simulations have shown that the five best performing estimators, namely, $\hat{\sigma}_{SW}^2$ of (21), $\hat{\sigma}_{NW}^2$ of (22), $\hat{\sigma}_{BW}^2$ of (23), $\hat{\sigma}_{P_X}$ and $\hat{\sigma}_{XS}$ work reasonably well in all the cases. However, if an inverse problem is only moderately ill-posed, the two new estimators outperform any of the existing methods. The estimator by Galatsanos & Katsaggelos (1992) is basically not able to produce a reasonable noise variance. The GML-based estimator works well...
Figure 16. The relative errors of the two dimensionless variance components as a function of the regularization parameter (in per cent). This figure is the same as Fig. 15, except for that the relative errors, instead of the absolute errors, are shown here.

with a ridge-type solution; however, its performance significantly deteriorates for a most general $\mathbf{R}$, if the regularization parameter is small. The simulations have also shown that the best estimators of the noise variance are extremely insensitive to the regularization parameter. In other words, if one would like to determine the regularization parameter by equating the estimator of noise variance to a given (prior) value, then a small change of the given value could result in a significant change in the regularization parameter.

Acknowledgments

This work is partly supported by a Grant-in-Aid for Scientific Research B19340129. The author would like to thank Prof A. Dermani and Prof. J. Kusche for their very constructive comments, and the Editor, Prof M. Diamant for his request for a substantial revision to explain geophysical relevance/implications and potential geophysical applications. The comments and the request have been fully implemented to substantially revise the paper and resulted in a significant improvement of the paper.

References


