Geophys. J. Int. (2004) **156,** 411–425 doi: 10.1111/j.1365-246X.2004.02190.x

A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems

Colin G. Farquharson and Douglas W. Oldenburg

UBC—Geophysical Inversion Facility, Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, B.C., V6T 1Z4, *Canada. E-mail: farq@eos.ubc.ca*

Accepted 2003 October 14. Received 2003 July 14; in original form 2002 April 23

SUMMARY

Two automatic ways of estimating the regularization parameter in underdetermined, minimum-structure-type solutions to non-linear inverse problems are compared: the generalized cross-validation and L-curve criteria. Both criteria provide a means of estimating the regularization parameter when only the relative sizes of the measurement uncertainties in a set of observations are known. The criteria, which are established components of linear inverse theory, are applied to the linearized inverse problem at each iteration in a typical iterative, linearized solution to the non-linear problem. The particular inverse problem considered here is the simultaneous inversion of electromagnetic loop–loop data for 1-D models of both electrical conductivity and magnetic susceptibility. The performance of each criteria is illustrated with inversions of a variety of synthetic and field data sets. In the great majority of examples tested, both criteria successfully determined suitable values of the regularization parameter, and hence credible models of the subsurface.

Key words: electromagnetic methods, inversion, regularization.

1 INTRODUCTION

The inverse problem of determining a plausible spatial variation of one or more physical properties within the Earth that is consistent with a finite set of geophysical observations can be solved by formulating it as an optimization problem in which an objective function such as

$$\Phi(\mathbf{m}) = \phi_\mathrm{d}(\mathbf{m}) + \beta\phi_\mathrm{m}(\mathbf{m}) \tag{1}$$

is minimized. The vector $\mathbf{m}$ contains the M parameters in the Earth model, ϕ_d is a measure of data misfit, ϕ_m is a measure of some property of the Earth model, and β is the regularization parameter that balances the effects of ϕ_d and ϕ_m. Here, the typical sum-of-squares misfit:

$$\phi_\mathrm{d}(\mathbf{m}) = \left\| \mathbf{W}_\mathrm{d}[\mathbf{d}^\mathrm{obs} - \mathbf{d}(\mathbf{m})] \right\|^2 \tag{2}$$

is considered, where $\mathbf{d}^\mathrm{obs} = (d_1^\mathrm{obs}, \ldots, d_N^\mathrm{obs})^T$ is the vector containing the observations, $\mathbf{d}(\mathbf{m})$ is the vector containing the data computed for the model $\mathbf{m}$, and $\| \cdot \|$ represents the l_2 norm. It is assumed that the noise in the observations is Gaussian and uncorrelated. The weighting matrix $\mathbf{W}_\mathrm{d}$ is therefore the diagonal matrix $\mathbf{W}_\mathrm{d} = \mathrm{diag}\{1/\sigma_1, \ldots, 1/\sigma_N\}$, where σ_i is the standard deviation of the noise in the ith observation. It is also assumed that the relative sizes of the standard deviations are known, with only their absolute sizes unknown. That is, σ_i can be expressed as $\sigma_0\tilde{\sigma}_i$, where the $\tilde{\sigma}_i (i = 1, \ldots, N)$ are known and the constant σ_0 is unknown.

Also, the following measure of the amount of structure in the model is considered:

$$\phi_\mathrm{m}(\mathbf{m}) = \alpha_s \left\| \mathbf{W}_s \left(\mathbf{m} - \mathbf{m}_s^\mathrm{ref}\right) \right\|^2 + \alpha_z \left\| \mathbf{W}_z \left(\mathbf{m} - \mathbf{m}_z^\mathrm{ref}\right) \right\|^2, \tag{3}$$

where $\mathbf{m}_s^\mathrm{ref}$ and $\mathbf{m}_z^\mathrm{ref}$ are two possibly different reference models. The elements of the weighting matrices $\mathbf{W}_s$ and $\mathbf{W}_z$ are obtained by substituting the discretized representation of the Earth into

$$\varphi_s(m) = \int_{z=0}^{\infty} \left| m(z) - m_s^\mathrm{ref}(z) \right|^2 dz, \tag{4}$$

and

$$\varphi_z(m) = \int_{z=0}^{\infty} \left| \frac{d}{dz} \left[m(z) - m_z^\mathrm{ref}(z) \right] \right|^2 dz, \tag{5}$$

respectively, and approximating the derivative in eq. (5) by a finite difference. The coefficients α_s and α_z enable the appropriate balance between the two components of ϕ_m to be achieved for a particular problem. It is also assumed that the discretization of the Earth model is sufficiently fine that the discretization does not regularize the problem. This invariable makes the discrete inverse problem of finding the parameters in the model underdetermined, thus mimicking the underlying inverse problem of finding the physical property as a function of position.

A crucial part of the solution process is deciding on a suitable value of the regularization parameter β. It should be chosen such that the observations are reproduced to a degree that is justified by the noise, and that there is not excessive structure in the constructed

model. If the standard deviations of the noise are known, the expectation, $E(\phi_d)$, of the misfit given by eq. (2) is equal to the number of observations, N. A straightforward univariate search can be used to find a value of β that results in $\phi_d \approx N$. This is known as the discrepancy principle (see, for example, Hansen 1997), and has been used extensively for geophysical inverse problems (see, for example, Constable *et al.* 1987). If, however, the noise in the observations is not well known, some means of automatically determining an appropriate value of the regularization parameter during the course of an inversion is required.

For linear inverse problems, several techniques have been developed for automatically estimating an appropriate regularization parameter when the observations are contaminated with Gaussian noise of uniform, but unknown, standard deviation. A number of authors, for example Wahba (1990) and Hansen (1997), choose the value of the regularization parameter that minimizes the generalized cross-validation (GCV) function. Hansen (1997) chooses the value corresponding to the point of maximum curvature on the 'L'-shaped curve obtained when ϕ_d is plotted as a function of ϕ_m for all possible values of the regularization parameter. And Akaike (1980) chooses the value that minimizes a Bayesian information criterion (ABIC) function.

There have been a few reports in the mathematical literature of the successful use of the GCV and L-curve criteria to choose the regularization parameter in non-linear problems. Vogel (1985) used the GCV criterion in a Newton-type solution to a simplified inverse scattering problem; Amato & Hughes (1991) used the GCV criterion in the inversion of the standard Fredholm integral equation of the first kind, which was non-linear because of their choice of an entropy measure as the regularization term; and Smith & Bowers (1993) investigated both the GCV and L-curve criteria in a quasi-Newton/trust-region inversion of the 1-D diffusion equation for a spatially varying diffusion coefficient.

Recently, there have been applications of the GCV and L-curve criteria to the solution of some non-linear inverse problems encountered in geophysics. Haber (1997) and Haber & Oldenburg (2000) proposed the use of the GCV technique in conjunction with a damped Gauss–Newton step in the solution of non-linear inverse problems, and applied their algorithm to the 1-D inversion of magnetotelluric data, and the inversion of gravity data for the depth to the interface between two layers of contrasting densities. Li & Oldenburg (2003) used the GCV technique in their solution to the linear inverse problem of constructing a 3-D susceptibility model from magnetic data, and used the obtained value of the regularization parameter in the subsequent non-linear inversion when positivity was imposed on the susceptibility. Li & Oldenburg (1999) also used the L-curve technique and a damped Gauss–Newton step in their 3-D inversion of DC resistivity data. Finally, Walker (1999) investigated the use of the GCV method, in conjunction with a damped Gauss–Newton step, in the inversion of electromagnetic loop–loop data for a 1-D conductivity model of the Earth. One of Walker's observations was that this approach sometimes put excessive structure in the model at early iterations, which was then difficult and time-consuming to remove at later iterations.

In addition, Akaike's Bayesian information criterion has been applied by Uchida (1993) and Mitsuhata *et al.* (2002) to the 2-D inversion of magnetotelluric and controlled-source electromagnetic data. We do not include this approach in our current discussion, but refer the interested reader to the above publications and references therein.

Here the use of the GCV and L-curve techniques, along with the discrepancy principle, are compared and contrasted. As a typical non-linear problem, the simultaneous inversion of frequency-domain electromagnetic loop–loop data to recover both the electrical conductivity and magnetic susceptibility of a 1-D Earth is considered. The solution is iterative, with a linearized approximation to the full non-linear inverse problem being solved at each iteration. The GCV- and L-curve-based techniques are applied to the solution of the linearized problems. First, the salient features of the iterative procedure are summarized, as well as the GCV and L-curve criteria, then their performance illustrated with 1-D inversions of both 1- and 3-D synthetic data sets, and a field data set.

2 THEORY

2.1 Iterative, linearized solution of the non-linear problem

The non-linear optimization problem consists of minimizing the objective function (from eqs 1–3):

$$\Phi(\mathbf{m}) = \left\|\mathbf{W}_d[\mathbf{d}^{obs} - \mathbf{d}(\mathbf{m})]\right\|^2 + \beta \sum_{k=1}^{2} \left\|\mathbf{W}_k \left(\mathbf{m} - \mathbf{m}_k^{ref}\right)\right\|^2, \qquad (6)$$

where $\mathbf{W}_1 = \sqrt{\alpha_s}\mathbf{W}_s$ and $\mathbf{W}_2 = \sqrt{\alpha_z}\mathbf{W}_z$, and $\mathbf{m}_1^{ref} = \mathbf{m}_s^{ref}$ and $\mathbf{m}_2^{ref} = \mathbf{m}_z^{ref}$. The usual procedure is followed (see, for example, Gill *et al.* 1981; Dennis & Schnabel 1996). Let $\mathbf{m}^{n-1}$ be the current model. A perturbation $\delta\mathbf{m}$ that will reduce $\Phi(\mathbf{m})$, and a value of the regularization parameter β that provides an optimal trade-off between misfit and model structure are sought. Assume for the moment that a suitable value, β^n, has been found for the regularization parameter for the current iteration. Consider also the linear Taylor series approximation of the data for the model to be found at this iteration:

$$\mathbf{d}^n \approx \mathbf{d}^{n-1} + \mathbf{J}^{n-1}\delta\mathbf{m}, \qquad (7)$$

where $\mathbf{d}^n = \mathbf{d}(\mathbf{m}^n)$, $\delta\mathbf{m} = \mathbf{m}^n - \mathbf{m}^{n-1}$, and $\mathbf{J}^{n-1}$ is the Jacobian matrix of sensitivities:

$$J_{ij}^{n-1} = \left.\frac{\partial d_i}{\partial m_j}\right|_{\mathbf{m}^{n-1}}. \qquad (8)$$

Substituting eq. (7) into the objective function in eq. (6) gives:

$$\Phi[\mathbf{m}^n] \approx \left\|\mathbf{W}_d(\mathbf{d}^{obs} - \mathbf{d}^{n-1} - \mathbf{J}^{n-1}\delta\mathbf{m})\right\|^2 + \beta^n \sum_{k=1}^{2} \left\|\mathbf{W}_k \left[\mathbf{m}^n - \mathbf{m}_k^{ref}\right]\right\|^2. \qquad (9)$$

Differentiating this expression with respect to the elements of $\delta\mathbf{m}$ and equating the resulting M derivatives to zero yields the following linear system of equations to solve:

$$\left(\mathbf{J}^{n-1T}\mathbf{W}_d^T\mathbf{W}_d\mathbf{J}^{n-1} + \beta^n \sum_{k=1}^{2} \mathbf{W}_k^T\mathbf{W}_k\right)\delta\mathbf{m} = \mathbf{J}^{n-1T}\mathbf{W}_d^T\mathbf{W}_d(\mathbf{d}^{obs} - \mathbf{d}^{n-1}) + \beta^n \sum_{k=1}^{2} \mathbf{W}_k^T\mathbf{W}_k \left(\mathbf{m}_k^{ref} - \mathbf{m}^{n-1}\right). \qquad (10)$$

The solution to eq. (10) is also equivalent to the least-squares solution of

$$\begin{pmatrix} \mathbf{W}_d\mathbf{J}^{n-1} \\ \sqrt{\beta^n}\mathbf{W}_1 \\ \sqrt{\beta^n}\mathbf{W}_2 \end{pmatrix}\delta\mathbf{m} = \begin{pmatrix} \mathbf{W}_d(\mathbf{d}^{obs} - \mathbf{d}^{n-1}) \\ \sqrt{\beta^n}\mathbf{W}_1 \left(\mathbf{m}_1^{ref} - \mathbf{m}^{n-1}\right) \\ \sqrt{\beta^n}\mathbf{W}_2 \left(\mathbf{m}_2^{ref} - \mathbf{m}^{n-1}\right) \end{pmatrix}. \qquad (11)$$

For the examples presented in this paper, eq. (10) is used when applying the GCV criterion, and eq. (11) when applying the L-curve criterion and the discrepancy prinicple.

Once the model update $\delta\mathbf{m}$ has been determined from the solution of either eq. (10) or eq. (11), the new model $\mathbf{m}^n$ is given by

$$\mathbf{m}^n = \mathbf{m}^{n-1} + \lambda\delta\mathbf{m}. \tag{12}$$

The step length λ is successively reduced, if required, by factors of 2 from its initial value of 1 to ensure that the objective function is decreased, that is, so that

$$\phi_{\rm d}^n(\lambda) + \beta^n\phi_{\rm m}^n(\lambda) < \phi_{\rm d}^{n-1} + \beta^n\phi_{\rm m}^{n-1}, \tag{13}$$

where $\phi_{\rm d}^n = \phi_{\rm d}[\mathbf{m}^n]$. This is the damped Gauss–Newton method.

Finally, two of the termination criteria of Gill *et al.* (1981), namely:

$$\Phi^{n-1} - \Phi^n < \tau(1 + \Phi^n), \tag{14}$$

$$\|\mathbf{m}^{n-1} - \mathbf{m}^n\| < \sqrt{\tau}(1 + \|\mathbf{m}^n\|), \tag{15}$$

are used to determine when the iterative procedure has reached convergence, where $\Phi^n = \Phi[\mathbf{m}^n]$, and τ is specified by the user (typically 0.01).

2.2 Choosing the regularization parameter: the discrepancy principle

The discrepancy principle is the method that has generally been used for choosing the regularization parameter in underdetermined inverse problems. Hansen (1997) gives a thorough discussion of it in the context of linear problems. It was used by Constable *et al.* (1987) in their Occam's inversion and by Smith & Booker (1988) for the inversion of magnetotelluric data for 1-D models of the electrical conductivity distribution within the Earth.

For the assumed uncorrelated Gaussian noise of zero mean and standard deviation σ_i, the measure of misfit given in eq. (2) is a χ^2 random variable. If the values used for the standard deviations when calculating $\phi_{\rm d}$ are known, the expectation of the misfit is equal to the number of observations, N, with a standard deviation of $\sqrt{2N}$. The value of the regularization parameter chosen according to the discrepancy principle is therefore one that results in a model for which $\phi_{\rm d} \approx N$.

For an iterative, linearized solution to a non-linear inverse problem, the above argument can be used to choose the regularization parameter at each iteration, with the misfit calculated using either the exact forward modelling, as done by Constable *et al.* (1987), or the linearized approximation given in eq. (7), as done by Smith & Booker (1988). The implementation of the discrepancy principle used for the examples presented in this paper uses the exact forward modelling. The target misfit of approximately N is not usually attainable at early iterations, in which case the value of the regularization parameter is typically chosen to be the one that gives the minimum misfit. However, if the regularization parameter is too small at early iterations, excessive structure can build-up in the model, which can then require a not insignificant number of additional iterations to remove later in the inversion. It is actually more efficient if the starting value of the regularization parameter is fairly large and restrictions are placed on its greatest allowed decrease, thus enforcing a slow but steady introduction of structure into the model. For the examples presented here, the discrepancy principle is therefore combined with an imposed cooling-schedule-type behaviour, with the value of β at the nth iteration chosen to be (Farquharson & Oldenburg 1993):

$$\beta^n = \max(c\beta^{n-1}, \beta^\star), \tag{16}$$

where $0.01 \leq c \leq 0.5$, and $\beta^\star$ is the value of the regularization parameter for which $\phi_{\rm d} \approx N$, or for which $\phi_{\rm d}$ is a minimum.

However, if the standard deviations of the noise in the observations are not known, as is often the case in reality, then neither is the expectation of the misfit. Consequently, there is no target misfit for which to aim in the inversion.

2.3 The GCV criterion

This criterion for defining an appropriate value of the regularization parameter is based on the following argument, the so-called 'leaving-out-one' lemma (Wahba 1990). Consider the linear inverse problem of finding the model, $\mathbf{m}$, which minimizes:

$$\sum_{i=1}^{N}\left[d_i^{\rm obs} - d_i(\mathbf{m})\right]^2 + \beta\|\mathbf{m}\|^2, \tag{17}$$

where $\mathbf{d}(\mathbf{m}) = \mathbf{L}\,\mathbf{m}$, and $\mathbf{L}$ is a matrix independent of $\mathbf{m}$. Assume also that the noise in every observation has the same standard deviation, that is, $\sigma_i = \sigma_0$ $(i = 1, \ldots, N)$. Consider inverting all but the kth observation using a trial value, $\hat{\beta}$, of the regularization parameter, that is, find the model $\mathbf{m}^k$ which minimizes:

$$\sum_{\substack{i=1\\ i\neq k}}^{N}\left[d_i^{\rm obs} - d_i(\mathbf{m})\right]^2 + \hat{\beta}\|\mathbf{m}\|^2. \tag{18}$$

For $\hat{\beta}$ to be considered a suitable value for the regularization parameter, the kth forward-modelled datum, $d_k[\mathbf{m}^k]$, should be close to the omitted observation, $d_k^{\rm obs}$. If this procedure is repeated leaving out each observation in turn, and all the forward-modelled data $d_k[\mathbf{m}^k]$ are close to their respective observations, $\hat{\beta}$ would be considered a suitable value of the regularization parameter for the whole set of observations. The most suitable value can therefore be defined as the one that minimizes the function:

$$V_0(\beta) = \sum_{k=1}^{N}\left\{d_k^{\rm obs} - d_k[\mathbf{m}^k]\right\}^2. \tag{19}$$

This is the ordinary cross-validation function. It can also be expressed in a more efficiently evaluated form that does not require explicit solution of the inverse problem for each omitted observation (Wahba 1990):

$$V_0(\beta) = \sum_{i=1}^{N}\frac{\left[d_i^{\rm obs} - d_i(\mathbf{m}_\beta)\right]^2}{[1 - A_{ii}(\beta)]^2}, \tag{20}$$

where $\mathbf{m}_\beta = (\mathbf{L}^T\,\mathbf{L} + \beta\,\mathbf{I})^{-1}\,\mathbf{L}^T\mathbf{d}^{\rm obs}$ is the solution of the inverse problem for the particular value of β, and A_{ii} is the ith element on the diagonal of the matrix $\mathbf{A}(\beta) = \mathbf{L}(\mathbf{L}^T\mathbf{L} + \beta\,\mathbf{I})^{-1}\mathbf{L}^T$.

The ordinary cross-validation function given in eqs (19) and (20) is not invariant under an orthogonal transformation of $\mathbf{L}$ and $\mathbf{d}^{\rm obs}$. The value of β that minimizes V_0 for the transformed problem will therefore not be the same as the one that minimizes V_0 for the original problem, leading to different inversion results for the two related problems. This should not be the case. A modification of eq. (20) gives the generalized cross-validation (GCV) function (Wahba 1990):

$$V(\beta) = \frac{\|\mathbf{d}^{\rm obs} - \mathbf{d}(\mathbf{m}_\beta)\|^2}{\{\mathrm{trace}[\mathbf{I} - \mathbf{A}(\beta)]\}^2}, \tag{21}$$

which is invariant under an orthogonal transformation.

For a non-linear problem solved using an iterative procedure, the GCV-based method described above can be applied to the linearized problem at each iteration. One would anticipate (assuming such a procedure converges) that at the final iterations for which the changes in the model are small and thus the linearized approximation

is an adequate description of the problem, the GCV criterion would exhibit the same success that it does for purely linear problems. This has been demonstrated by Haber & Oldenburg (2000) who showed for two example non-linear problems that, at convergence, the obtained value of the regularization parameter was a good estimate of what was expected given the noise in their synthetic data sets. In addition, they found that the estimates of the regularization parameter at the early iterations were close to its ultimate value, implying that the GCV-based method, and the leave-out-one lemma on which it is based, can distinguish between the Gaussian noise in a set of observations and the errors arising from the linear approximation.

For the problem considered here, the GCV function for the nth iteration is given by (from eq. 10, by analogy with eq. 21):

$$V^n(\beta) = \frac{\left\| \mathbf{W}_\mathrm{d}\hat{\mathbf{d}} - \mathbf{W}_\mathrm{d}\mathbf{J}^{n-1}\mathbf{M}^{-1}\left(\mathbf{J}^{n-1T}\mathbf{W}_\mathrm{d}^T\mathbf{W}_\mathrm{d}\hat{\mathbf{d}} + \mathbf{r}\right)\right\|^2}{\left[\mathrm{trace}\left(\mathbf{I} - \mathbf{W}_\mathrm{d}\mathbf{J}^{n-1}\mathbf{M}^{-1}\mathbf{J}^{n-1T}\mathbf{W}_\mathrm{d}^T\right)\right]^2}, \qquad (22)$$

where:

$$\mathbf{M}(\beta) = \left(\mathbf{J}^{n-1T}\mathbf{W}_\mathrm{d}^T\mathbf{W}_\mathrm{d}\mathbf{J}^{n-1} + \beta\sum_{i=1}^{2}\mathbf{W}_i^T\mathbf{W}_i\right), \qquad (23)$$

$$\mathbf{r} = \beta\sum_{i=1}^{2}\mathbf{W}_i^T\mathbf{W}_i\left(\mathbf{m}_i^{\mathrm{ref}} - \mathbf{m}^{n-1}\right), \qquad (24)$$

and $\hat{\mathbf{d}} = \mathbf{d}^{\mathrm{obs}} - \mathbf{d}^{n-1}$. The data weighting matrix $\mathbf{W}_\mathrm{d}$ contains the estimates of the relative amounts of noise in the observations: $\mathbf{W}_\mathrm{d} = \mathrm{diag}\{1/\tilde{\sigma}_1, \ldots, 1/\tilde{\sigma}_N\}$.

Although the GCV-based estimates of the regularization parameter at the early iterations can be close to what its final value will be, it has been shown by Walker (1999) that using these estimates can cause too much structure to appear too early in the model. The GCV criterion is therefore combined with a cooling-schedule-type behaviour, just as for the discrepancy principle, choosing the value of β at the nth iteration to be:

$$\beta^n = \max(c\beta^{n-1}, \beta^\star), \qquad (25)$$

where $0.01 \leq c \leq 0.5$, and $\beta^\star$ is the minimizer of the GCV function given in eq. (22).

2.4 The L-curve criterion

Consider again the linear inverse problem introduced in Section 2.3. If solutions are computed for all values of the regularization parameter β, the graph, using log–log axes, of the misfit $\|\mathbf{d}^{\mathrm{obs}} - \mathbf{L}\,\mathbf{m}(\beta)\|^2$ versus the model norm $\|\mathbf{m}(\beta)\|^2$ tends to have a characteristic 'L' shape (Hansen 1997). At the corner of the L-curve, a change in the value β results in changes of roughly equal significance in both the misfit and model norm. In contrast, on either of the two branches of the L-curve, a change in β results in either a small decrease in the misfit and a large increase in the model norm (for β greater than that at the corner), or a small increase in the model norm and a large decrease in the misfit (for β smaller than that at the corner). The corner of the L-curve can therefore be taken as indicating the value of the regularization parameter that gives the best balance between the two opposing components of the objective function.

For a non-linear problem, a value for the regularization parameter can be selected by applying the above ideas to the linearized inverse problem at each iteration (Li & Oldenburg 1999). Again, the expectation is that at convergence, when the linearized approximation is an adequate description of the full non-linear problem, the value of the regularization parameter chosen for the linearized problem is also suitable for the non-linear problem. At each iteration, the misfit, $\phi_\mathrm{d}^{\mathrm{lin}}$, computed using the linearized approximation of the forward modelling given in eq. (7) is plotted against ϕ_m. The data weighting matrix $\mathbf{W}_\mathrm{d}$ is comprised of the relative uncertainties in the observations, just as for the GCV-based criterion. The curvature of the L-curve is computed using the formula (Hansen 1997):

$$C(\beta) = \frac{\zeta'\eta'' - \zeta''\eta'}{[(\zeta')^2 + (\eta')^2]^{3/2}}, \qquad (26)$$

where $\zeta = \log\phi_\mathrm{d}^{\mathrm{lin}}$, and $\eta = \log\phi_\mathrm{m}$. The prime denotes differentiation with respect to $\log\beta$. Just as for the implementations of the discrepancy principle and the GCV-based approach, the regularization parameter is chosen at the nth iteration according to the expression:

$$\beta^n = \max(c\beta^{n-1}, \beta^\star), \qquad (27)$$

where $0.01 \leq c \leq 0.5$, and $\beta^\star$ is the maximizer of the curvature of the L-curve given by eq. (26).

3 EXAMPLES

3.1 A synthetic 1-D example

The abilities of the GCV- and L-curve-based methods for choosing the regularization parameter are first illustrated with a synthetic example: the simultaneous inversion of electromagnetic loop–loop data for 1-D models of both electrical conductivity and magnetic susceptibility (see also Farquharson *et al.* 2003). The three-layered Earth model for which the data were generated is shown by the dashed lines in Figs 1(b) and (c). A 'max–min'-type survey configuration was considered: a vertical magnetic dipole transmitter 1 m above the surface of the model, and observations of the vertical component of the secondary (i.e. total minus free-space) magnetic field, also 1 m above the surface, at a distance of 50 m from the transmitter. The real and imaginary parts of the secondary field (normalized by the free-space field, and expressed as a percentage) were computed for ten frequencies ranging from 110 Hz to 56 kHz. Gaussian noise of zero mean and standard deviation equal to 5 per cent of the magnitude of an observation, or 0.001 per cent, whichever was larger, was added to give the data set to be inverted. The χ^2 measure of the actual amount of noise introduced in this example was 21.8. The data are shown by the error bars (which are equal in size to the standard deviations of the added noise) in Fig. 1(a).

The synthetic data set described above was inverted for both conductivity and susceptibility using the GCV- and L-curve-based methods to choose the regularization parameter. The Earth models comprised 50 layers of increasing thickness, with uniform conductivity and susceptibility in each layer. The parameters sought in the inversions were the logarithms of the layer conductivities and the layer susceptibilities. To ensure positivity of the recovered susceptibility, a logarithmic barrier term was included in the objective function (see, for example, Wright 1997; Farquharson *et al.* 2003; Li & Oldenburg 2003). This issue will not be pursued further here, except to note that both the GCV and L-curve methods were successful even with this additional non-linear term in the objective function.

For all the following inversions, the coefficients α_s and α_z for the conductivity half of the model were equal to 0.001 and 1, respectively, and those for the susceptibility half were equal to 0.05 and 1. The reference conductivity was a homogeneous half-space of 0.001 S m^{-1}, and the reference susceptibility was a half-space of 0 SI units. The starting model was the best-fitting half-space, and

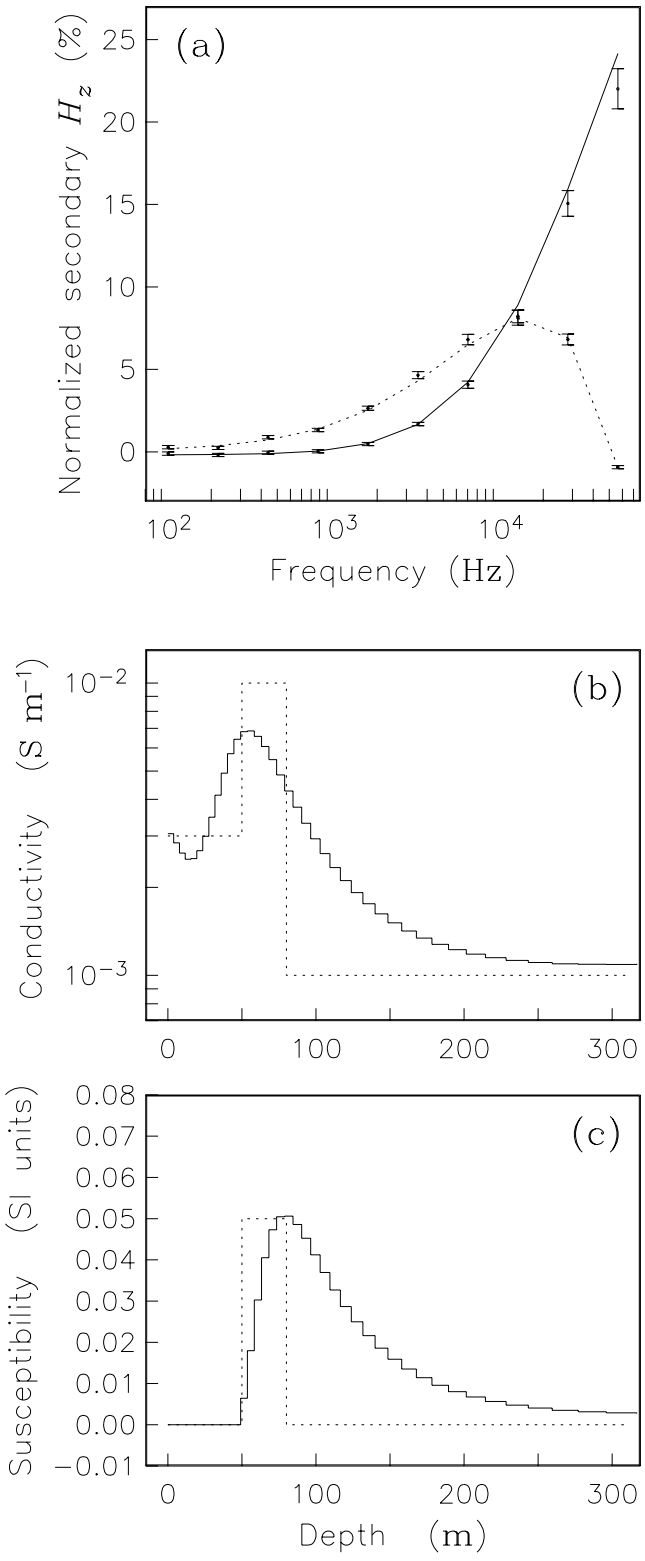

Figure 1. (a) The synthetic data set (error bars) for the first example inversion and the forward-modelled data (solid line, inphase; dashed line, quadrature) for the model produced by the inversion using the GCV-based method for choosing the regularization parameter. (b) The conductivities in the model constructed by the inversion, and (c) the susceptibilities in the constructed model. The dashed lines in (b) and (c) indicate the model for which the data were generated.

the initial value of the regularization parameter, β^0, was equal to $N/\phi^\dagger{}_m$, where $\phi^\dagger{}_m$ was the model norm computed for a two-layer model of typical conductivities and susceptibilities (the top ten layers of 0.02 S m^{-1} and 0.02 SI units; the remainder of 0.01 S m^{-1} and 0 SI units). The constant c in eqs (25) and (27) was taken to be 0.5.

The conductivity and susceptibility models produced using the GCV-based method are shown in Figs 1(b) and (c). Their smeared-out character is due to the smoothing regularization that was used, in particular the choice of an l_2 norm, and the loss of resolution with depth. The corresponding forward-modelled data are shown in Fig. 1(a). It is clear that they reproduce the observations well. The value of misfit is equal to 19.3, which is slightly less than the amount of noise (=21.8) introduced into the data set.

The GCV function at each iteration in the inversion is shown in Fig. 2. The minimum in the GCV function at each iteration is listed

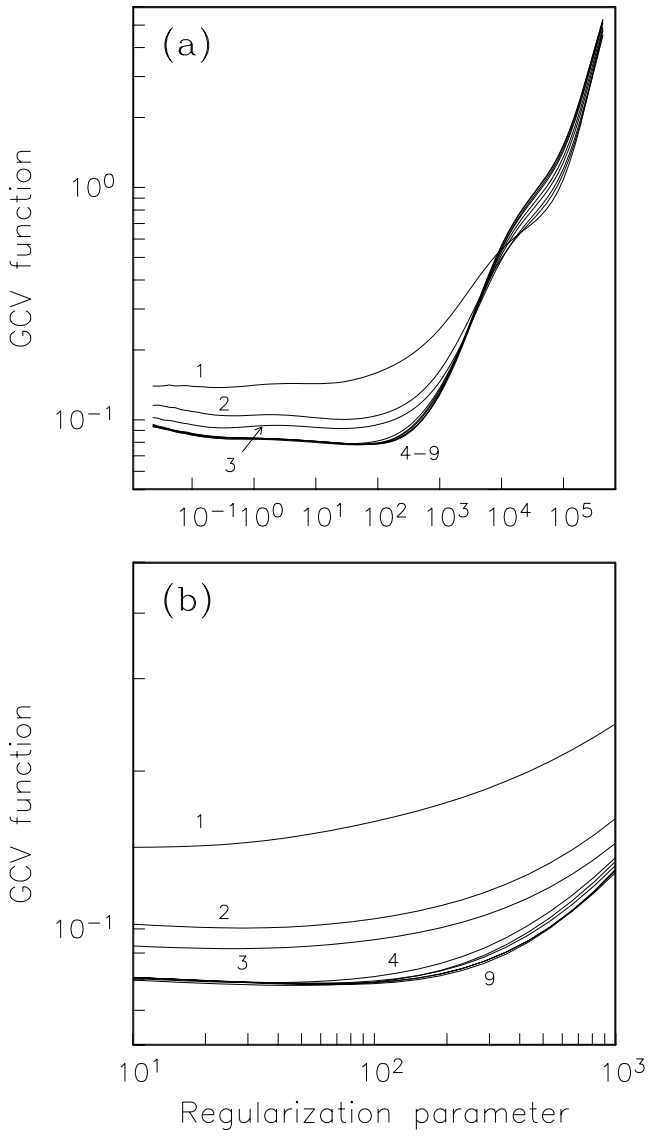

Figure 2. The variation of the GCV function with the regularization parameter at each iteration of the inversion of the synthetic data set shown in Fig. 1. Panel (a) shows the variation for a range of values of the regularization parameter. Panel (b) emphasizes the variation around the value (60.0) to which the inversion converged. The iteration to which each curve corresponds is indicated by the labels.

Table 1. The values of the regularization parameter at the minimum of the GCV function ($\beta^\star_{\mathrm{GCV}}$) and the values used ($\beta^n_{\mathrm{GCV}}$) at each iteration ($n$) of the inversion of the first synthetic data set using the GCV-based method. Also the values corresponding to the point of maximum curvature of the L-curve ($\beta^\star_{\mathrm{L}}$) and the values used (β^n_{L}) during that inversion.

n	$\beta^\star_{\mathrm{GCV}}$	β^n_{GCV}	$\beta^\star_{\mathrm{L}}$	β^n_{L}
1	0.23	228.0	56.1	228.0
2	28.4	114.0	37.3	114.0
3	28.5	57.1	40.0	57.1
4	34.7	34.7	32.5	32.5
5	46.2	46.2	36.8	36.8
6	53.2	53.2	40.0	40.0
7	65.0	65.0	48.4	48.4
8	60.0	60.0	45.7	45.7
9	60.0	60.0	36.6	36.6

in Table 1. The values of the regularization parameter used at each iteration according to eq. (25) with $c = 0.5$ are also listed in Table 1. It can be seen that the restriction on the decrease in β came into effect at the first three iterations. It can also be seen from Table 1 that the minimizer of the GCV function at all iterations except the first is a fair estimate of the value of the regularization parameter at convergence and that it becomes successively closer as the iterations proceed.

The models produced by the inversion of the first synthetic data set using the L-curve-based method are shown in Figs 3(b) and (c). Just as for the results of the inversion using the GCV-based method, the constructed models are in concordance with the models from which the data were generated. The forward-modelled data for the constructed models are shown in Fig. 3(a). Their misfit is equal to 18.9, which is less than the amount of noise added to the data (=21.8), and slightly less than the misfit at convergence of the GCV-based inversion (=19.3).

The L-curves for the iterations in the inversion of the first data set are shown in Fig. 4(a). The values of the (linearized) misfit, the model norm and the curvature of the L-curves as functions of the regularization parameter are shown in Figs 4(b)–(d). The values of the regularization parameter at the point of maximum curvature on the L-curves are listed in Table 1, along with the actual values chosen at each iteration according to eq. (27) with $c = 0.5$. From Fig. 4 it can be seen that the L-curve, and hence its curvature, changes substantially between the early iterations. However, the L-curves for the later iterations coalesce into a single curve as the iterative procedure converges to the solution of the non-linear inverse problem.

The final values of the regularization parameter for the inversions using the GCV and L-curve criteria (60.0 versus 36.6—see Table 1) are different despite the closeness of the corresponding values of misfit (19.3 and 21.8). This reflects the slow variation of the inversion results with the regularization parameter, as illustrated by the wide, indistinct minimum of the GCV function in Fig. 2(b) and the wide maximum in the curvature of the L-curve shown in Fig. 4(b).

As a final note for this example data set, its inversion using the discrepancy principle did converge to the target misfit of 20.0 (for which $\beta = 95.6$) and produced conductivity and susceptibility models essentially the same as those produced using the GCV and L-curve criteria.

3.2 Different noise realizations

The results of repeating the above inversions for eight other realizations of the noise used to make the synthetic data are briefly given

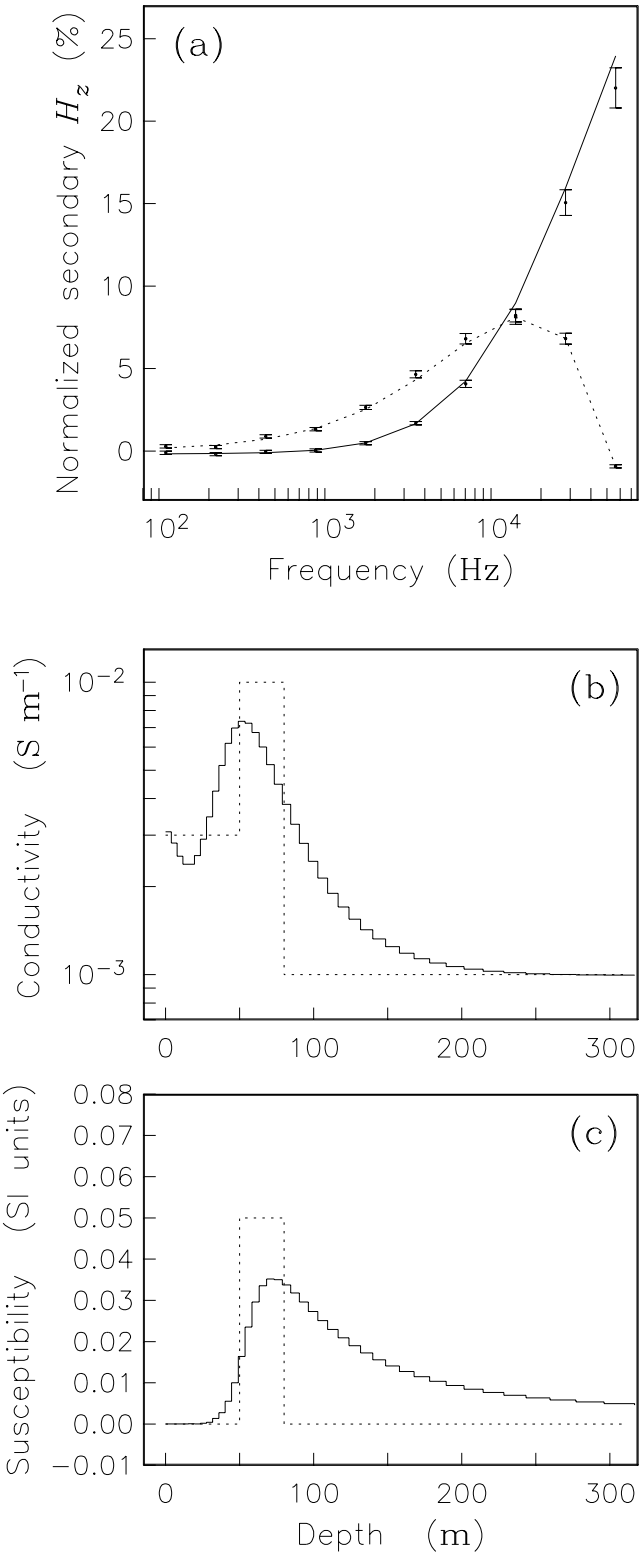

Figure 3. (a) The forward-modelled data (solid line, inphase; dashed line, quadrature) for the result of inverting the first synthetic data set (error bars) using the L-curve-based method for choosing the regularization parameter. (b) The constructed conductivity model, and (c) the constructed susceptibility model. The dashed lines in (b) and (c) indicate the model for which the data were generated.

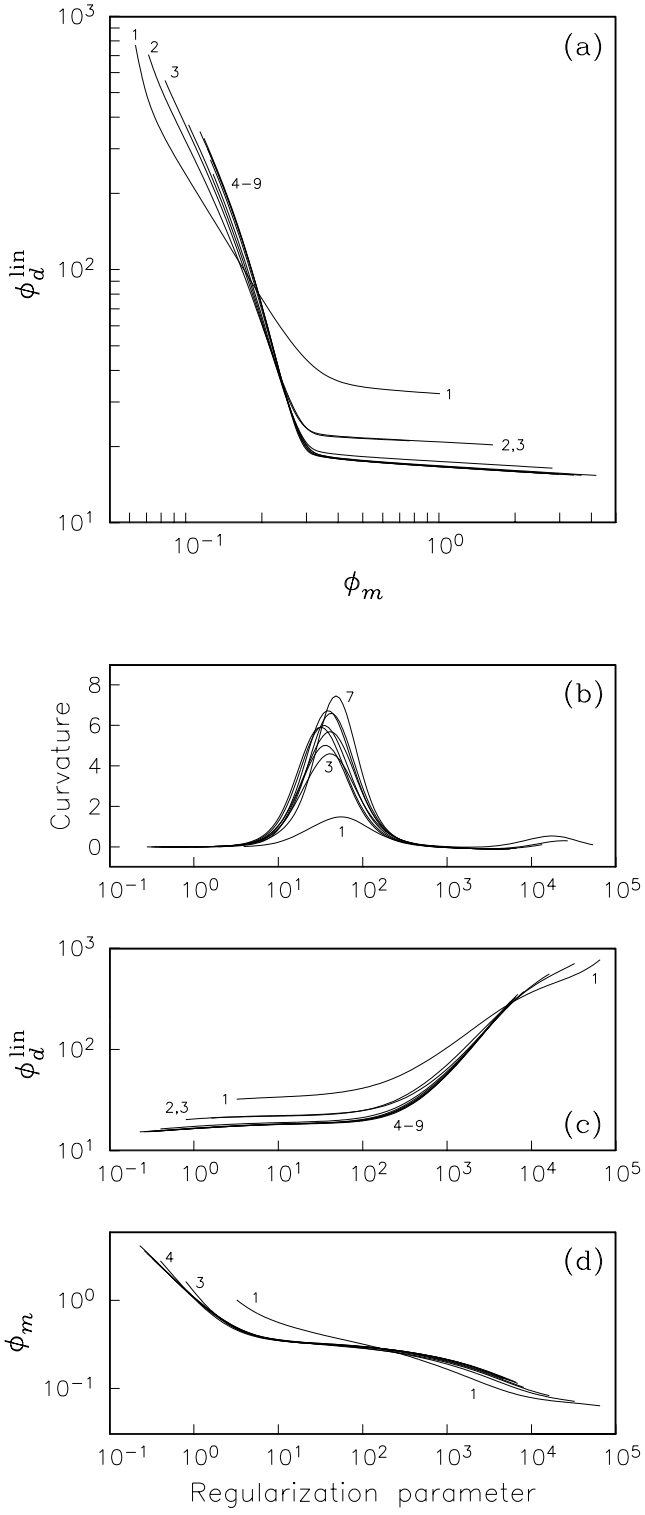

Figure 4. (a) The L-curves of the linearized misfit plotted as a function of the model norm at each iteration in the inversion of the first synthetic data set, the results of which are shown in Fig. 3. The labels indicate the iteration to which each of the more distinct L-curves corresponds. (b) The curvature of the L-curves as a function of the regularization parameter. (c) The (linearized) misfit as a function of the regularization parameter, and (d) the model norm at each iteration in the inversion. The labels indicate to which iteration the more distinct lines correspond.

Table 2. The final value of the regularization parameter, β, and the corresponding value of the misfit, $\phi_{\rm d}$, for the inversions of the synthetic data sets described in Section 2.4, which differ only in the realization of the added noise. The superscripts and subscripts 'GCV' and 'L' indicate the results of the inversions using the GCV and L-curve criteria, respectively. The amount of noise added for each noise realization (NR) is denoted by $\phi_{\rm d}^{\star}$.

NR	$\phi_{\rm d}^{\star}$	$\beta_{\rm GCV}$	$\phi_{\rm d}^{\rm GCV}$	$\beta_{\rm L}$	$\phi_{\rm d}^{\rm L}$
1	22.7	86.8	22.0	35.0	21.4
2	12.7	63.0	10.4	15.9	10.0
3	23.7	100.0	19.4	41.1	18.8
4	21.7	97.0	13.8	24.3	13.6
5	19.3	83.2	13.5	23.9	13.0
6	15.5	0.48	12.8	42.3	14.6
7	24.0	35.3	17.0	32.8	16.9
8	12.6	72.2	10.9	27.9	10.6

here. The amount of noise added in each case is given in Table 2. Each data set was inverted using both the GCV and L-curve criteria. The values of the inversion parameters were the same as those in the previous section. The final values of the regularization parameter and the corresponding values of the misfit are given in Table 2. The final conductivity and susceptibility models are shown in Figs 5 and 6.

As the numbers in Table 2 show, both the GCV and L-curve criteria gave final values of the misfit that were close to, although consistently less than, the amount of added noise. For the cases where the added noise was significantly different from the number of observations (realizations 2, 6 and 8), the results achieved by both the GCV and L-curve criteria are more appropriate than those that would have been achieved using the discrepancy principle. There was one major failure. For the sixth realization, the GCV criterion significantly underestimated the value of the regularization parameter resulting in extreme conductivity and susceptibility models (see Fig. 5). This is representative of our experience from applying the GCV criterion to many data sets, both synthetic and genuine: occasionally (although less often than the one-in-eight suggested by this example), it produces a value of the regularization parameter that overfits the observations to such an extent that the constructed model is unacceptable.

3.3 1-D inversions of synthetic 3-D data

The effectiveness of the GCV and L-curve criteria is further illustrated by presenting the results of performing 1-D inversions of the data at each observation location in a synthetic 3-D data set. The errors introduced by the 1-D approximation are correlated, and are not unlike the linearization errors at the early iterations of the 1-D example analysed in Section 3.1. It is therefore anticipated that the GCV and L-curve criteria will be able to discriminate between the Gaussian random noise and the correlated errors arising from the 1-D approximation, just as they were mostly successful at discriminating between random noise and linearization errors.

Synthetic data computed using the programme of Newman & Alumbaugh (1995) for the two-prism model shown in Fig. 7 is considered. This data set has also been used by Zhang & Oldenburg (1999). The shallower prism (prism 1) had a conductivity of 0.1 S m^{-1} and a susceptibility of 0.1 SI units, and the deeper prism (prism 2) had a conductivity of 0.5 S m^{-1} and a susceptibility of 0.2 SI units.

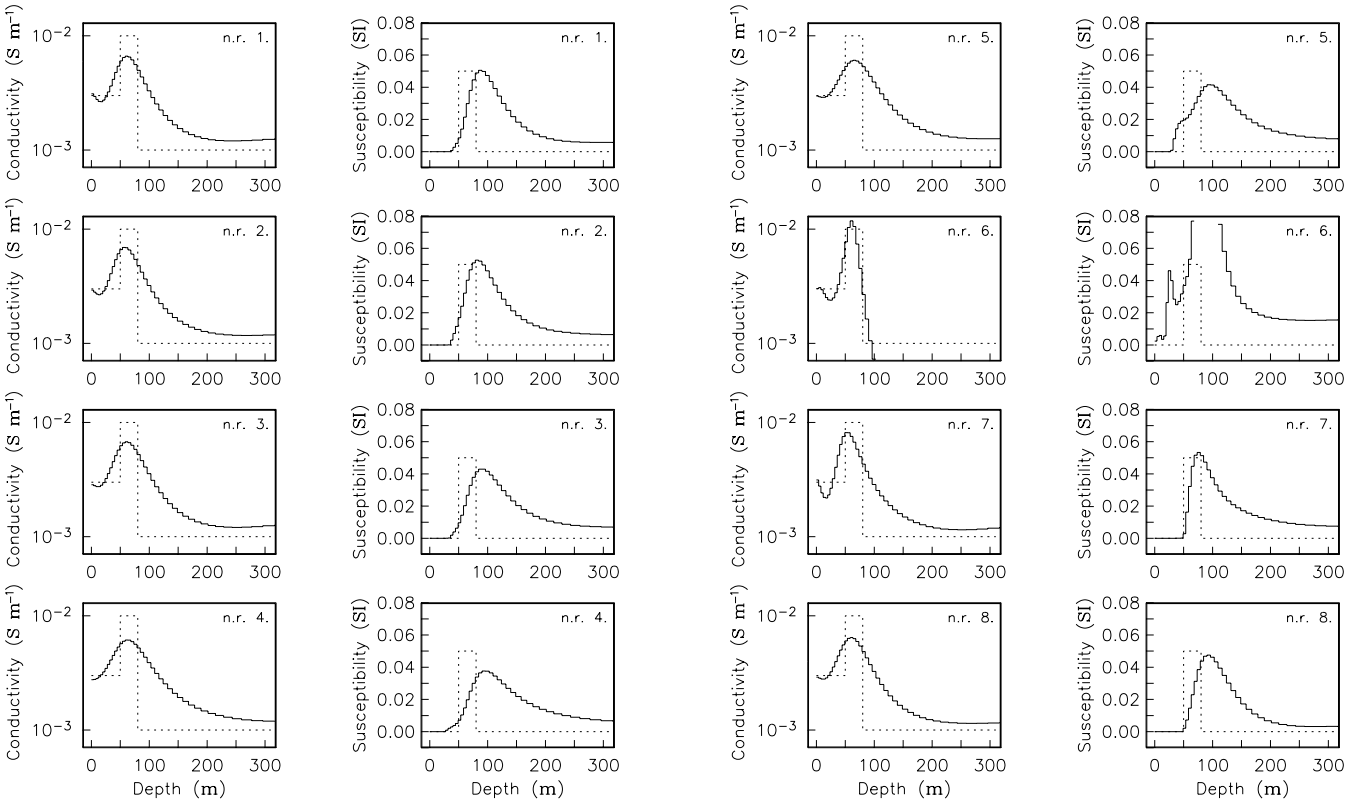

Figure 5. The conductivity and susceptibility models produced by the inversions, using the GCV criterion, of synthetic data sets differing only in the realization of the added noise (see Section 3.2).

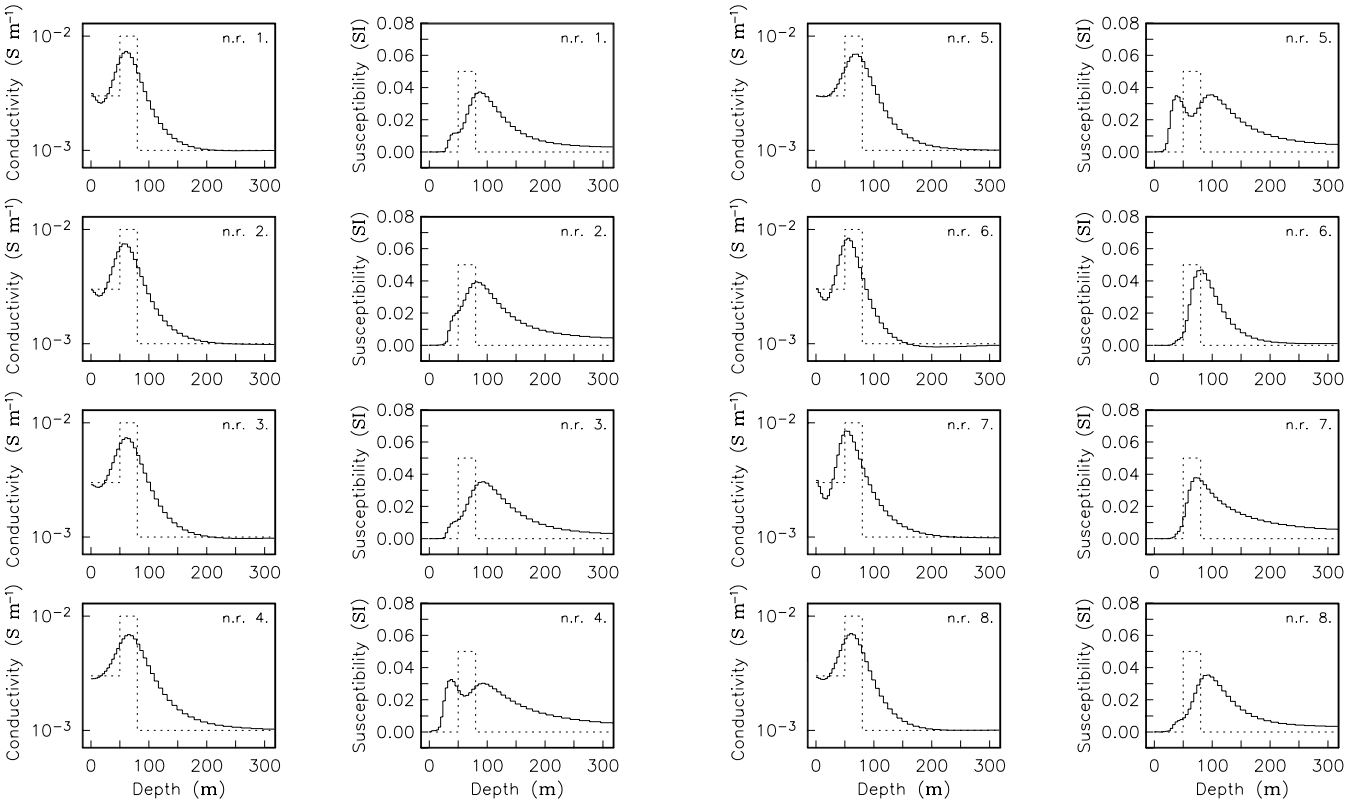

Figure 6. The conductivity and susceptibility models produced by the inversions, using the L-curve criterion, of synthetic data sets differing only in the realization of the added noise (see Section 3.2).

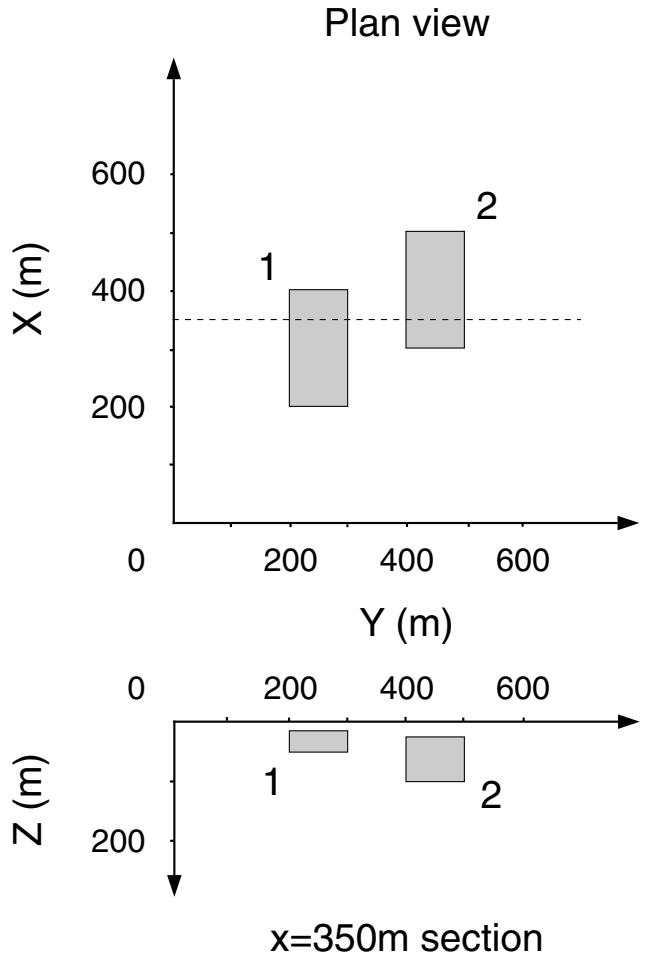

Figure 7. The two prism model of Section 3.3. Prism 1 has a conductivity of 0.1 S m^{-1} and a susceptibility of 0.1 SI units, and prism 2 has a conductivity of 0.5 S m^{-1} and a susceptibility of 0.2 SI units. The background is a non-susceptible half-space of conductivity 0.01 S m^{-1}. The line of data that was inverted is indicated by the dashed line.

The background was a homogeneous half-space of conductivity 0.01 S m^{-1} and zero susceptibility. The secondary magnetic field for an airborne-type configuration (a vertical magnetic dipole transmitter at a height of 30 m, and measurements of the vertical component of the magnetic field 10 m from the transmitter in the y-direction and also at a height of 30 m) was computed at 10 frequencies (110, 220, 440, 880, 1760, 3520, 7040, 14 080 and 56 320 Hz) for 29 locations of the transmitter–receiver pair along the line $x = 350$ m (see Fig. 7). Gaussian noise of standard deviation equal to 1 per cent of the magnitude of a datum, or 0.5 ppm, whichever was larger, was added to the computed field values to give the synthetic data set to be inverted. The synthetic data are indicated by the error bars in Fig. 8(a). The size of each error bar corresponds to the standard deviation of the noise added to that datum.

The 20 data (inphase and quadrature parts at the ten frequencies) at each location along the line $x = 350$ m were simultaneously inverted for 1-D models of both conductivity and susceptibility. There were 25 layers in the models. The values of the coefficients α_s and α_z in the conductivity part of the model norm were 0.001 and 1, respectively, and the values of the coefficients in the susceptibility part were 0.1 and 10, respectively. The reference conductivity model was a half-space of 0.01 S m^{-1}, and the reference susceptibility model was a non-susceptible half-space.

The results of using the L-curve criterion to determine the regularization parameter in the 1-D inversions at each location are shown in Fig. 8. Panels (c) and (d) show the images of the conductivity and susceptibility beneath the line of data created by stitching together the 1-D models for each location along the line. There are clearly artefacts in these images arising from inverting data containing 3-D effects for 1-D models, most notably the smearing outwards and downwards from the true locations of the prisms of the conductivity and to a lesser extend the susceptibility, the underestimation of the vertical extents of the prisms, and the overestimation of the resistivity directly below the prisms. However, these artefacts are not excessive, and the images are quite interpretable. It is therefore clear that the L-curve criterion has been successful in achieving an appropriate fit to the observations all along the line. At each end of the line, the attained misfits are close to their expected value of 20, and the corresponding parts of the conductivity and susceptibility images agree with the corresponding parts of the true model. Over the prisms, the values of the regularization parameter chosen by the L-curve criterion are sufficiently small that the data are fit reasonably well and there is meaningful structure in the model, but not so small that the artefacts in the models are extreme.

Fig. 9 shows the results of repeating the previous example, but using the discrepancy principle to choose the regularization parameter. At each end of the line, the target misfit of 20 was attained. However, over the prisms the 3-D effects in the data could not be fit by 1-D models, and so the discrepancy principle resorted to choosing the value of the regularization parameter that gave the smallest possible misfit. As can be seen from panel (b) in Figs 8 and 9, the final values of misfit over the two prisms are significantly less than those from the inversions using the L-curve criterion. The conductivity and susceptibility images in Figs 9(c) and (d) display the greater amount of structure associated with these smaller misfits. The artefacts in the conductivity and susceptibility images in Fig. 9 are not at all bad, but are definitely stronger than in the images produced using the L-curve criterion.

The results of using the GCV criterion for this example are shown in Fig. 10. The achieved misfits over the prisms are somewhat less than those for the L-curve criterion, and the amount of structure in the conductivity and susceptibility images is greater. However, like the L-curve criterion, the GCV criterion has done something sensible when confronted with having to fit data containing 3-D effects with a 1-D model, giving better results than those obtained using the discrepancy principle.

3.4 A field example

Finally, the results of inverting a field data set using both the GCV and L-curve criteria are presented. The observations are from an airborne survey for which measurements were made of the inphase and quadrature parts of the secondary magnetic field at three frequencies (1325, 4925 and 11 025 Hz) for the horizontal coplanar transmitter–receiver combination (vertical magnetic dipole source, and measurements of the vertical component of the magnetic field 5.1 m from the transmitter along the flight direction). The flight height varied slowly along the line, not deviating more than a couple of metres up or down from 30 m. The line of data considered here was composed of measurements made at 105 locations. As is typical for field data, no estimates of the measurement uncertainties were available. Uncertainties of 1 per cent of the value of a datum, or 1 ppm, whichever was greater, were therefore assigned to the observations. Panel (a) of Fig. 11 displays the observations

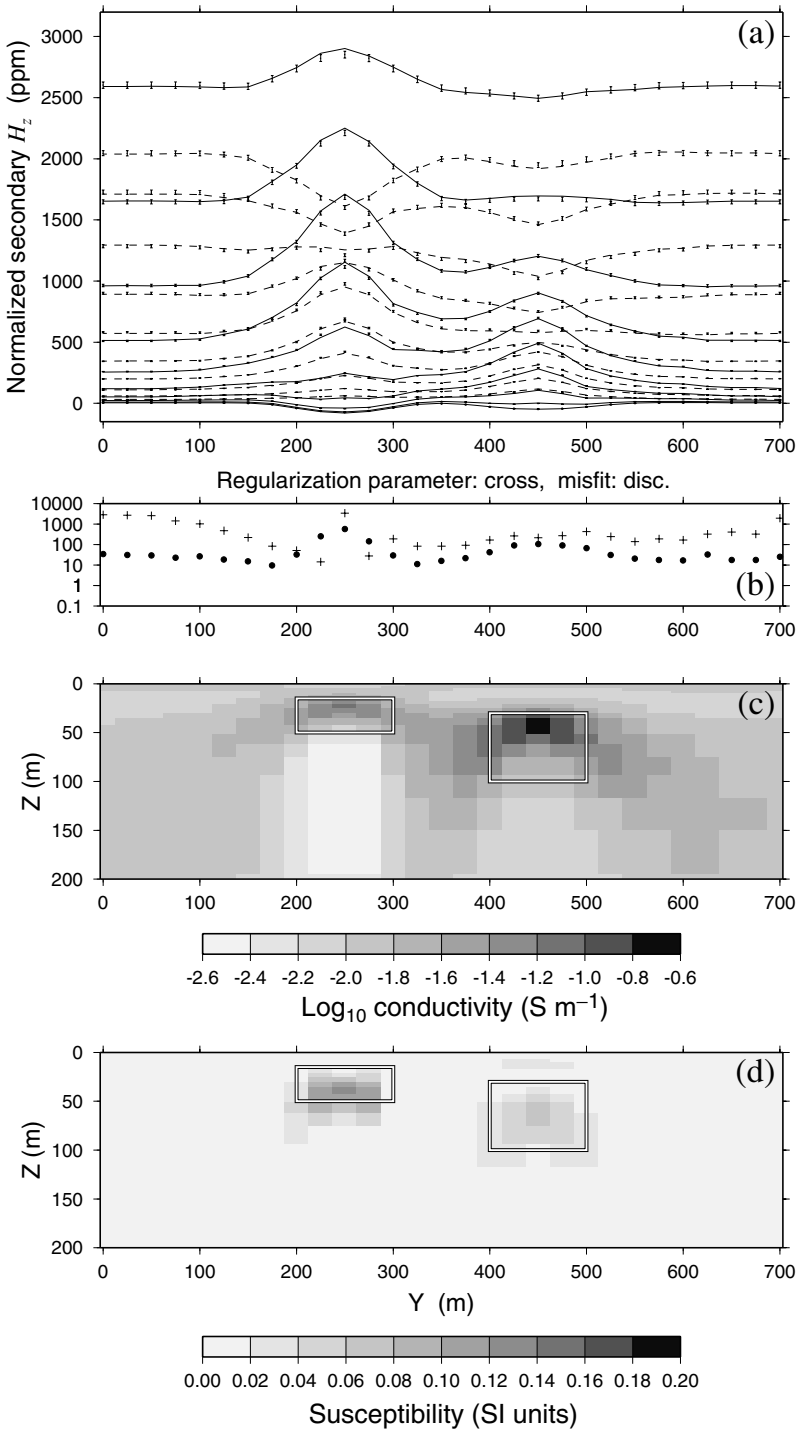

Figure 8. (a) The synthetic data set (error bars) created from the model shown in Fig. 7 and the forward-modelled data (solid, inphase; dashed, quadrature) from the 1-D models constructed using the L-curve criterion. (b) The regularization parameter and misfit for the final model for each sounding. (c) The image created by concatenating the constructed 1-D conductivity models for all soundings along the line. (d) The susceptibility image. The outlines in (c) and (d) indicate the location of the two prisms.

and the uncertainties. The purely ad hoc choice for the absolute size of the assigned uncertainties was based on the smooth variation of the data along the line, which suggested that the amount of noise in the observations was small. The use of an uncertainty equal to a percentage of the size of each datum was thought to be a realistic representation of how the amount of noise varied with frequency at any one location. It is in exactly this kind of real-life situation in which the noise in a data set is not well known that automatic means of determining the regularization parameter such as the GCV and L-curve criteria come into their own.

Figs 11 and 12 show the results of inverting the line of data using the L-curve and GCV criteria, respectively. Just as for the synthetic two-prism example described in Section 3.3, the data at each location were simultaneously inverted for 1-D conductivity and

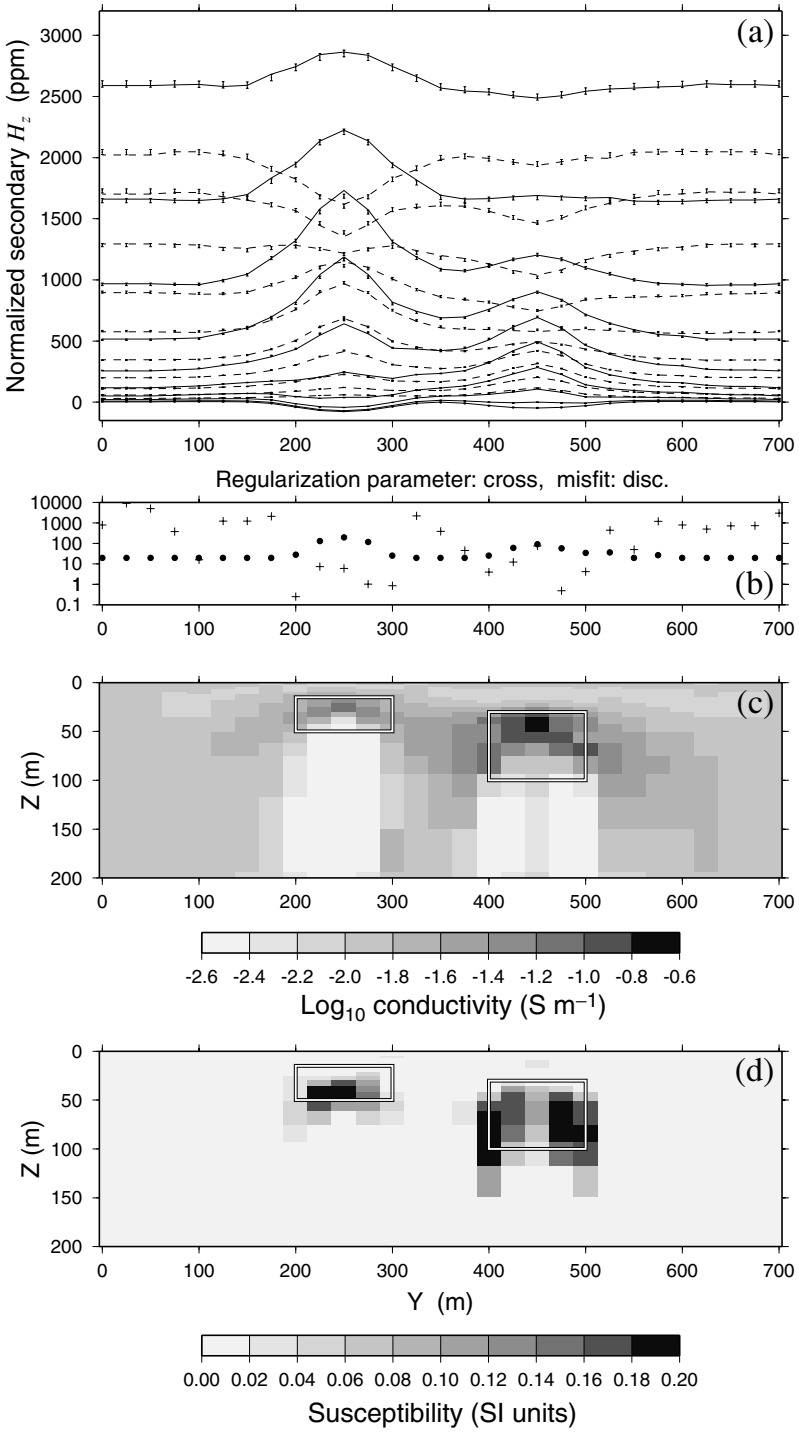

Figure 9. (a) The synthetic 3-D data set (error bars) and the forward-modelled data from the 1-D models constructed using the discrepancy principle. (b) The regularization parameter and misfit for the final model for each sounding. (c) The conductivity image. (d) The susceptibility image.

susceptibility models. These models were then stitched together to give the 2-D images of the subsurface beneath the flight line. Both inversions have revealed essentially the same variations in the thickness, conductivity and susceptibility of the overburden layer. The GCV-based inversion gave somewhat smaller values of misfit along most of the line, and slightly more structure in the conductivity and susceptibility images. However, both the GCV & L-curve criteria have resulted in models that give a misfit significantly smaller than six at each location, which is equal to the number of observations, and hence what the target misfit would have been for an inversion using the discrepancy principle. It is clear that the GCV and L-curve criteria are suggesting that the uncertainties we assigned in lieu of their observed values are actually too large for at least half of this survey line. If the discrepancy principle had been used with these assumed, incorrect uncertainties, the information content of the data along these stretches of the line would not have been fully extracted. Unfortunately, no information concerning the true structure of the subsurface beneath this survey line is presently available.

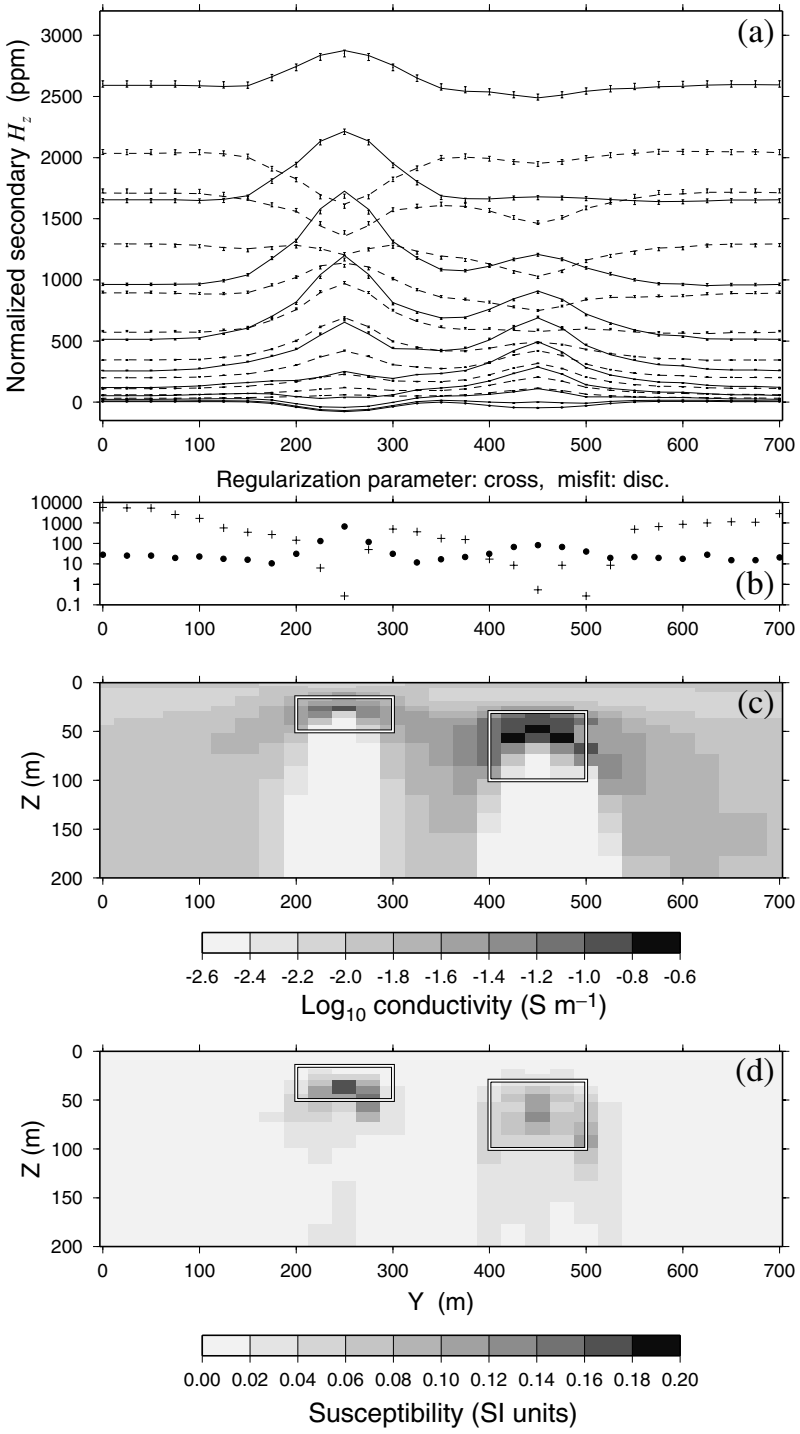

Figure 10. (a) The synthetic 3-D data set (error bars) and the forward-modelled data from the 1-D models constructed using the GCV criterion. (b) The regularization parameter and misfit for the final model for each sounding. (c) The conductivity image. (d) The susceptibility image.

4 DISCUSSIONS

We chose to investigate the performance of the GCV and L-curve criteria on a 1-D inverse problem so that computation time would not be an issue. (For example, the inversions using the discrepancy principle, the GCV criterion and the L-curve criteria of the line of data over the two-prism model in Section 3.3 took 8.5, 3.5 and 7.5 min, respectively, on a 500 MHz Pentium III computer.) However, there are a few computational practicalities that are relevant to larger problems, and so we briefly discuss them here. For the discrepancy principle, the line search along the misfit versus regularization parameter curve can be made efficient and accurate. Our implementation steps along this curve (in terms of the logarithm of the regularization parameter) starting at the value of the regularization parameter for the previous iteration until the target misfit, or the minimum in the misfit, is bracketed. The search then contracts to the target or minimum using a bisection or golden section search (see, for example, Press *et al.* 1992), respectively. However, the

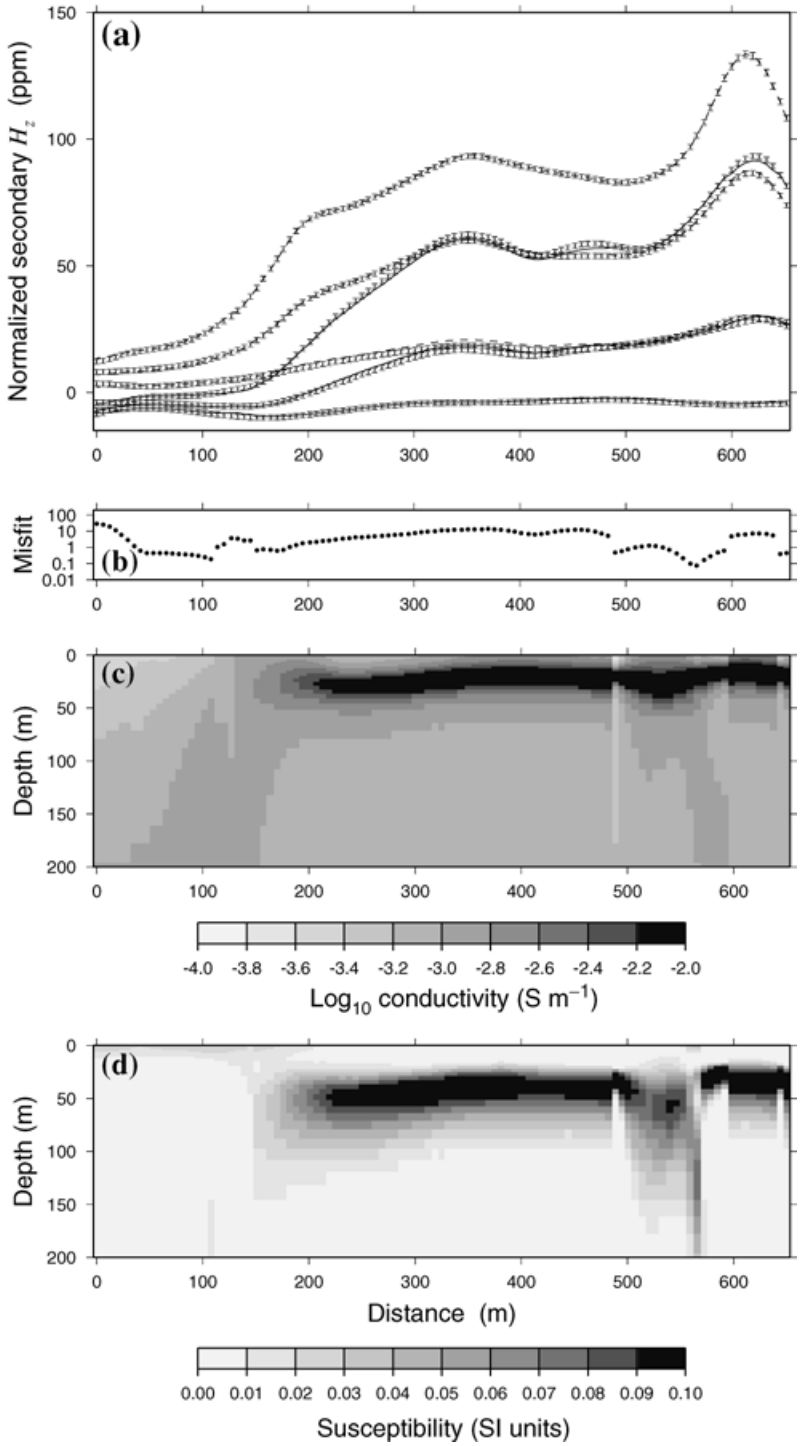

Figure 11. (a) The line of airborne data (error bars) discussed in Section 3.4 and the data (solid lines, inphase; dashed lines, quadrature) forward modelled from the 1-D models constructed by the inversions, using the L-curve criterion, of the six data (inphase and quadrature parts at the three frequencies 1325, 4925 and 11 025 Hz) at each horizontal location. (b) The final misfits attained by the inversions at each location. (c) The image created by concatenating the 1-D conductivity models constructed by the L-curve inversions for each location. (d) The corresponding image of the susceptibility.

forward modelling required to compute the misfit for each value of the regularization parameter is relatively time-consuming. For the example here, this made the discrepancy principle slower than the implementations of the GCV or L-curve criteria.

The line search over the GCV function versus (the logarithm of) the regularization parameter curve can also be made efficient, stepping along the curve from the value chosen at the previous iteration until the minimum is bracketed, and then contracting down into the minimum using a golden section search. The one limiting factor as far as speed is concerned is that the inverse of the matrix is required to compute the GCV function (see eq. 22).

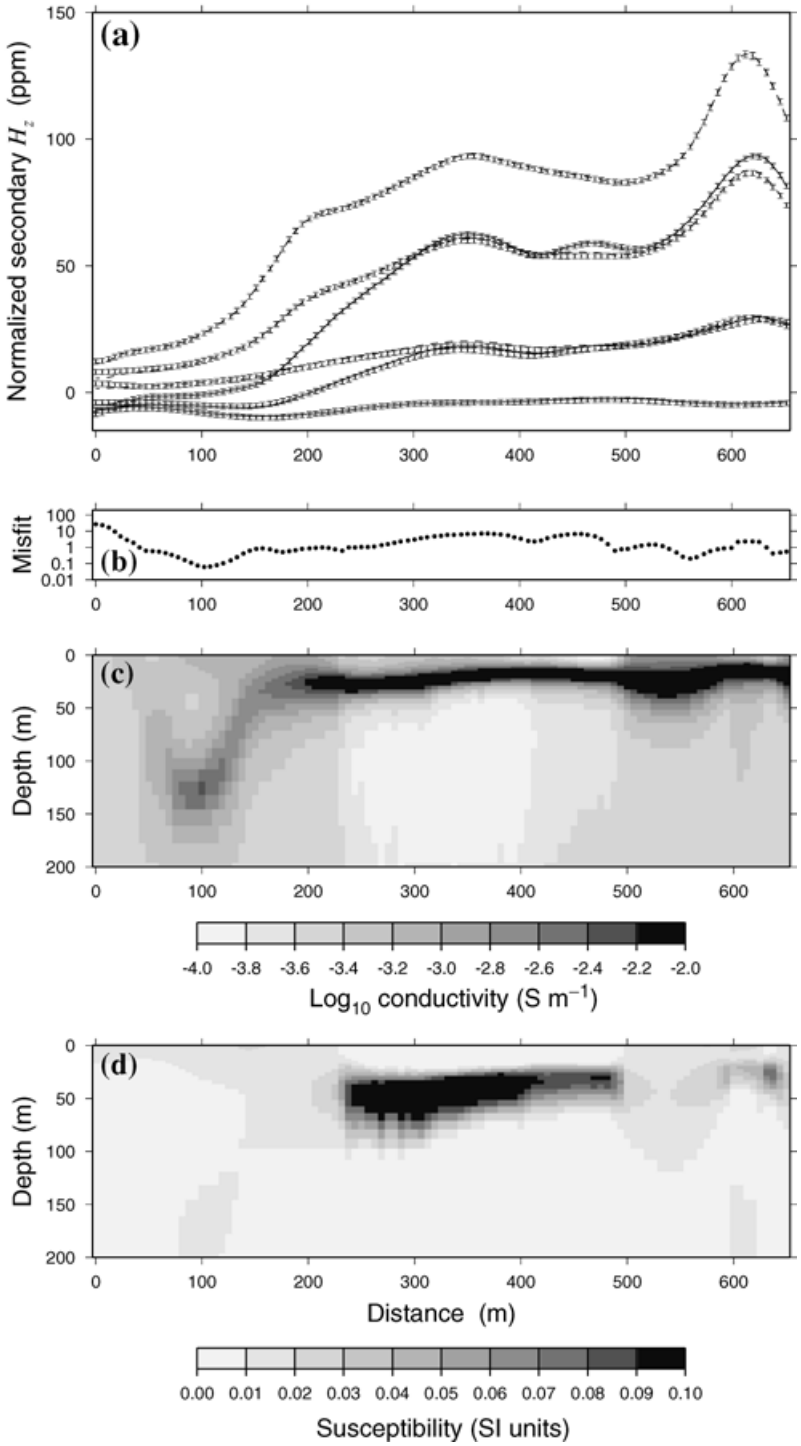

Figure 12. The line of airborne data discussed in Section 3.4 and the results of their inversion using the GCV-criterion. (a) The observations (error bars) and the forward-modelled data from the constructed 1-D models (solid, inphase; dashed, quadrature). (b) The misfits achieved at each location. (c) The conductivity image made by stitching together the 1-D models for every location. (d) The susceptibility image.

Implementation of the line search over the curvature of the L-curve is not so straightforward. It is susceptible to numerical noise because of the two numerical differentiations required (see eq. 26), and the slow variation with the regularization parameter of the functions being differentiated (the logarithms of the model norm and linearized misfit—see Fig. 4). It is therefore important to make the line search robust. We sample the curvature as a function of the regularization parameter at equally spaced values of the logarithm of the regularization parameter over a couple of orders of magnitude centred on the value of the regularization parameter from the previous iteration. We fit a parabola through the maximum of the sampled values and its two neighbouring points, and take the value of the regularization parameter corresponding to the maximum of this parabola. This implementation of the line search generally makes

the L-curve criterion slower than the GCV criterion, but has the desired robustness.

Finally, in this discussion of the details, we mention that the statistical basis of the GCV criterion would suggest that it cannot really be expected to work as well when the number of observations is small (less than ∼20, Haber 1997). However, this is somewhat contradicted by our experience: the GCV criterion has continued to be successful on numerous examples with as few as six (for example, the airborne data set in Section 3.4) or even four data.

5 CONCLUSIONS

We have illustrated the use of the GCV and L-curve criteria for automatically estimating the regularization parameter in iterative, minimum-structure, underdetermined inversion algorithms by incorporating them in the solution of a representative and yet computationally tractable non-linear geophysical inverse problem. In our experience, both criteria perform well, giving appropriate values of the regularization parameter in the vast majority of situations. Such automatic methods are particularly valuable in real-life situations for which it is too often the case that estimates of the measurement uncertainties are not acquired, meaning that we can only guess at the noise levels, and hence the target misfit in the traditional discrepancy principle method for choosing the regularization parameter.

ACKNOWLEDGMENTS

We would like to thank Robert Smith, formerly of Rio Tinto, now of Greenfields Geophysics, for providing us with the line of airborne data. The work presented here was funded by NSERC and the 'IMAGE' Consortium, of which the following were members: AGIP, Anglo American, Billiton, Cominco, Falconbridge, INCO, MIM, Muskox Minerals, Newmont, Placer Dome and Rio Tinto. We are grateful for their participation. We would also like to thank two anonymous reviewers and the editor Karsten Bahr for their comments that led to significant improvements in this paper. Finally, we would like to thank Eldad Haber for many useful discussions on inverse theory and Roman Shekhtman for his assistance with all things computer-related.

REFERENCES

Akaike, H., 1980. Likelihood and the Bayes procedure, in, Bayesian Statistics: Proc. 1st Int. Meeting, pp. 143–166, eds Bernardo, J.M., DeGroot, M.H., Lindley, D.V. & Smith, A.F.M., University Press, Valencia.

Amato, U. & Hughes, W., 1991. Maximum entropy regularization of Fredholm integral equations of the first kind, *Inverse Problems,* **7,** 793–808.

Constable, S.C., Parker, R.L. & Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics,* **52,** 289–300.

Dennis, J.E. & Schnabel, R.B., 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* SIAM, Philadelphia.

Farquharson, C.G. & Oldenburg, D.W., 1993. Inversion of time-domain electromagnetic data for a horizontally layered Earth, *Geophys. J. Int.,* **114,** 433–442.

Farquharson, C.G., Oldenburg, D.W. & Routh, P.S., 2003. Simultaneous 1D inversion of loop–loop electromagnetic data for magnetic susceptibility and electrical conductivity, *Geophysics*, **68,** 1857–1869.

Gill, P.E., Murray, W. & Wright, M.H., 1981. *Practical Optimization,* Academic, London.

Haber, E., 1997. Numerical strategies for the solution of inverse problems, *PhD thesis,* University of British Columbia.

Haber, E. & Oldenburg, D.W., 2000. A GCV based method for nonlinear ill-posed problems, *Comput. Geosci.,* **4,** 41–63.

Hansen, P.C., 1997 *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion,* SIAM, Philadelphia.

Li, Y. & Oldenburg, D.W., 1999. 3-D inversion of DC resistivity data using an L-curve criterion, *Expanded Abstracts of the 69th Ann. Int. Meeting of the Society of Exploration Geophysicists,* pp. 251–254. Society of Exploration Geophysicists, Tulsa, OK, USA.

Li, Y. & Oldenburg, D.W., 2003. Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method, *Geophys. J. Int.,* **152,** 251–265.

Mitsuhata, Y., Uchida, T. & Amano, H., 2002. 2.5-D inversion of frequency-domain electromagnetic data generated by a grounded-wire source, *Geophysics,* **67,** 1753–1768.

Newman, G.A. & Alumbaugh, D.L., 1995. Frequency-domain modelling of airborne electromagnetic responses using staggered finite differences, *Geophys. Prospect.,* **43,** 1021–1042.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P., 1992. *Numerical Recipes in FORTRAN: The Art of Scientific Computing,* Cambridge University Press, Cambridge.

Smith, J.T. & Bookear, J.R., 1988. Magnetotelluric inversion for minimum structure, *Geophysics,* **53,** 1565–1576.

Smith, R.C. & Bowers, K.L., 1993. Sinc–Galerkin estimation of diffusivity in parabolic problems, *Inverse Problems,* **9,** 113–135.

Uchida, T., 1993. Smooth 2-D Inversion for Magnetotelluric Data Based on Statistical Criterion ABIC, *J. Geomagnetism Geoelectricity,* **45,** 841–858.

Vogel, C.R., 1985. Numerical solution of a non-linear ill-posed problem arising in inverse scattering, *Inverse Problems,* **1,** 939–403.

Wahba, G., 1990. *Spline Models for Observational Data,* SIAM, Philadelphia.

Walker, S.E., 1999. Inversion of EM data to recover 1-D conductivity and a geometric survey parameter, *MSc thesis*, University of British Columbia.

Wright, S.J., 1997. *Primal-Dual Interior Point Methods,* SIAM, Philadelphia.

Zhang, Z. & Oldenburg, D.W., 1999. Simultaneous reconstruction of 1-D susceptibility and conductivity from electromagnetic data, *Geophysics,* **64,** 33–47.