JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 96, NO. B3, PAGES 3905–3922, MARCH 10, 1991

Rapid Inversion of Two- and Three-Dimensional Magnetotelluric Data

J. TORQUIL SMITH AND JOHN R. BOOKER

Geophysics Program, University of Washington, Seattle

We have developed an efficient iterative inversion method applicable to both two-dimensional (2D) and three-dimensional magnetotelluric data. The method approximates horizontal derivative terms with their values calculated from the fields of the previous iteration. The equations at each horizontal coordinate then become uncoupled. At each iteration this allows separate inversions for the improved conductivity profile beneath each measurement site. Resultant profiles are interpolated to form a new multidimensional model for which the fields for the next iteration are calculated. The method is extremely fast, and tests with 2D data show very promising results.

INTRODUCTION

Exploiting the full potential of magnetotellurics (MT) will require inversion of large quantities of data at many sites for structure which varies in three dimensions. Traditional multidimensional inversions (which we will call "standard") proceed as follows: The conductivity of the Earth is parametrized by values at a number of nodes or in a number of predefined elements. A starting model is guessed and a matrix, **F**, of partial derivatives of the data with respect to small changes in the parameters is calculated. This involves solutions of multiple forward problems. Singular value decomposition (SVD) or some other damped generalized inverse of $\mathbf{FF}^T$ is then used to predict the conductivity perturbations that should best improve the fit to the data. These perturbations are added to the initial guess to produce a new starting model. The forward problem is solved one more time to calculate new data residuals, and the whole process is repeated until a satisfactory fit to the data has been obtained. The computer resources necessary to generate and store **F** and to predict the model perturbations grow rapidly with the size of the model and the data set.

We have developed a different type of iterative scheme, which is applicable in both two and three dimensions and have implemented it for the two-dimensional (2D) case. The crucial step in this new method is to approximate the lateral gradients of the electric and magnetic fields in the model by their values from the previous iteration. One can then calculate a conductivity perturbation beneath each measurement site by solving an inverse problem that is closely related to the one-dimensional (1D) inverse problem. One need only solve one forward problem to calculate the data residuals and the fields inside the model for the next iteration. This forward problem can be very efficiently solved using an iterative method because the fields from the previous iteration are a good approximation to the new fields. Furthermore, the primary memory required is to store the model and the fields at each node. This will always be much less than required to store **F**. As long as our method converges in a reasonable number of iterations, it will be much more efficient than a standard method and its superiority will increase as data sets and models become larger. The next section presents a brief overview of the theoretical background for both two and three dimensions. This is followed by a discussion of numerical implementation and synthetic examples.

Paper number 90JB02416.
0148-0227/91/90JB-02416$05.00

THEORY

TE Mode

We consider first the transverse electric (TE) mode in which the electric field is parallel to the strike of the structure. Maxwell's equations in a good isotropic conductor assuming a time dependence of $e^{-i\omega t}$, no ferromagnetism and ignoring displacement currents reduce to

$$\nabla\times\mathbf{H} = \sigma\mathbf{E} \qquad \nabla\times\mathbf{E} = i\omega\mu_0\mathbf{H}$$

where **E** and **H** are the electric and magnetic fields, σ is the conductivity and μ_0 is the magnetic permeability of free space. For 2D σ, with x aligned with strike, y perpendicular to strike and z positive downward, these equations become

$$\nabla^2 E_x = -i\omega\mu_0\sigma(y,z)\, E_x \tag{1}$$

$$\frac{\partial E_x}{\partial z} = i\omega\mu_0 H_y \tag{2}$$

To simplify the remainder of this discussion, all variables, such as x, y, z or ω will be shown only when needed for clarity.

In order to solve the inverse problem, it is first necessary to determine the sensitivity of the data to changes in the conductivity. This can be done by a perturbation analysis closely related to that used for the 1D MT problem [*Oldenburg*, 1978]. Equation (1) can be conveniently rewritten

$$\frac{1}{E}\frac{\partial^2 E}{\partial z^2} + \left\{\frac{1}{E}\frac{\partial^2 E}{\partial y^2}\right\} + i\omega\mu_0\,\sigma = 0 \tag{3}$$

Without the term in braces, this would be the equation for induction in a 1D model. The data are defined to be the surface ($z = 0$) values of the variable

$$V = \frac{1}{E}\frac{\partial E}{\partial z} = i\omega\mu_0\frac{H_y}{E_x}$$

Note that V is closely related to the MT impedance ($Z_{xy} = E_x/H_y$). From this definition it is easy to show that

$$\frac{1}{E}\left(\frac{\partial^2 E}{\partial z^2}\right) = \frac{\partial V}{\partial z} + V^2$$

and that (3) becomes

$$\frac{\partial V}{\partial z} + V^2 + \left\{\frac{1}{E}\frac{\partial^2 E}{\partial y^2}\right\} + i\omega\mu_0\,\sigma = 0 \tag{4}$$

Note that the lateral gradients of the fields are confined to the term in braces.

Now suppose that V_0 and E_0 satisfy equation (4) when $\sigma = \sigma_0$. Let $\sigma = \sigma_0 + \delta\sigma$ and $V = V_0 + \delta V$. Since vertical field gradients are generally larger than horizontal gradients due to the skin depth effect, we make the approximation

$$\frac{1}{E}\frac{\partial^2 E}{\partial y^2} = \frac{1}{E_0}\frac{\partial^2 E_0}{\partial y^2}$$

Substituting in equation (4), subtracting the zeroth-order equation reduces this to a Ricatti equation for δV. Neglecting second-order terms gives the first-order linear differential equation

$$\frac{\partial}{\partial z}\delta V + 2V_0\delta V + i\omega\mu_0\,\delta\sigma = 0$$

which can be made an exact differential using an integrating factor. Vertical integration at each site y_i and use of the definition of V finally yields

$$\delta V(y_i,0) = \frac{i\omega\mu_0}{E_0^2(y_i,0)}\int E_0^2(y_i,z)\,\delta\sigma(z)\,dz \qquad (5)$$

The limits of integration are the surface and a depth which is sufficiently large that the electric field has effectively decayed to zero. This expression is complex and thus represents two equations at each frequency at each site.

The above expression is easily modified if one wants to use $\ln\sigma(z)$ as the model, because, for infinitesimal variations, $\delta\sigma = \sigma_0\delta(\ln\sigma)$. Another useful modification is to redefine the data as

$$d_{xy} = \ln\left[-i\omega\mu_0\left[\frac{H_y}{E_x}\right]^2\right] = \ln\left(\frac{V^2}{-i\omega\mu_0}\right) \qquad (6)$$

because

$$Re\,[d_{xy}] = -\ln\rho_a \qquad Im\,[d_{xy}] = 3\pi/2 - 2\phi$$

where ρ_a and ϕ are the apparent resistivity and impedance phase commonly reported for MT measurements. By differentiating (6) one can easily conclude that

$$\delta d_{xy} = \frac{2}{V(y_i,0)}\delta V = \int\frac{2\sigma_0(z)E_0^2(y_i,z)}{E_0(y_i,0)H_0(y_i,0)}\delta(\ln\sigma)\,dz \qquad (7)$$

where $H_0(y_i,0)$ is calculated from $E_0(y_i,z)$ using (2).

If the residuals on the left side of equations (5) or (7) are the difference between the measured data at each frequency and that predicted by σ_0, these equations can be inverted for a conductivity perturbation directly beneath the site. We will call these site inversions "pseudo-1D", because these "pseudo-Frechet" derivatives differ from the Frechet derivatives for the 1D case only because the field $E_0(y_i,z)$ used to compute them must satisfy the 2D equation (1) with $\sigma = \sigma_0(y,z)$.

The complete iterative cycle is as follows: An initial guess is made for $\sigma_0(y,z)$ and equation (1) is solved for $E_0(y,z)$. These fields are used to estimate the pseudo-Frechet kernels and the data residuals at each site. Equations (5) or (7) are then inverted using a smoothness criterion that guards against introducing superfluous structure. Next, $\sigma_0(y_i,z)$ is updated (usually with only a percentage of $\delta\sigma$ or $\delta(\ln\sigma)$ and interpolated between sites to produce a new $\sigma(y,z)$. Equation (1) is then solved again for new estimates of the 2D fields using a finite difference approximation and a fast, sparse matrix method. The cycle is repeated until the data residuals become sufficiently small. This scheme, where one holds one part of the problem fixed while operating on another, is reminiscent of relaxation techniques for solving differential equations, and we have therefore dubbed the method the "Rapid Relaxation Inverse" or more simply RRI.

TM Mode

Using the same coordinate system as the TE mode and the resistivity $\rho = 1/\sigma$, the relevant equations for the transverse magnetic (TM) mode, in which the electric currents flow perpendicular to the strike (i.e., in the y - z plane), become

$$\nabla^2 H_x + \nabla\rho\cdot\nabla H_x = -i\omega\mu_0 H_x \qquad (1')$$

$$\rho\frac{\partial H_x}{\partial z} = E_y \qquad (2')$$

Equation (1′) can be be rewritten

$$\frac{1}{H_x}\frac{\partial}{\partial z}\rho\frac{\partial H_x}{\partial z} + \left\{\frac{1}{H_x}\frac{\partial}{\partial y}\rho\frac{\partial H_x}{\partial y}\right\} + i\omega\mu_0 = 0 \qquad (3')$$

If we again drop the subscript on H_x and define the new datum

$$U = \frac{\rho}{H}\frac{\partial H}{\partial z} = \frac{E_y}{H_x} = Z_{yx}$$

the calculation of the approximate pseudo-Frechet derivative closely parallels the TE mode. We omit the details. The final result is

$$\delta U = \frac{1}{H_0^2(y_i,0)}\int E_0^2(y_i,z)\,\delta\sigma\,dz \qquad (5')$$

where E_0 is determined using (2′) and the field H_0 satisfying (1′) with $\rho = \rho_0$. Re-defining the TM data as

$$d_{yx} = \ln\left(-i\omega\mu_0\left[\frac{H_x}{E_y}\right]^2\right) = \ln\left(\frac{-i\omega\mu_0}{U^2}\right) \qquad (6')$$

we obtain the TM analog of the pseudo-Frechet derivative (7)

$$\delta d_{yx} = \frac{2}{U(y_i,0)}\delta U = \int\frac{-2\sigma_0(z)\,E_0^2(y_i,z)}{E_0(y_i,0)H_0(y_i,0)}\delta(\ln\sigma)dz \qquad (7')$$

The approximate pseudo-Frechet derivatives (5′) or (7′) can be inverted independently from the TE mode or can be combined with the TE mode pseudo-Frechet derivatives and data for a simultaneous inversion.

Extension to Three Dimensions

The same approach may be applied in three dimensions. The x and y components of (1) can be written

$$\frac{1}{E_x}\frac{\partial^2 E_x}{\partial z^2} + \left\{\frac{1}{E_x}\left(\frac{\partial^2 E_x}{\partial y^2} - \left[\frac{\partial^2 E_y}{\partial x\partial y} + \frac{\partial^2 E_z}{\partial x\partial z}\right]\right)\right\} + i\omega\mu_0\,\sigma = 0 \qquad (8)$$

$$\frac{1}{E_y}\frac{\partial^2 E_y}{\partial z^2} + \left\{\frac{1}{E_y}\left(\frac{\partial^2 E_y}{\partial x^2} - \left[\frac{\partial^2 E_x}{\partial x\partial y} - \frac{\partial^2 E_z}{\partial y\partial z}\right]\right)\right\} + i\omega\mu_0\,\sigma = 0 \qquad (9)$$

If the terms in braces are approximated using the field from the previous iteration, equations (8) and (9) have the same structure as equation (3) and can be perturbed and integrated in exactly the same way. This leads to two pseudo-Frechet derivatives, which both have the same form as equation (5), but involve different polarizations of the electric field. Magnetic field analogues of equations (8) and (9) also exist. There are thus four pseudo-Frechet derivatives that can be used independently or simultaneously for pseudo-1D inversions. After obtaining conductivity perturbations under each site, the three-dimensional

(3D) model is updated by interpolation and improved fields computed with a 3D forward calculation.

Alternatively, at least two approaches can be formulated using 2D inversions corrected for 3D effects along slices through a 3D structure. The numerical experiments of *Wannamaker et al.* [1984] suggest that a proper choice of strike and polarization can significantly reduce 3D effects on a 2D slice. Specifically, the source polarization with currents perpendicular to the long axis of a 3D body has impedances very close to TM mode impedances calculated for a 2D structure that is the same as the cross section of the 3D model. Thus slice approaches may have better convergence properties than the fully 3D approach using pseudo-1D inversions.

One slice method is to divide the iterative process into two levels. In the inner iteration, the terms in equations (8) or (9) (or their magnetic analogues) involving derivatives perpendicular to a 2D slice (the terms in square brackets) are held fixed, while a relaxation inversion of the type already discussed is performed for the slice. (Note, however, that the cross-slice derivative terms in square brackets must be treated as source terms in solving the 2D forward problems.) These "pseudo-2D" inversions are made for several slices which are then interpolated to form a 3D model. In the outer iteration, the cross-slice derivatives are updated using fields computed with a fully 3D forward calculation.

A second slice approach can be constructed by analogy to the 2D problem in which the pseudo-1D pseudo-Frechet derivatives are algebraically identical to 1D derivatives, but use 2D fields in place of the 1D fields. For three dimensions, standard 2D partial derivatives are calculated for each slice, with the 2D fields along the slice replaced by 3D fields. The details are left to Appendix A. These pseudo-2D partial derivatives are used to invert for a 2D perturbation along each slice and the improved 3D model is obtained by interpolating between slices. This method involves only one level of iteration.

COMPUTATIONAL CONSIDERATIONS

Pseudo-One-Dimensional Inversions

Our philosophy for this nonlinear inverse problem is to find models which are globally extreme in some sense. In principle, any 1D inversion could be used for the pseudo-1D inverses. However, the MT inverse problem is known for generating structure which is either unlikely or not required by the data [*Parker*, 1980; *Oldenburg et al.*, 1984]. Structure not required by the data can be misinterpreted and could make our 2D algorithm unstable. We therefore choose to look for models with minimum structure.

In 1D work it has been found that models which minimize mean-square derivatives of ln (σ) for a given misfit to the data [*Constable et al.*, 1987; *Smith and Booker*, 1988] are quite useful. For accurate synthetic data these inversions appear to converge to the actual structure as the misfit decreases, although no proof of convergence currently exists. For data with noise, these inversions produce models which look like smoothed versions of the actual structure as long as the data are not overfit (i.e., as long as the required misfit is greater than or equal to the expected misfit based on the noise in the data). If the data are overfit, the models tend toward Parker's *D*+ model, consisting of a series of positive delta functions. We shall refer to this as *D*+ behavior. It is a property of any 1D inversion of noisy data [*Parker*, 1980].

Appendix B generalizes a functional for measuring structure, found successful in one dimension, to incorporate a penalty for rapid horizontal variations. The specific form used here is the norm of a scaled Laplacian of the model m,

$$Q(y_i) = \int (z+z_0)^3 \left[\frac{\partial^2 m}{\partial z^2} + g(z) \frac{\partial^2 m}{\partial y^2} \right]^2 dz \qquad (10)$$

where the $(z+z_0)^3$ factor arises from taking derivatives and integrating with respect to log depth, and the $g(z)$ factor allows for trading-off between penalizing horizontal and vertical structures. When approximated with finite differences and a quadrature rule, this can be written in the form

$$Q_i = (\mathbf{R}\mathbf{m}_i - \mathbf{c}_i)^T (\mathbf{R}\mathbf{m}_i - \mathbf{c}_i)$$

where $\mathbf{m}_i$ is the discretely sampled model beneath the site y_i; $\mathbf{R}$ is a "roughening" operator which essentially performs a numerical vertical second derivative on $\mathbf{m}_i$; $\mathbf{c}_i$ is a vector that arises from penalizing the horizontal second derivative, it expresses our bias that structure under y_i must be similar to that beneath adjacent sites; and T indicates the matrix transpose.

Since the measured data are in error, we do not want the model minimizing Q_i that exactly fits the data. Instead, we want to minimize

$$W_i = Q_i + \beta_i e_i^2 \qquad (11)$$

The variances of the measurement errors have been used to scale the data and pseudo-Frechet derivatives so that the squared misfit Σe_i^2 is the standard χ^2 statistic. The constant β_i is a trade-off parameter between model structure and misfit. For the linear inverse problem, Appendix C presents an efficient and stable method for choosing β_i and finding $\mathbf{m}_i$ such that e_i^2 takes on a prescribed level. Since the actual inverse problem is nonlinear, it is necessary to reduce e_i^2 to its ultimate goal in a series of steps, small enough that the linearization inherent in (5) and (7), or (5′) and (7′) remains valid. In our work on the 1D inverse problem [*Smith and Booker*, 1988] we found that to keep the step size small, it is desirable to reduce the goal for Σe_i^2 in steps rather than always aiming for the final desired misfit. This results in a series of β_i variables increasing initially as the misfit gets smaller, and then settling down as the final desired misfit is approached. This behavior is also be seen in an example presented by *Constable et al.* [1987] which shows the optimal sequence of damping parameter $1/\beta$ for a specific 1D inversion.

However, a complication arises in two dimensions because W_i can increase due to a conductivity change under another site. This is a more serious problem for the TM mode, where electric currents flow along the profile than it is for the TE mode, where they flow across the profile. The success of a step must therefore be judged on a more global measure than the behavior of W_i at the single site. The single site objective functions can be combined to form a global measure in many different ways. The global measure of acceptability we prefer is

$$W_G = \sum_{i=1}^{n_{\text{sites}}} \frac{W_i}{\beta_i + \beta_{\text{median}}} \qquad (12)$$

This scales the single site objective functions so that the coefficient multiplying the squared misfit e_i^2 is of order 1, and W_G avoids being dominated by sites with very large or very small β_i. Since β_i and β_{median} change with iteration, the value of W_G for a candidate model is compared to its value for the last accepted model calculated using the β_i variables and β_{median} from the current iteration. This makes the functional measuring acceptability identical for both models. Experience indicates

that accepting a step only when W_G does not increase, rejects fewer steps and leads the iterative process astray no more often than any other global measure we have tried.

Step size control involves several interrelated issues and is the most difficult practical problem facing any implementation of RRI. We have adopted an ad hoc approach to step size control, which, although not rigorously proven to lead to convergence, works well in practice. The technique that we will outline may seem unnecessarily involved. It is important to remember, however, that the time for RRI is dominated by the forward problem. It is worth going to considerable effort to avoid a failed step. Our method is entirely automatic and has proven quite competent, making progress reasonably rapidly and seldom resulting in a failed step until we get close to the best model that can be achieved.

The most important issue is that the misfit at any iteration is due partly to statistical measurement errors and partly to inaccuracy of the pseudo-Frechet derivatives. Attempting to decrease e_i^2 to its ultimate goal prior to buildup of essentially correct 2D structure and, hence, reasonably accurate pseudo-Frechet derivatives, results in very rough pseudo-1D inversions, i.e., *D+* behavior. If the *D+* spikes occur at different depths under adjacent sites, horizontal model gradients will also be large, with bad effects on both the roughness matrix **R** and bias vector **c** and also on the accuracy of the forward calculation for the 2D fields used for the pseudo-Frechet derivatives. Once excessive structure of this kind has been introduced, it is very difficult to get rid of in subsequent iterations.

We control step size by controlling the factor by which the present misfit e_0^2 at site i is to be reduced at the current step, calling this factor ν_i. At the most fundamental level, we control step size by never trying to reduce the misfit by more than 50% in a single step ($0.5 \le 1.$), and by never allowing ν_i to be less than $(e_{\min}^2/e_0^2)^{\varepsilon})]$, where $e_{\min}^2$ is the smallest misfit that can be achieved by the linear inversion (see Appendix C). (We use $\varepsilon = 1/3$ in this paper, but have found values from 1/4 to 1/2 to be useful.) Within this range, ν_i is adjusted according to the success of the previous iterations. Whenever there are two successful 2D forward steps in a row ν_i is allowed to decrease toward its lower limit (0.5). The ν_i is increased only on certain failed 2D forward steps (see discussion of how to deal with failed steps, below).

After we have have found β_i at all the sites, we calculate their median β_{median} excluding sites where $\beta = 0$. (β_i becomes zero whenever smoothing the structure in from adjacent sites would result in a misfit less than the current goal at a site.) Sites with β_i greater than β_{median} are reinverted, setting β_i to the geometric mean of the original β_i and β_{median}. This roll-back of the large β_i variables is limited to the value $\beta_i^{\text{smoothest}}$ that results when we minimize (11) keeping the misfit at e_0^2. This prevents making a step for which the misfit is predicted to be worse, although it requires doing yet another pseudo-1D inversion.

The roll-back of large β variables just described causes 2D structure in the model to grow more slowly and more evenly and in a sense couples the sites together. We have generally found it useful to couple the sites additionally in a more explicit way. We horizontally smooth the residuals, using a centrally peaked window extending over five sites. This has the effect of fitting the long horizontal wavelengths in the data in early iterations and shorter wavelengths in later iterations.

The remedy for a failed step depends on what caused the failure. We compare the new model beneath each site to the the model associated with $\beta_i^{\text{smoothest}}$ to determine if the step is primarily trying to improve the misfit or to smooth the structure. (See *Smith and Booker* [1988] for further discussion of how this comparison is made.) If the former is the case, we try a less ambitious step with ν_i increased by taking its square root. If the latter is the case, or we have already tried the less ambitious step, we halve the mixing coefficient ξ_i in $\mathbf{m} = \mathbf{m}_0 + \xi_i \delta\mathbf{m}$. For the TE mode, $\xi_i \le 1$, but we have found that progress for the TM mode and for both modes together has fewer failures if its maximum is 0.8. The mixing coefficient is not increased again until at least two 2D forward steps in a row have been successful, and even then, we only increase it by a factor of $\sqrt{2}$.

Interpolation

Our horizontal interpolation of σ between the pseudo-1D inversion step and the 2D forward calculation employs cubic polynomials using four points (two on each side) flanking the region of interest. The polynomials are constrained so that when pieced together the first derivative is continuous everywhere. These have most of the benefits of ordinary cubic splines but are determined more locally. The time spent on interpolation is negligible.

Forward Problem

We solve the forward problem with a finite difference method. One could also implement RRI using a finite element method. The important point is not how the forward problem is solved but that the method be both fast and accurate. This is because the time involved in RRI is dominated by the forward problem and inaccuracies may lead to structure not required by the data, or worse, prevent convergence of the iterative algorithm.

Accurate finite difference approximations (FDA) have caused some difficulty for forward calculations when the model consists of discrete blocks separated by discontinuities. The interpolation scheme just described guarantees that both σ and its gradient are continuous everywhere in our model except at the air-Earth interface (and there only for the TE mode). This greatly simplifies finding an accurate FDA. The 2D Laplacian operator on the left side of (1) is approximated in the standard way using centered second differences computed from the central and four surrounding nodes, and the right side is approximated with the conductivity and electric field at the central node. The TM mode must also approximate the term in (3′) involving $\nabla\rho$ using centered first differences of the field and the resistivities at the adjacent nodes. In the TE mode the discontinuity at the Earth surface is handled in the way suggested by *Brewitt-Taylor and Weaver* [1978] and rigorously justified by *Smith* [1988]. This simply involves replacing the interface values of σ with $\sigma/2$, but otherwise treating the model as if it were continuous.

At the side edges of our model, we require that E and B be equal to their values in (potentially different) 1D models. An alternate condition on the horizontal gradients could easily be implemented. At the bottom of the model, $z = z_{\max}$, we use an impedance boundary condition

$$\frac{\partial E_x}{\partial z} = \sqrt{-i\omega\mu_0\sigma}\, E_x$$

When the bottom of the model is one dimensional and sufficiently far below any 2D structure, this extends the model to infinite depth. In the more general case with $\sigma(y)$ at the bottom, this condition has no simple physical analogy, but it should be better than demanding an infinitely conducting half-space with $E = 0$ at $z = z_{\max}$. For forward calculations in which σ is constant in the deep portion of the model, we have found that this boundary condition can be applied at surprisingly shallow depths

without significantly degrading the accuracy. For instance, for the "diamond model" described later, making the 10-ohm-m basal half-space only 1 km thick and containing only two nodes, perturbs TE or TM data at all periods by less than 1% over a model in which the half-space is well discretized to a depth of 200 km. However, for inversion the bottom of the model should be placed fairly deep as large, deep, lateral gradients may result in significant inaccuracies.

The top of the model is at the ground surface for the TM mode and at the top of an air layer for the TE mode. For both modes we make the horizontal magnetic field constant at the top of the model. In the TE mode, this implies

$$\frac{\partial E_x}{\partial z} = -i\omega\mu_0 H_y$$

at $z = z_{source}$. This condition is simple, consistent with our vertical edge conditions and geophysically reasonable.

The FDA plus the boundary conditions lead to a matrix equation $\mathbf{Ax} = \mathbf{b}$ in which A has only five nonzero diagonals. We solve this very sparse system iteratively using incomplete LU factorization (ILU) with orthomin acceleration [*Behie and Vinsome*, 1982]. Other methods such as multigridding [*Brandt*, 1977] and conjugate gradient methods [*Kershaw*, 1978] exist which may have advantages. Undoubtedly, further improvements, particularly on machines capable of large-scale parallel processing will become available and can be installed in the algorithm simply.

For each new RRI forward step, we initialize the fields with their values from the previous step. This considerably reduces the number of orthomin iterations required for a given accuracy, because these fields are a good approximation to the fields in the new model. We routinely use a maximum of only six or eight ILU iterations per RRI step.

Particularly for the TM mode, we have found it desirable to use 64 bit arithmetic in order to achieve adequate accuracy in the solution of the forward problem. This actually yields a speed improvement in many modern computers, whose floating point hardware is optimized for double precision. It does necessitate more storage, however. One may be able improve the accuracy of single precision forward calculations by reformulating them in terms of deviations from some average 1D model as exploited by *Wannamaker et al.* [1986].

Effort and Storage

A major reason for the relative slowness of a standard inversion method is that the partial derivative matrix **F** cannot be computed analytically. Conceptually, the simplest way to obtain **F** numerically is to place a small perturbation successively at each cell and solve the forward problems. However, the reciprocity of the Green's functions allows the necessary derivatives to be extracted if the perturbations are placed at the measurement sites rather than in the model cells [*Weidelt*, 1975]. This is implied by *Rodi*'s [1976] derivation of the 2D partial derivatives although he did not specifically state the result in this manner. Since there are typically more model parameters than sites, this organization of the calculation will usually be more efficient. Thus one must solve one forward problem for each site at each period. Application of a roughening matrix **R**, which is required if one is to minimize a spatial derivative rather than the mean square size of the perturbation doubles the number of forward problems to solve.

There are two ways to solve the multiple forward problems for the standard method. If the forward problem is initially solved using a complete LU decomposition of **Ax**, the additional forward problems necessary for Rodi's formulation and roughening require considerably less calculation than the initial decomposition, because only the right-hand side of the system $\mathbf{Ax} = \mathbf{b}$ needs to be changed. On the other hand, if the forward problem is solved with an iterative technique such as ILU, one has to completely solve the forward problem twice for each site. As model size increases, the efficiency of an iterative forward solution should eventually win out over complete decomposition, especially if fields from a previous iteration are used to start the ILU. Furthermore, the storage necessary for the complete decomposition of **Ax** exceeds that required for an iterative solver, so that ILU or its relatives may be desirable for standard inversions even when it is slower.

Table 1 compares the operation counts and memory for the various portions of one iteration in a 2D inversion. The "actual problem" is the "rift" example to be described later. The "large problem" is motivated by the largest MT data sets that presently exist and probably represents the size of problem that one should expect to routinely handle in the near future. The operation counts reported for ILU assume eight iterations. We have found that the iterative solution typically converges in less than eight iterations in the later RRI steps where the field changes are small and we have obtained good results when the number of ILU iterations is limited to as few as four for all RRI steps. Thus the operation counts in the table are conservative. The most obvious conclusion for the standard 2D inversions is that the time is dominated by forming **F** and by forming and factoring $\mathbf{HH}^T$ and the storage is dominated by the partials **F**. (The matrix **H** is a matrix obtained by repartitioning the unroughened partial derivative matrix $\mathbf{F}\tilde{\mathbf{R}}^{-1}$. **H** and $\tilde{\mathbf{R}}$ are defined in Appendix C.) For the large problem, time for forming and factoring $\mathbf{HH}^T$ dominates that needed to calculate **F**. Thus the efficiency of an iterative solution to the forward problem has little effect for standard 2D inversions. RRI, on the other hand, is completely dominated by the time and storage for the forward problem. RRI for the actual problem is predicted to be about 70 times faster than a standard inversion and require 5% of the storage. Corresponding figures for the large problem are more than 1000 and 1%.

On a workstation taking 3.0 s per double precision Mflop (1 Mflop = 10^6 multiplications plus 10^6 additions plus 10^6 assignments plus loop overhead), the iterative forward algorithm for the actual problem is predicted to take about 123 s for 12 frequencies. The observed time is 72 s, which is substantially less than predicted because the average RRI step required about four ILU iterations.

It is important to remember that the inverse problem is nonlinear and that the times in Table 1*a* are for calculation of a perturbation to a starting model. This perturbation must then be added to the starting model and the cycle repeated until a satisfactory fit to the data has been obtained and the global measure of acceptability (12) is small. A comparison of standard inversions with RRI is incomplete without considering the number of these outer iterations required. Figure 1 shows the rms misfit as a function of RRI iterations for the TM mode inversions of the "Prism" example described in detail in the next section. Plots for the other inversion examples of this paper are all very similar. The log misfit decreases approximately linearly until it reaches its specified goal, where it almost immediately levels out. The rate of decrease is independent of the noise in the data. Thus noisy data can be fit more rapidly than accurate data.

One would expect that fewer outer iterations are likely to be

TABLE 1*a*. Multiplications (Real Numbers) for Major Steps of a 2D Inversion

	Forward Problem $\mathbf{Ax} = \mathbf{b}$	"Forward" Problems for Partials $\mathbf{F}$	Roughening $\mathbf{FR}^{-1}$*	Forming and Factoring $\mathbf{HH}^T$*	Total
			General Form		
Standard (LU)	$4n_y n_z^3 n_f$	$12n_y n_z^2 n_f n_s$	$n_y n_z^3 + 6n_y n_z^2 n_f n_s$	$2n_y n_z n_f^2 n_s^2 + 4n_f^3 n_s^3$	
Standard (ILU)	$(72+112n_i)n_y n_z n_f$	$(72+112n_i)n_y n_z n_f n_s$	$(36+56n_i)n_y n_z n_f n_s$	$2n_y n_z n_f^2 n_s^2 + 4n_f^3 n_s^3$	
RRI	$(72+112n_i)n_y n_z n_f$	--	--	$2n_z n_f^2 n_s + 4n_f^3 n_s$	
			Actual Problem†		
Standard (LU)	$3.8\ 10^8$	$7.4\ 10^8$	$3.8\ 10^8$	$1.2\ 10^9$	$2.7\ 10^9$
Standard (ILU)	$4.0\ 10^7$	$1.2\ 10^9$	$6.2\ 10^8$	$1.2\ 10^9$	$3.1\ 10^9$
RRI	$4.0\ 10^7$	--	--	$6.4\ 10^5$	$4.1\ 10^7$
			Large Problem‡		
Standard (LU)	$1.7\ 10^{10}$	$5.0\ 10^{10}$	$2.5\ 10^{10}$	$3.6\ 10^{11}$	$4.5\ 10^{11}$
Standard (ILU)	$4.1\ 10^8$	$4.1\ 10^{10}$	$2.0\ 10^{10}$	$3.6\ 10^{11}$	$4.2\ 10^{11}$
RRI	$4.1\ 10^8$	--	--	$2.9\ 10^7$	$4.4\ 10^8$

*See Appendix C.
†Grid is $n_y = 72$ by $n_z = 48$ with $n_f = 12$ frequencies at $n_s = 31$ sites; orthomin iterations $n_i = 8$.
‡Grid is $n_y = 140$ by $n_z = 100$ with $n_f = 30$ frequencies at $n_s = 100$ sites; orthomin iterations $n_i = 8$.

required by a standard method. However, *deGroot-Hedlin and Constable* [1990] report needing 10 outer iterations to fit data with an accuracy of 3% and 5 more iterations to remove remaining structure for a minimum structure 2D standard inversion. Thus the speed superiority of RRI is largely realizable. Furthermore, the standard method envisioned in Table 1 is the most efficient we can visualize. It does not actually exist. Implementations of the standard 2D inverse such as *deGroot-Hedlin and Constable* [1990] (see Appendix C) and *Jupp and Vozoff* [1977] are much slower.

Madden and Mackie [1989] sketch a conjugate gradient approach to solving the MT inverse problem which does not require solving for the partial derivative matrix **F**. They report good results with five conjugate gradient iterations per inversion iteration requiring the solution of 11 forward problems per iteration (plus an unspecified amount of additional calculation), so their method should be substantially more efficient than the standard method of Table 1, but still slower than RRI.

Figure 1 also plots a global structure functional $Q_G = \sum_{i=1}^{n_{sites}} Q_i$ against RRI iteration number. This functional initially rises as structure is added to the very smooth starting model (a half-space). It reaches a maximum and then declines. The highest value in the early iterations occurs for the noisiest data. This is due to local graininess of the structure associated with the noise. As convergence approaches, Q_G decreases with increasing noise, because less structure is required to fit noiser data within their errors. Q_G levels out at an iteration number somewhat higher than the convergence of the rms misfit. This reflects the fact that the algorithm continues to remove unnecessary structure after it has fit the data. One can continue the process of removing very minor details in the structure for many tens of iterations. As a practical matter, the structure changes very little after 20 iterations for all levels of error, even when the global measure continues to change slowly.

One must note that Q_G does not always continue to decrease monotonically on approaching convergence, as can be seen in Figure 1. This is because the data at each site are inverted to minimize only a single Q_i and not Q_G: smoothing the model below one site may require making it rougher below an adjacent site to continue fitting the data at the adjacent site. One could

TABLE 1*b*. Storage Requirements (Real Numbers) for Major Arrays of a 2D Inversion

	Forward Problem $\mathbf{Ax} = \mathbf{b}$	Starting Fields $\mathbf{x}_0$	Partials $\mathbf{F}$	$\mathbf{HH}^T$	Total
		General Form			
Standard (LU)	$6n_y n_z^2$	--	$2n_y n_z n_f n_s$	$4n_f^2 n_s^2$	
Standard (ILU)	$30n_y n_z$	$2n_y n_z n_f$	$2n_y n_z n_f n_s$	$4n_f^2 n_s^2$	
RRI	$30n_y n_z$	$2n_y n_z n_f$	$2n_z n_f$	$4n_f^2$	
		Actual Problem			
Standard (LU)	$1.0\ 10^6$	--	$2.6\ 10^6$	$5.5\ 10^5$	$4.1\ 10^6$
Standard (ILU)	$1.0\ 10^5$	$8.3\ 10^4$	$2.6\ 10^6$	$5.5\ 10^5$	$3.3\ 10^6$
RRI	$1.0\ 10^5$	$8.3\ 10^4$	$1.2\ 10^3$	$5.8\ 10^2$	$1.9\ 10^5$
		Large Problem			
Standard (LU)	$8.4\ 10^6$	--	$8.4\ 10^7$	$3.6\ 10^7$	$1.3\ 10^8$
Standard (ILU)	$4.2\ 10^5$	$8.4\ 10^5$	$8.4\ 10^7$	$3.6\ 10^7$	$1.2\ 10^8$
RRI	$4.2\ 10^5$	$8.4\ 10^5$	$6.0\ 10^3$	$3.6\ 10^3$	$1.3\ 10^6$

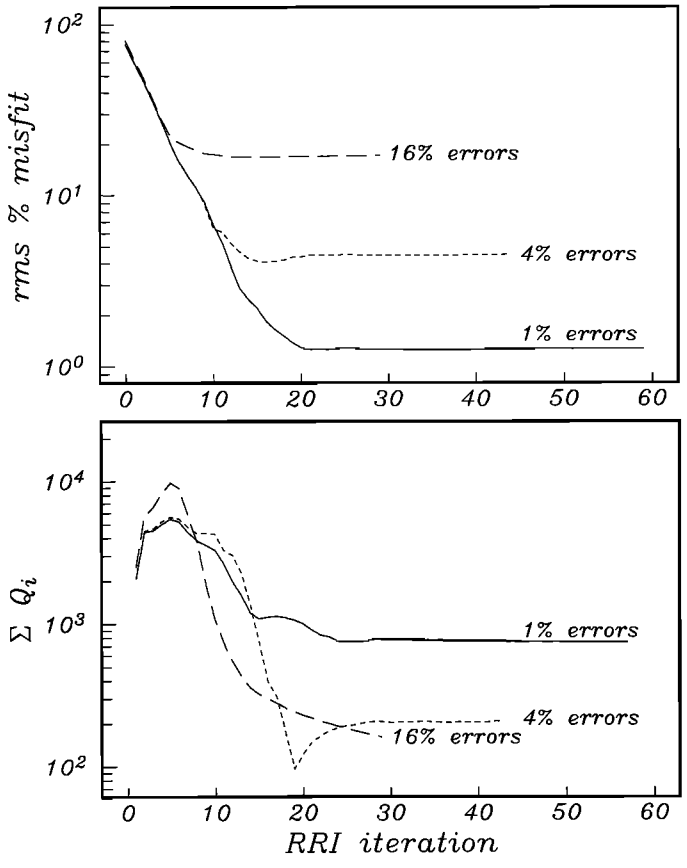

Fig. 1. (Top) The rms percent misfit as a function of iteration for TM mode inversions of Prism example of Figure 3. (Bottom) Global sum of individual site roughness norms (B1) $Q_G = \Sigma_{i=1}^{n_{sites}} Q_i$ as a function of iteration, for the same inversions.

explicitly minimize Q_G by inverting all the sites together in a simultaneous inversion. This would, undoubtably, weight the data and the model roughness somewhat differently. It is questionable whether this would often be merited since computation time for the linearized inversion increases as the third power of the number of data inverted simultaneously.

The relative effort spent on the different steps is somewhat different in the 3D case. Estimates of multiplications and storage for 3D inversions of a single polarization are shown in Table 2. Even with a modest 40 x 40 x 40 nodal mesh, the forward time now dominates the standard inverse with LU decomposition. Thus an iterative solution of the forward problem is warranted for all cases. Iterative solution of the 3D forward problem requires more storage and multiplications than for a 2D problem with the same number of nodes, because the number of diagonals in A increases to 12. For this larger number of diagonals, incomplete Cholesky decomposition with conjugate gradient preconditioning (ICCG) [*Kershaw*, 1978] requires fewer operations. Table 2 therefore assumes ICCG. Table 2 also assumes that twice as many iterations will be needed for the 3D forward problem than were required in two dimensions. We feel that this is a conservative position, but it remains to be demonstrated. With these assumptions, the standard inverse with the iterative forward method is expected to be 17 times faster and require one fifth the memory of complete LU decomposition.

RRI appears to offer speed increases of about 200 over the iterative standard method with ICCG and 4000 over LU. The order of magnitude decrease in memory is also important, because the standard inversions would be difficult to implement even on a supercomputer. Comparison of the pseudo-1D and pseudo-2D versions of RRI is not straightforward, despite the rough equality of the total times in Table 2a. Since the time needed for an iteration is of the order of the 3D forward time, the best method may be the one that results in the smallest number of 3D outer iterations. As noted earlier, there are good reasons to believe that this will be a 2D slice method. Using the pseudo-2D kernels of Appendix A is equivalent to doing a 2D standard inversion for each slice. The other slice method described earlier is equivalent to a 2D RRI inversion for the slice. It would produce the slice structure much more rapidly, but this increase efficiency would be illusory if it resulted in many more 3D outer iterations.

RESULTS FOR SYNTHETIC DATA

We present inversions of synthetic data for three different 2D models. The structures used to generate the data are shown in Figure 2. We shall refer to them as the Prism, Diamond and Rift

TABLE 2*a*. Multiplications (Real Numbers) for Major Steps of a 3D Inversion

	Forward Problem $\mathbf{Ax} = \mathbf{b}$	"Forward" Problems for partials F	Roughening $\mathbf{FR}^{-1}$*	Forming and Factoring $\mathbf{HH}^T$*	Total
			General Form		
Standard (LU)	$4n_x n_y^3 n_z^3 n_f$	$12n_x n_y^2 n_z^2 n_f n_s$	$n_x n_y^3 n_z^3 + 6n_x n_y^2 n_z^2 n_f n_s$	$2n_x n_y n_z n_f^2 n_s^2 + 4n_f^3 n_s^3$	
Standard (ICCG)	$(224+186n_i)n_x n_y n_z n_f$	$(224+186n_i)n_x n_y n_z n_f n_s$	$(37+31n_i)n_x n_y n_z n_f n_s$	$2n_x n_y n_z n_f^2 n_s^2 + 4n_f^3 n_s^3$	
RRI (pseudo-1D)	$(224+186n_i)n_x n_y n_z n_f$	--	--	$2n_x n_f^2 n_s + 4n_f^3 n_s$	
RRI (slice)†	$(224+186n_i)n_x n_y n_z n_f$	$4n_y n_z^3 n_f n_{s_x} + 12n_y n_z^2 n_f n_s$	$n_y n_z^3 + 6n_y n_z^2 n_f n_s$	$2n_y n_z n_f^2 n_{s_y}^2 n_{s_x} + 4n_f^3 n_{s_y}^2 n_{s_x}$	
			Sample Problem‡		
Standard (LU)	$6.6\ 10^{12}$	$1.2\ 10^{12}$	$7.8\ 10^{11}$	$1.3\ 10^{11}$	$8.7\ 10^{12}$
Standard (ICCG)	$2.0\ 10^{9}$	$2.0\ 10^{11}$	$3.4\ 10^{10}$	$1.3\ 10^{11}$	$3.7\ 10^{11}$
RRI (pseudo-1D)	$2.0\ 10^{9}$	--	--	$1.2\ 10^{6}$	$2.0\ 10^{9}$
RRI (slice)	$2.0\ 10^{9}$	$1.8\ 10^{9}$	$3.9\ 10^{8}$	$3.6\ 10^{8}$	$4.6\ 10^{9}$

*See Appendix C.
†With 2D partials; see Appendix A.
‡Grid is $n_x = 40$ by $n_y = 40$ by $n_z = 40$ with $n_f = 10$ frequencies at $n_s = 100$ total sites in $n_{s_y} = 10$ lines of $n_{s_x} = 10$ sites; ICCG iterations $n_i = 16$

TABLE 2*b*. Storage Requirements (Real Numbers) for Major Arrays of a 3D Inversion

	Forward Problem $\mathbf{Ax} = \mathbf{b}$	Starting Fields $\mathbf{x}_0$	Partials $\mathbf{F}$	$\mathbf{HH}^T$	Total
		General Form			
Standard (LU)	$6n_x n_y^2 n_z^2$	--	$2n_x n_y n_z n_f n_s$	$4n_f^2 n_s^2$	
Standard (ICCG)	$46n_x n_y n_z$	$6n_x n_y n_z n_f$	$2n_x n_y n_z n_f n_s$	$4n_f^2 n_s^2$	
RRI (pseudo-1-D)	$46n_x n_y n_z$	$6n_x n_y n_z n_f$	$2n_z n_f$	$4n_f^2$	
RRI (slice)	$46n_x n_y n_z$	$6n_x n_y n_z n_f$	$2n_y n_z n_f n_s + 6n_y n_z^2$	$4n_f^2 n_{s_y}^2$	
		Sample Problem			
Standard (LU)	$6.1\ 10^8$	--	$1.3\ 10^8$	$4.0\ 10^6$	$7.5\ 10^8$
Standard (ICCG)	$2.9\ 10^6$	$3.8\ 10^6$	$1.3\ 10^8$	$4.0\ 10^6$	$1.4\ 10^8$
RRI (pseudo-1D)	$2.9\ 10^6$	$3.8\ 10^6$	$8.0\ 10^2$	$4.0\ 10^2$	$6.8\ 10^6$
RRI (slice)	$2.9\ 10^6$	$3.8\ 10^6$	$3.6\ 10^6$	$4.0\ 10^4$	$1.0\ 10^7$

models. We generated synthetic data for all models using the finite element program of *Wannamaker et al.* [1986]. We decided to use a completely different forward algorithm in generating and inverting the synthetic data to help reveal any major systematic problems with our forward calculations. The 12 periods from 5 to 800 s are equally spaced logarithmically. In each case the bandwidth of the data is such that the skin depth at the shortest period is less than the scale of the surface structure while the skin depth of the longest period is much larger than the large scale 2D structure. The sites are spaced more closely than the scale lengths of the anomalies to ensure adequate sampling. The models for generating the data are all very well discretized, and the errors are judged to be about 1%, with the TE data being somewhat more accurate than the TM data.

Normally distributed noise at various levels was generated and added to the accurate synthetic data. As might be expected, the individual sites' variances vary somewhat from the overall variance. If we assumed that each site had the overall variance, the pseudo-1D inversions tended to overfit the sites with high noise and underfit the sites with low noise. This led to *D*+ behavior at the overfit sites. We therefore calculated the actual variances at each site and used these to set the ultimate misfit goals for the individual sites. We rather quickly found that trying to reduce the misfit to $e^2_{i\,\text{expected}} = n_{\text{data}}$ at each site led to unnecessarily rough models. Backing off on the ultimate goal to 1.5 $e^2_{i\,\text{expected}}$, which corresponds to the 95% confidence level and an rms misfit only 1.2 times the expected misfit, usually produced models which were much smoother and had little or no extraneous details.

The grids used to invert the data are different from those used to generate them. For generating synthetic data it is necessary to ensure that node spacing is sufficiently small to properly parameterize rapid field variations near boundaries of the discrete model elements. For inverting the data, vertical node spacing at the surface is made less than one tenth of the skin depth deduced from the lowest observed ρ_a at the shortest period. The spacing of the rest of the vertical nodes is designed to keep the grid as uniform as possible to take advantage of the higher accuracy of the FDA on a uniform grid, yet sparse enough to avoid ill-conditioning the system $\mathbf{Ax} = \mathbf{b}$ and consequent poor convergence of the ILU solution. We know that the horizontal conductivity variations must be quite smooth. The horizontal nodes are therefore made uniform where there are sites and grow logarithmically outside of the region with data. We have found that spacing comparable to the site spacing usually works well. Adding nodes between the sites only makes a significant difference if the sites are very far apart or quite uneven.

We have started our inversions from a wide variety of models. There is some tendency for the final result to depend on the starting model, especially if it contains significant 2D structure that is not required by the data. We have found that a 1D minimum structure inversion of the horizontally averaged data makes an extremely effective starting model. (We average TE data 6 for TE mode inversions or TM data (6′) for TM mode inversions, weighting the individual data by their attributed errors, so that the 1D starting model will be the smooth 1D model that best fits the 2D data.) The inversions of the Diamond and Rift data were started this way. The Prism model was started from a 50-ohm-m half-space.

The Prism model is about as simple a test case as one can devise. The conductive and resistive prisms are embedded in a half-space and are far enough apart that their anomalies have little interaction. Figure 3 shows the TM results for added noise levels of 1, 4 and 16%. The figure is made by filling boxes centered on each node with the appropriate gray level. Thus the distribution of boxes give direct information about the grid. As one would hope, the structure fades as the noise level increases, but little extraneous structure appears.

Figure 4 compares inversions with 1% added noise for TE, TM and TE joint with TM. The TE mode images the conductive prism well but is barely sensitive to the resistive prism. The TM mode images both prisms about equally well in the horizontal direction but tends to stretch them out vertically. This vertical stretching can be decreased by reducing the parameter η that controls the depth dependence of the horizontal structure penalty in the norm (see Appendix B). However, it becomes difficult to fit the data to high accuracy if $\eta<1$. The joint inversion appears to image both prisms somewhat better than either mode alone. However, the joint inversion could not attain its ultimate misfit goal, and the side lobes below the conductive prism suggest that we are overfitting systematic errors. Systematic biases of order 1% are quite likely in the forward algorithms used to generate the data and to invert them, and there is considerable reason to expect these biases to be different for the two modes.

The Diamond model reduces the high symmetry of the Prism model. It also places the anomalous bodies close enough to interact with one another and, by introducing a conductive basement, it has significant 2D anomalous currents flowing at some distance from the 2D structure. Figure 5 shows inversions with

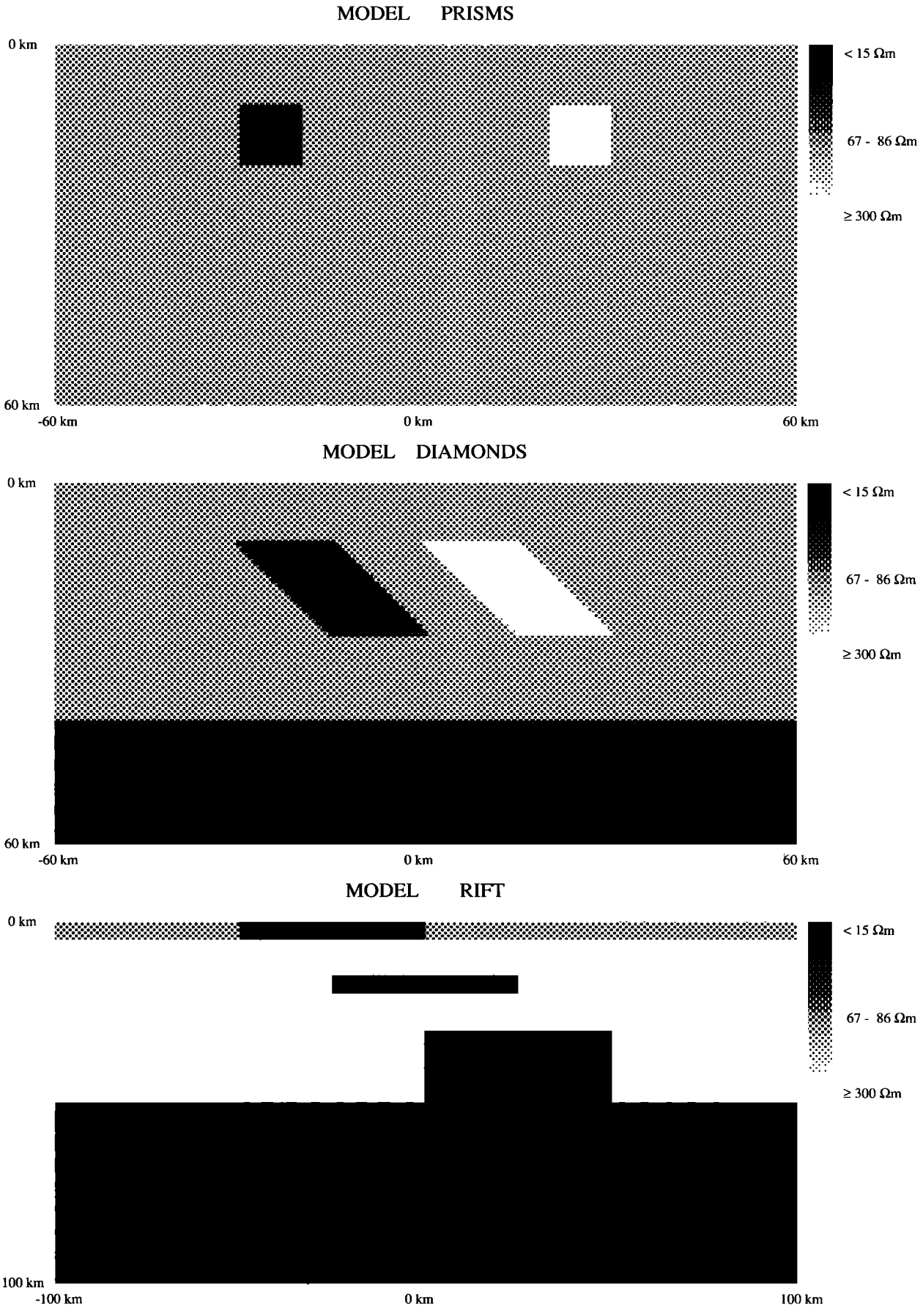

Fig. 2. The models used to generate synthetic data. The finite element meshes for each model are Prisms - 97 horizontal x 58 vertical + 8 air, 400 km wide x 300 km deep; Diamonds - 117x73+8, 400x200 km; Rift - 122x78+9, 600x200 km. The sawtooth edge of the Diamonds was actually modeled by a straight edge with the triangular elements of the finite element program. The resistivities of the elements are black - 10 ohm-m; gray - 100 ohm-m; white - 1000 ohm-m. Note the different scale for the third model. These pictures have the same scales as the inversions in the other figures.

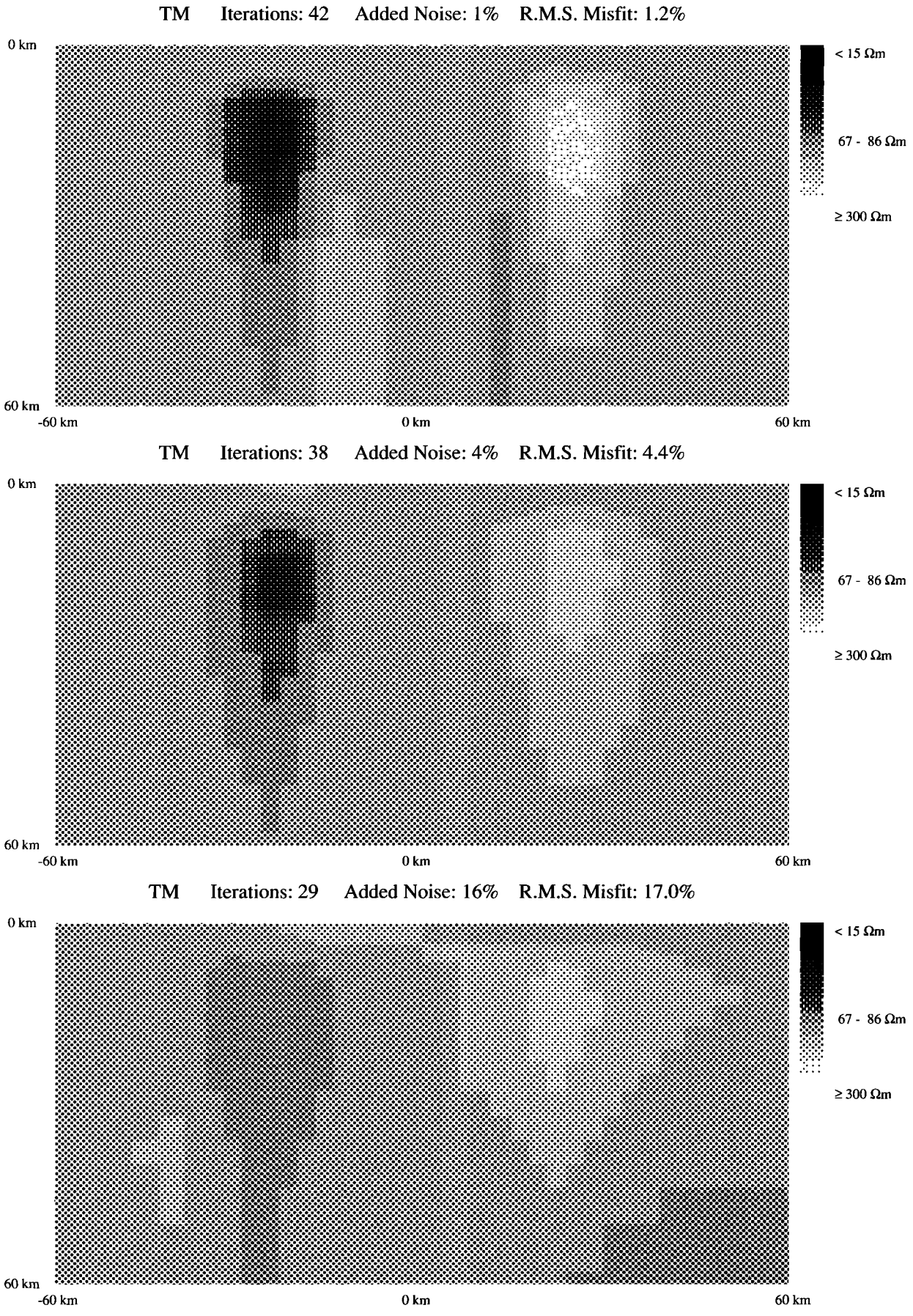

Fig. 3. Inversion of synthetic TM mode data for the Prisms model with increasing added noise. There are 30 sites mostly spaced 3 km apart. Another inversion with 6-km site spacing produced a result with slightly less focus. The forward modeling grid has 44x35+8 nodes, is 400 km wide and extends to 100 km depth. The vertical node spacing is 500 m at the surface and 10 km at the bottom. The display uses boxes centered on the nodes.

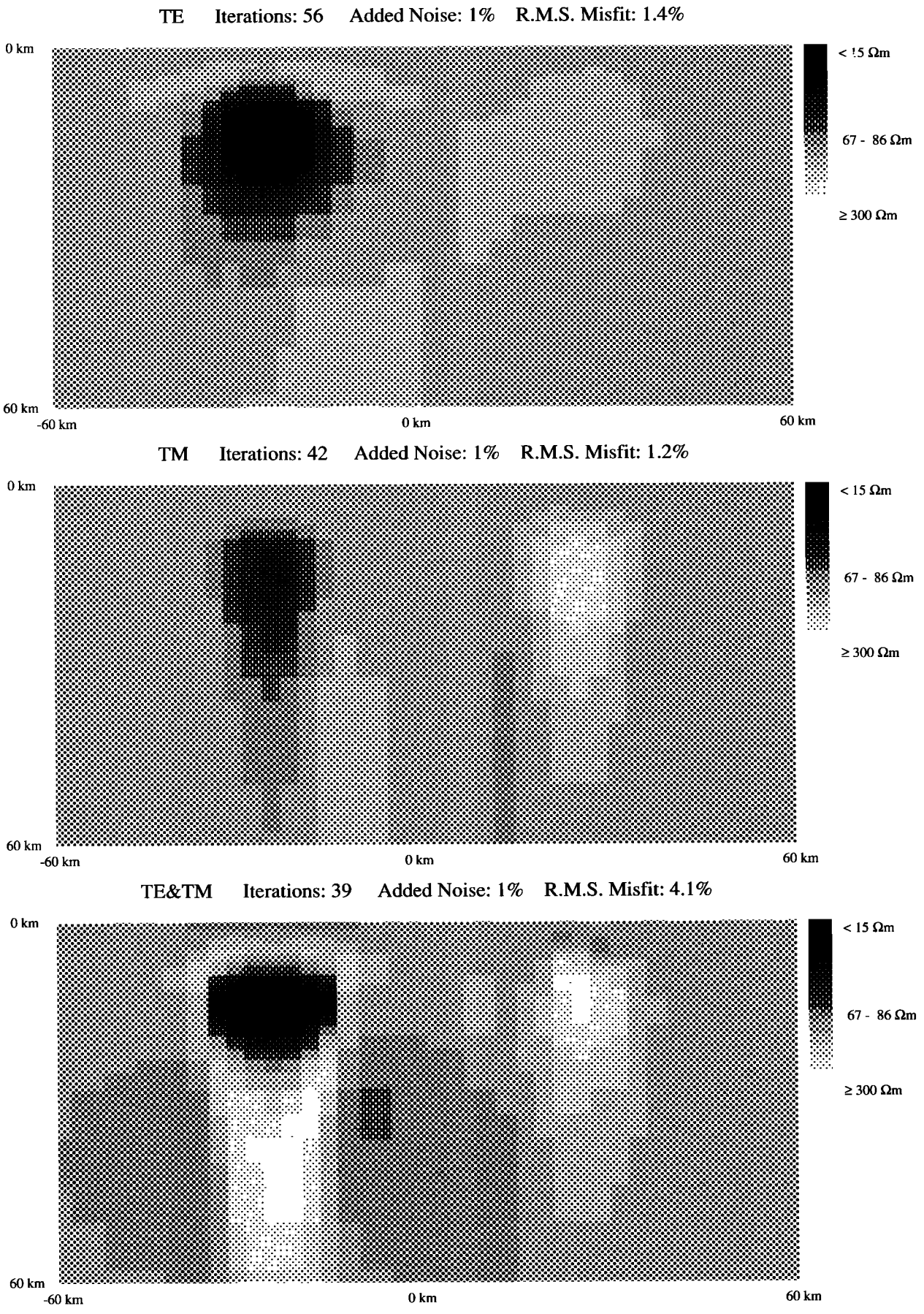

Fig. 4. Inversion of synthetic TE, TM and TE joint with TM mode data for the Prisms model with 1% added noise.

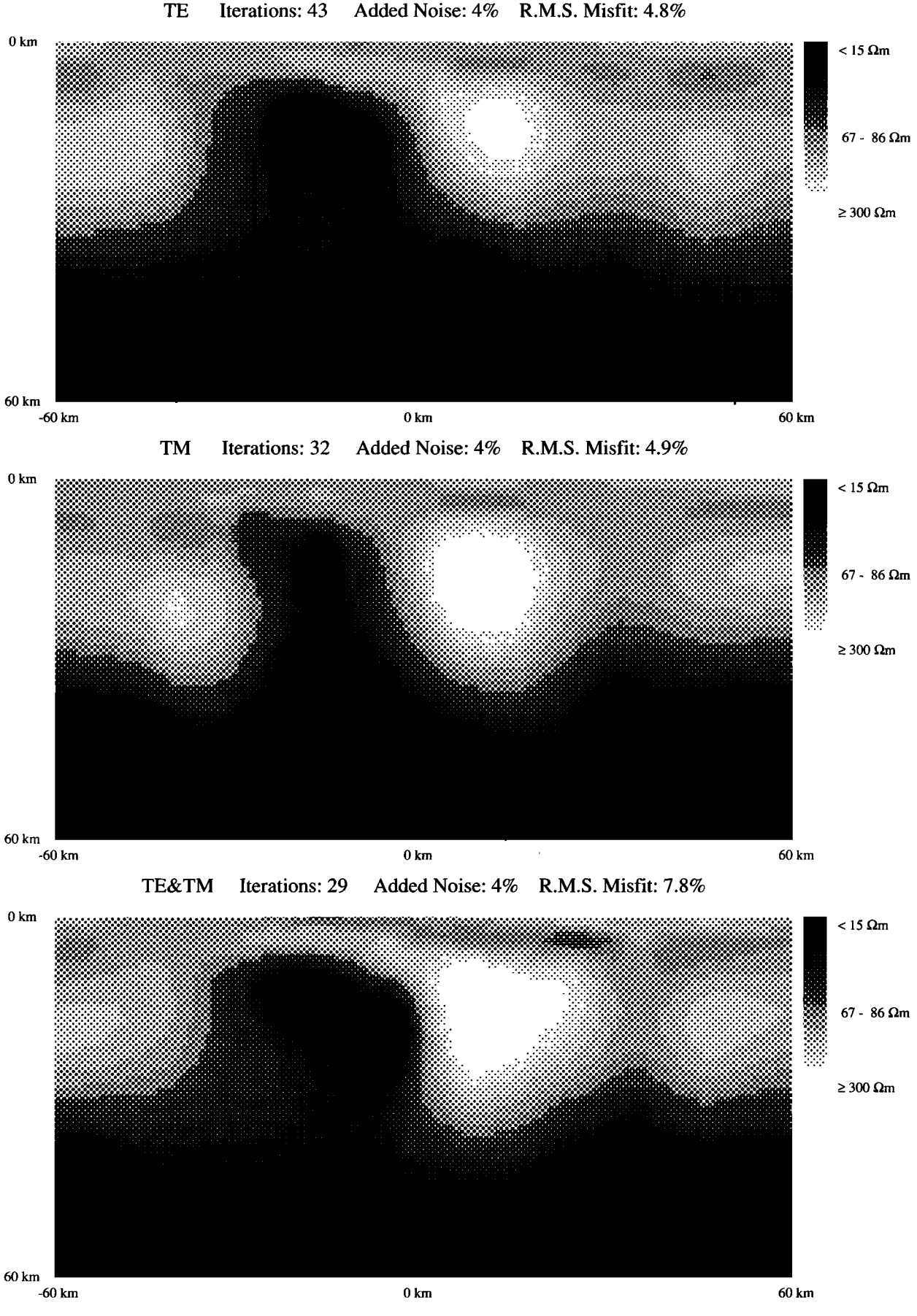

Fig. 5. Inversion of synthetic TE, TM and TE joint with TM mode data for the Diamonds model with 4% added noise. There are 27 sites mostly spaced 4 km apart. The forward modeling grid has 42x40+8 nodes, is 400 km wide and extends to 100 km depth. The vertical node spacing is 500 m at the surface and 5 km at the bottom. The display differs from Figures 3 and 4 in having boxes on a 1-km grid that are interpolated using the same algorithm that interpolates the pseudo-1D inversions.

4% added noise for TE, TM and TE joint with TM. This figure was made by interpolating the pseudo-1D inversions onto a 1-km grid using the same algorithm that we use to interpolate them onto the grid for the forward problem. The results are presented with $\eta = 1.25$ (all other results use $\eta = 1.5$), which suppresses some unwanted, small-scale horizontal variations in the basement. Further decrease of η made accurate fits difficult and led to step failure without further improvement in fidelity of the model. All models show evidence of the tilt of the conductive diamond and clearly detect the existence of the resistive diamond. There is essentially no evidence in Figure 5 of the tilt of the resistive diamond. Thus despite the impression gained from the Prisms model, both modes have lower resolution of resistive structure. Once again, the achievable misfit for the joint inversion is 3% higher than for either mode alone but produces a model which appears more focussed than either mode alone.

The final model is a structure that one might find near a continental rift. There is a widespread shallow crust of moderate resistivity over a resistive basement. A sediment-filled rift valley overlies a horizontal magma chamber at middepth and an upwarping of a deeper asthenospheric conductor. The three elements of the model are horizontally offset to make it easier to understand their interaction. To make the problem more difficult, the conductance of the valley (conductivity times depth) is the same as the middepth feature. Furthermore, the thickness of the valley and the magma chamber is only slightly larger than the skin depth at the shortest period. Conventional wisdom would say that one would have difficulty imaging the resistive region under the valley and one could not image the middepth feature under the valley. This is because fields whose skin depths are comparable to the thickness of the middepth feature will be strongly attenuated by the surface structure.

Figure 6 shows TE mode inversions for added noise of 0, 1 and 4% using the forward problem grid for the display. For additional added noise from 8 to 32%, the model changes remarkably little from the 4% model. The ultimate goal for inverting the 0% data was 0.3% at all sites. The inversion of this high-accuracy data has done a remarkable job of imaging all the important features of the actual structure. Not only have we very clearly imaged the base of the valley and the top of the middepth conductor, we have actually imaged the base of the middepth conductor. The middepth conductor is still detected by the 1% data but is no longer required under the valley at 4%. One expects that at a given level of noise the required conductivity gradient at depth will be lower where the fields have been attenuated by shallower structure. This is very evident under the valley, where the high gradient at the top of the basement gets spread out vertically as the noise increases giving the impression of an apparent downward deflection of the basement. These results demonstrate that the conventional wisdom about the resolving power of induction data is wrong. Skin depth does play a role in resolution, but if you have accurate enough data, you can image details, whose scale is much less than the skin depth.

The TM mode inversion for 4% added noise is shown in the top panel of Figure 7. This model has imaged the valley and the modestly conducting upper crust about as well as the TE mode, but it has several other clearly undesirable features. To begin with, the structure emerged in its final form in about 10 iterations. Thereafter, almost 75% of all forward steps failed. Although the final result is quite smooth and has an acceptable misfit, inverting data with less noise does not result in a significantly different model or a better misfit. There is also no obvious evidence of a midcrustal conductor and the asthenospheric upwelling is in the wrong place. In fact, its position appears to be controlled by the valley. We have tried a variety of experiments to try to understand what is happening. One was to start the TM mode from the TE results for accurate data. The algorithm rather quickly converged to a model almost identical to the top panel of Figure 7. This makes it unlikely that we have simply found a local minimum. Another experiment, in which we removed the valley, is shown in the middle panel of Figure 7. Now the basement upwelling has about the right horizontal and vertical scales, but it is centered on the position of the middepth conductor. This is clearly seen by comparing it with the TE inversion for highly accurate data without the valley, shown in the lower panel. The individual TM data residuals are quite interesting. With the valley, they are dominated by the two sites that straddle the left end of the middepth conductor. Furthermore, the only unusual residuals at these sites are for $\ln(\rho_a)$ at periods longer than 50 s. The same is true for the model without a valley except now the anomalous sites are above the left step of the basement upwelling. It appears in each case that our functional has found a very smooth model which is in strong disagreement with only 2% of the data. This appears to be further evidence that we are not trapped in a local minimum of the object function. We think that what is actually happening is that the 2D structure at middepth and below has small TM response relative to the shallow structure. Two kinds of models have this property: the smooth ones of Figure 7 and thin, horizontal conductors of limited horizontal extent. The possible improvement in misfit of this second class of models (which in this case is the truth), while locally large, is simply too small overall to outweigh their much higher roughness as measured by our functional. It remains to be seen whether some modification of the penalty function can overcome this kind of problem.

CONCLUSION

We have presented a method of inverting multidimensional MT data which is orders of magnitude faster than standard methods that compute the full partial derivative of the data with respect to model elements. The inversion works well on noisy synthetic 2D data. It produces an image that in most cases appears as an out-of-focus view of the actual structure. The focus improves as the noise level decreases, and in some cases one can recover surprising detail. An important question for RRI and probably standard 2D minimum structure inversions is whether one can find an appropriate object function that can deal reasonably with structure that produces little TM response because it is vertically thin and of limited horizontal extent.

APPENDIX A: PSEUDO-PARTIAL DERIVATIVES FOR PSEUDO-TWO-DIMENSIONAL INVERSIONS ALONG SLICES OF A THREE-DIMENSIONAL MODEL

Following *Rodi* [1976] and *Madden* [1972], the elements of the partial derivative matrix $\mathbf{F}$ at each frequency for a standard 2D TM mode inversion are given by

$$\frac{\partial d_i}{\partial \sigma_j} = \frac{2}{\mathbf{u}_i^T \mathbf{x}_0}\left\{\frac{\partial \mathbf{u}_i^T}{\partial \sigma_j}\mathbf{x}_0 + \mathbf{u}_i^T \mathbf{A}^{-1}\left[\frac{\partial \mathbf{b}}{\partial \sigma_j} - \frac{\partial \mathbf{A}}{\partial \sigma_j}\mathbf{x}_0\right]\right\} - \frac{2}{\mathbf{v}_i^T \mathbf{x}_0}\left\{\frac{\partial \mathbf{v}_i^T}{\partial \sigma_j}\mathbf{x}_0 + \mathbf{v}_i^T \mathbf{A}^{-1}\left[\frac{\partial \mathbf{b}}{\partial \sigma_j} - \frac{\partial \mathbf{A}}{\partial \sigma_j}\mathbf{x}_0\right]\right\} \quad \text{(A1)}$$

where d_i is the complex datum (6′) at the site y_i, σ_j is the conductivity at the jth node and T denotes the matrix transpose. The vector $\mathbf{x}_0$ contains values of the magnetic field H_x at the nodes and is the solution to the TM forward problem $\mathbf{Ax} = \mathbf{b}$ for the starting

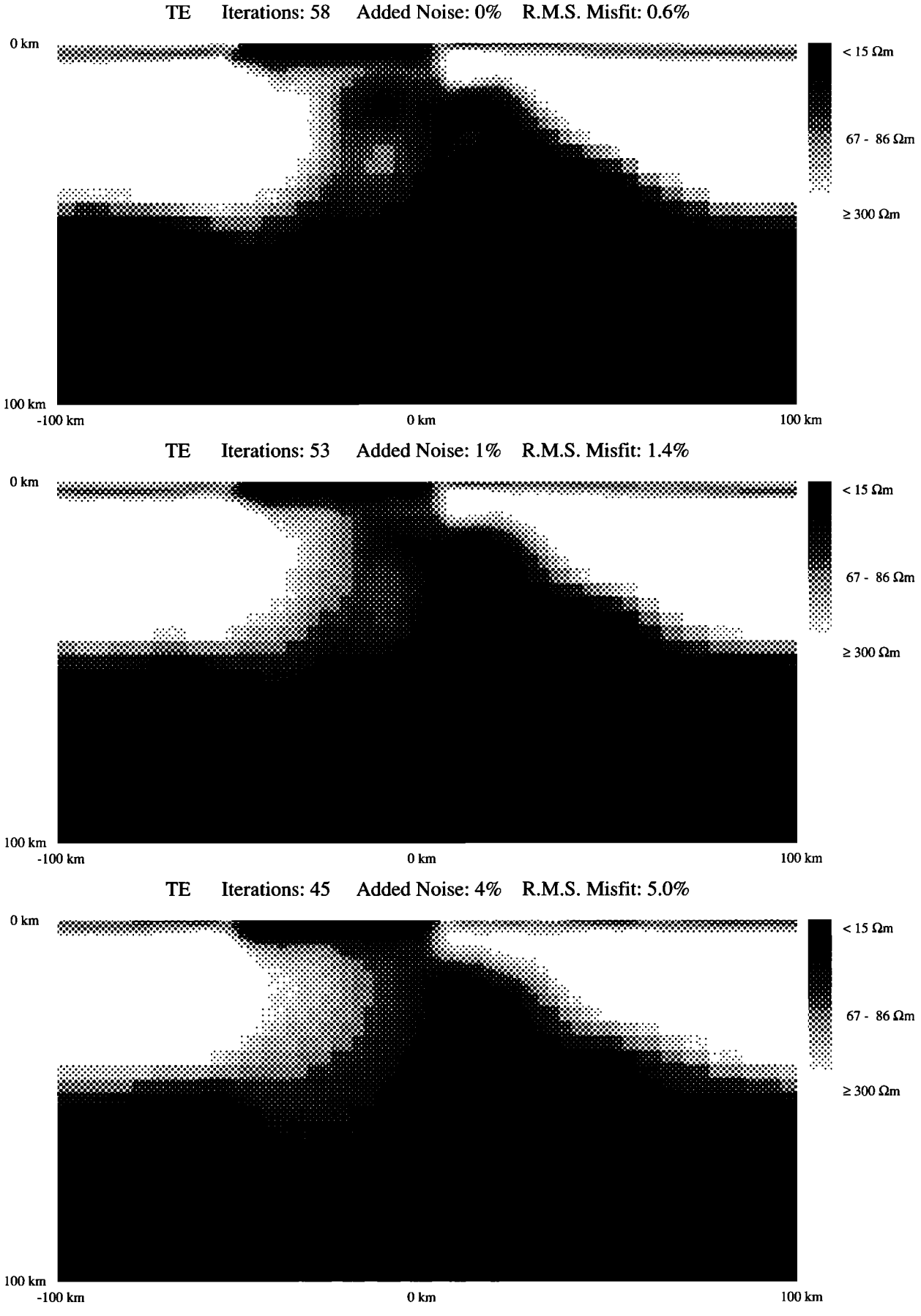

Fig. 6. Inversion of synthetic TE mode data for the Rift model with increasing added noise. There are 31 sites mostly spaced 4 km apart. Other inversions, with higher site densities in the vicinity of the ends of the valley and the midcrustal conductor, were essentially identical. The forward modeling grid has 72x40+8 nodes, is 500 km wide and extends to 100 km depth. The vertical node spacing is 500 m at the surface and 5 km at the bottom. The display uses boxes centered on the nodes.

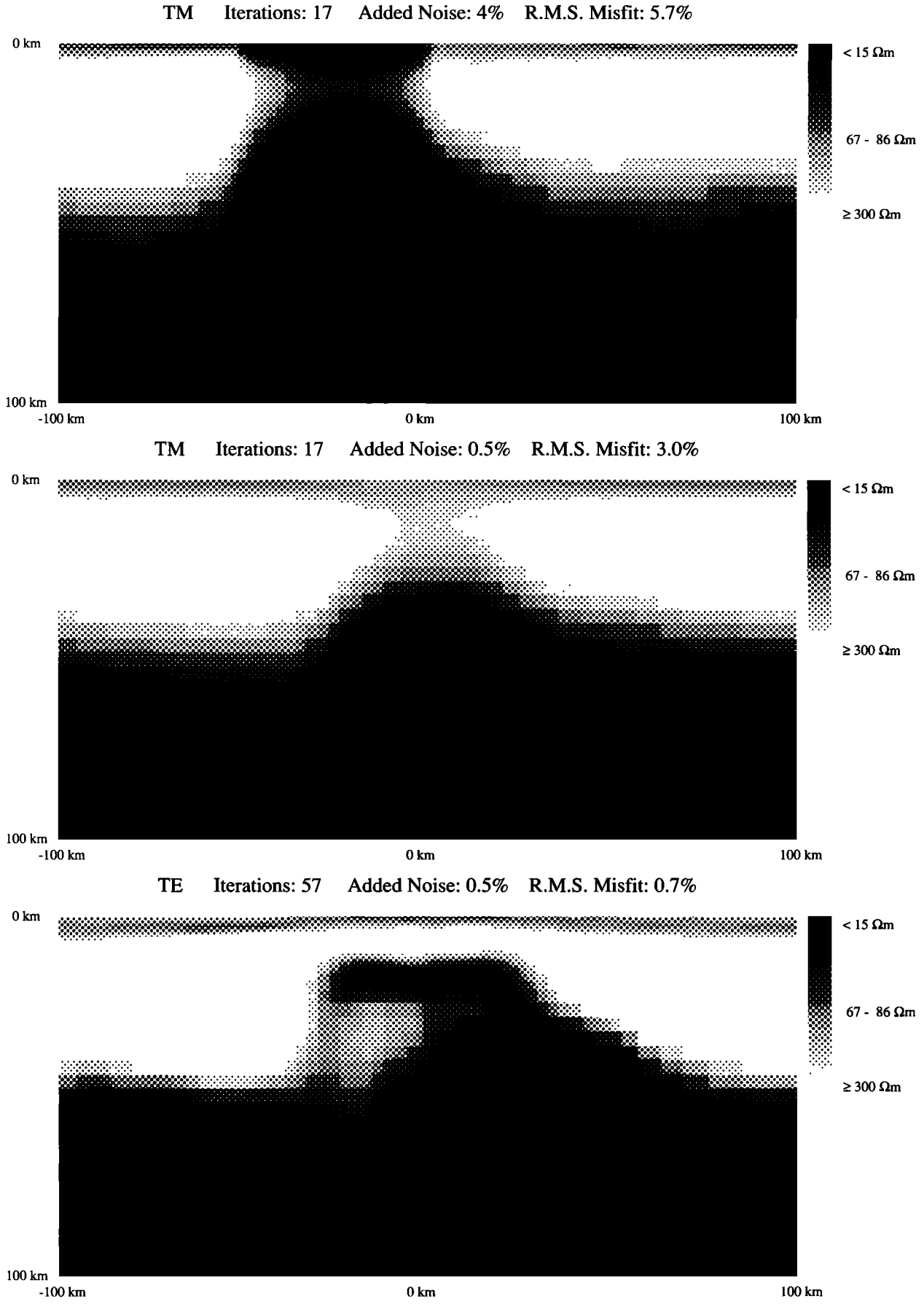

Fig. 7. (Top) Inversion of synthetic TM mode data for the Rift model with 4% noise. The sites are the same as the TE mode. Increasing sight density at the ends of the valley resulted in an essentially identical model. (Middle) Inversion of synthetic TM mode data for the Rift model with no valley and 0.5% added noise. There is one less site, but otherwise the grid is almost identical to the case with the valley. (Bottom) Inversion of synthetic TE mode data for the Rift model with no valley and 0.5% added noise. All the displays use boxes centered on the nodes.

model σ_0. The $\mathbf{v}_i$ and $\mathbf{u}_i$ are vectors such that $\mathbf{v}_i^T\mathbf{x}$ and $\mathbf{u}_i^T\mathbf{x}$ give H_x and E_y at the site y_i. For efficient computation of these derivatives when the number of nodes is large, the vector

$$\left[\frac{2\mathbf{u}_i^T}{\mathbf{u}_i^T\mathbf{x}_0} - \frac{2\mathbf{v}_i^T}{\mathbf{v}_i^T\mathbf{x}_0}\right]\mathbf{A}^{-1}$$

is computed only once for each site at each frequency.

To adapt (A1) for use in a pseudo 2D inversion along a slice of a 3D model, $\mathbf{x}$ is replaced by the vector of values of H_x computed at the nodes of the slice by solving the 3D forward problem; $\mathbf{A}$ and $\mathbf{b}$ are replaced by the matrix and source terms for solving the 2D TM forward problem for the slice; and $\mathbf{u}_i$ and $\mathbf{v}_i$ are unchanged.

APPENDIX B: STRUCTURE FUNCTIONAL

A fairly general form for a functional to minimize at a site y_i, which penalizes both horizontal and vertical structure for a model m is

$$Q(y_i) = \int\left[\frac{\partial^2 m(y_i,z)}{\partial f^2(z)} + g(z)\frac{\partial^2 m(y,z)}{\partial y^2}\bigg|_{y=y_i}\frac{\partial^2 z}{\partial f^2(z)}\right]^2 df(z)$$

This is a norm of a scaled version of the Laplacian beneath the site. The function $f(z)$ controls the scale length used for measuring structure at different depths, and $g(z)$ allows for varying the relative weight placed on horizontal structure. The dz^2/df^2 factor in the second term keeps the dimensions of the two terms the same. In our 1D work [*Smith and Booker*, 1988] we found $m = \ln(\sigma) = -\ln(\rho)$ and $f = \ln(z+z_0)$ to be particularly useful, because they tend to produce misfits that are fairly uniform across the frequency spectrum. They also guarantee a finite, nonnegative conductivity and apply a greater penalty on small-scale structure as depth increases. We have therefore used the same choices in this 2D work. The small constant z_0 keeps the criterion from diverging at the surface and is usually chosen to be the skin depth at the highest frequency in the surface material. Using the above choice for $f(z)$ and reexpressing the integral and derivatives in terms of linear depth, it is easy to show that a nearly equivalent functional is

$$Q = \int_0^{z_{max}} (z+z_0)^3\left[\frac{\partial^2 m}{\partial z^2} + g(z)\frac{\partial^2 m}{\partial y^2}\bigg|_{y=y_i}\right]^2 dz \qquad \text{(B1)}$$

The increase in penalty with depth inherent in choosing $f(z)=\ln(z+z_0)$ showns up explicitly in the $(z+z_0)^3$ factor. A suitably general choice of g is

$$g(z) = \alpha\left[\frac{\Delta_i}{z+z_0}\right]^\eta$$

where α and η are constants. As η increases from 0 to 3/2, the horizontal structure penalty changes from being the same as on vertical structure to being independent of depth. Δ_i is the distance between the sites flanking y_i and is introduced to keep the weight function nondimensional and to increase the penalty on horizontal structure where the sites are far apart. We have found that $\alpha = 4$ and $\eta = 3/2$ works well in most cases we have looked at.

Sampling the continuous model m at discrete points z_j, $j = 1$ to n_z, approximating the integral with a quadrature rule, using centered second differences to approximate the vertical derivatives, and using second differences and the values of the model at the adjacent sites from the previous iteration to approximate the horizontal derivatives, (B1) can be approximated by

$$Q = (\mathbf{Rm} - \mathbf{c})^T(\mathbf{Rm} - \mathbf{c}) \qquad \text{(B2)}$$

Where $\mathbf{R}$ is the so-called roughening matrix for the site; $\mathbf{m}$ is the sampled model; $\mathbf{c}$ is a bias vector expressing the fact that minimizing the horizontal derivative pulls the model toward its values at the adjacent sites; and the superscript T denotes the matrix transpose. If q_j are the coefficients in the quadrature rule, the elements of $\mathbf{R}$ and $\mathbf{c}$ at a site y_i are

$$R_{j\,j-1} = q_j^{1/2}(z_j+z_o)^{3/2}\frac{2}{(z_j-z_{j-1})(z_{j+1}-z_{j-1})}$$

$$R_{j\,j+1} = q_j^{1/2}(z_j+z_o)^{3/2}\frac{2}{(z_{j+1}-z_j)(z_{j+1}-z_{j-1})}$$

$$a_j = g(y_i,z_j)\,q_j^{1/2}(z_j+z_o)^{3/2}\frac{2}{(y_i-y_{i-1})(y_{i+1}-y_{i-1})}$$

$$b_j = g(y_i,z_j)\,q_j^{1/2}(z_j+z_o)^{3/2}\frac{2}{(y_{i+1}-y_i)(y_{i+1}-y_{i-1})}$$

$$R_{j\,j} = -R_{j\,j-1} - R_{j\,j+1} - a_j - b_j$$

$$c_j = -a_j m_j(y_{i-1}) - b_j m_j(y_{i+1})$$

for $j = 2$ to n_z-1. All other elements of $\mathbf{R}$ and $\mathbf{c}$ are zero. $\mathbf{R}$ is tridiagonal and its first and last rows are all zeros. This corresponds to the fact that the second derivative is not sensitive to the average value or average first derivative of $\mathbf{m}$. It also means that $\mathbf{R}$ is singular, a problem that is dealt with in Appendix C.

In practice, using the values of the model at adjacent sites from the previous iteration to approximate the horizontal derivative term in (B1) appears to over-estimate changes in the derivative at depth. This is clearly the case when changes to the model at adjacent sites are similar and large: the horizontal derivative is approximated as if the model at adjacent sites would remain unchanged. An effect of this is that for $\eta \approx 0$ so much weight is placed on horizontal smoothing at depth that the bottom of a model does not perceivably change from a 1D starting model. Choosing $\eta = 3/2$ seems to be a reasonable compromise between allowing the models to change at depth and placing a penalty on horizontal structure. One could avoid this difficulty for $\eta \approx 0$ by inverting for the structure beneath all stations simultaneously. It would be interesting to compare the resultant models to further evaluate the effect of this approximation. However, simultaneous inversion of the different sites would greatly increase the computation time, so we have not implemented this.

APPENDIX C: LINEARIZED INVERSION

Sampling the continuous model at discrete points z_j, $j = 1$ to n_z beneath a site y_i and approximating the integral with a quadrature rule, and replacing the infinitesimals δd and δm with finite differences, (5), (5′), (7) or (7′) can be split into n_d separate equations for real and imaginary parts and written as

$$(\mathbf{d} - \mathbf{e}) - \mathbf{d}_0 = \mathbf{Fm} - \mathbf{Fm}_0 \qquad \text{(C1)}$$

where $\mathbf{m}$ and $\mathbf{m}_0$ are n_z element vectors containing the new and starting models, $\mathbf{d}$ and $\mathbf{d}_0$ are n_d element vectors containing the measured and calculated data and $\mathbf{F}$ is an $n_z \times n_d$ pseudo-Frechet derivative matrix. The measured data are in error and we only want to fit them within some specified tolerance $\mathbf{e}^T\mathbf{e}$. We assume that the covariances of the measured data are known and have already been used to scale the data and pseudo-Frechet derivatives so that $\text{cov}(\mathbf{e}) = \mathbf{I}$, an identity matrix and $\mathbf{e}^T\mathbf{e}$ is the standard χ^2

misfit statistic. The equations (C1) are to be inverted separately at each site. Note, however, that (C1) would also hold for a standard 2D inversion with appropriately dimensioned matrices. Thus the efficient inversion of (C1) to be described in this appendix is equally applicable to a standard inversion.

We want to find the model satisfying (C1) which simultaneously minimizes a structure functional like the one described in Appendix B. We can do this by finding a constant β such that when $\mathbf{m}$ minimizes the object function

$$W = (\mathbf{Rm} - \mathbf{c})^T(\mathbf{Rm} - \mathbf{c}) + \beta\, \mathbf{e}^T\mathbf{e} \tag{C2}$$

the mean square misfit, $\mathbf{e}^T\mathbf{e}$, equals a predetermined goal (such as the expected value of χ^2).

If β is known a priori and we define

$$\tilde{\mathbf{d}} = \mathbf{d} - \mathbf{d}_0 + \mathbf{F}\mathbf{m}_0 \tag{C3}$$

the model minimizing (C2) when $\mathbf{c} = 0$ is

$$\mathbf{m} = \left(\frac{1}{\beta}\mathbf{R}^T\mathbf{R} + \mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\tilde{\mathbf{d}} \tag{C4}$$

Constable et al. [1987] use this result in a 1D minimum structure inversion. They choose β by forward modelling with a series of successive guesses to β. While this is not particularly onerous in 1D, because the forward problem can be solved very quickly, *deGroot-Hedlin and Constable* [1990] use the same approach for a minimum structure 2D standard inversion. They report needing 6 to 14 2D forward modelling steps to find β at each iteration of their algorithm. This is particularly time consuming because they fully factor each forward problem, rather than using an iterative forward method which could take advantage of the similarity of the fields in the models for the different trial β variables.

Our approach, based on the linearized equations (C1), results in a direct relation between $\mathbf{e}^T\mathbf{e}$ and β, which is much more efficient numerically and for 1D problems shows no tendency to be caught in the local minima reported by *Constable et al.* [1987]. It is described in detail by *Smith* [1988] and is a generalization of the case $\mathbf{R}$ nonsingular and $\mathbf{c} = 0$ by *Shure et al.* [1982].

It is first necessary to transform the roughening matrix $\mathbf{R}$ so that it is not singular. The elements of this transformed matrix $\tilde{\mathbf{R}}$ are identical to $\mathbf{R}$ except for the four elements

$$\tilde{R}_{1\,1} = \quad R_{2\,2}$$

$$\tilde{R}_{1\,2} = -R_{2\,2}$$

$$\tilde{R}_{n_z\,n_z} = R_{n_z-1\,n_z-1}$$

$$\tilde{R}_{n_z\,n_z-1} = -R_{n_z-1\,n_z-1}$$

The first and last elements of $\tilde{\mathbf{R}}\mathbf{m} - \mathbf{c}$ are a scaled finite difference approximation of the vertical derivative at the top and bottom of the model and the sum of the squares of the other elements gives the roughness.

Using (C1) and (C3) and the fact that $\tilde{\mathbf{R}}$ is not singular, we can write

$$\tilde{\mathbf{d}} = \mathbf{F}\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{R}}\mathbf{m} + \mathbf{e} \tag{C5}$$

Since $\tilde{\mathbf{R}}$ is still tridiagonal, finding $\mathbf{F}\tilde{\mathbf{R}}^{-1}$ involves a negligible amount of computation. The rest of the solution is simpler if we make the following partitions:

$$\mathbf{F}\tilde{\mathbf{R}}^{-1} = \begin{bmatrix}\mathbf{g}_1 & \mathbf{H} & \mathbf{g}_{n_z}\end{bmatrix} \qquad \tilde{\mathbf{R}}\mathbf{m} = \begin{bmatrix}p_1 \\ \tilde{\mathbf{m}} \\ p_{n_z}\end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix}c_1 & \tilde{\mathbf{c}} & c_{n_z}\end{bmatrix}$$

where $\mathbf{g}_1$ and $\mathbf{g}_{n_z}$ are the right and left columns of $\mathbf{F}\tilde{\mathbf{R}}^{-1}$ and p_1, p_{n_z}, c_1 and c_{n_z} are the first and last elements of $\tilde{\mathbf{R}}\mathbf{m}$ and $\mathbf{c}$ respectively. Defining $\mathbf{G} = [\mathbf{g}_1\ \mathbf{g}_{n_z}]$ and $\mathbf{p} = [p_1\ p_{n_z}]$, we can rewrite (C3) as

$$\tilde{\mathbf{d}} = \mathbf{H}\tilde{\mathbf{m}} + \mathbf{G}\mathbf{p} + \mathbf{e} \tag{C6}$$

The problem has now been stated in the form of the mixed inverse problem treated by *Pavlis and Booker* [1980, 1983]. The inverse problem for the sampled continuous model $\tilde{\mathbf{m}}$ is underconstrained, because n_{data} is usually less than n_z, while that for $\mathbf{p}$ would be overconstrained if $\tilde{\mathbf{m}}$ were known. It is worth noting that additional constraints or parameters are easily incorporated at this point. For instance, the apparent resistivity commonly has an unknown, frequency-independent multiplier because of static distortion of the surface electric field (see *Jones* [1988] for a clear discussion of the problem). This implies that the real part of the logarithmic data (6) or (6′) have an unknown offset at each site. We can incorporate simultaneous determination of this offset by redefining $\mathbf{p} = [p_1\ p_{n_z}\ s]$ and $\mathbf{G} = [\mathbf{g}_1\ \mathbf{g}_{n_z}\ \mathbf{i}]$, where s is the static shift constant and the elements of $\mathbf{i}$ are 1 or 0 depending on whether they correspond to an equation for the apparent resistivity or phase data.

The object function (C2) can now be minimized subject to the n_d constraints (C6) by introducing n_d Lagrange multipliers. After some algebra, this yields

$$\tilde{\mathbf{m}} = \tilde{\mathbf{c}} + \beta\mathbf{H}^T\mathbf{S}\mathbf{a} \tag{C7}$$

where

$$\mathbf{S} = (\beta\mathbf{H}\mathbf{H}^T + \mathbf{I})^{-1}$$

$$\mathbf{a} = \tilde{\mathbf{d}} - \mathbf{H}\tilde{\mathbf{c}} - \mathbf{G}\mathbf{p} \tag{C8}$$

Note that (C7) requires factoring (inverting) an $n_d \times n_d$ matrix while (C4) requires factoring an $n_z \times n_z$ matrix. Since the number of model parameters will almost always substantially exceed the number of data, (C7) will be much faster to calculate.

The scaled surface and bottom derivatives, $\mathbf{p}$ can be set to desired values or chosen to further minimize the object function (C2). In the latter case, (C7) and the definitions of $\mathbf{p}$ and $\tilde{\mathbf{m}}$ are used to eliminate $\mathbf{m}$ from (C2). Then taking the partials with respect to the elements of $\mathbf{p}$ and setting the results to zero yields

$$\mathbf{p} = \boldsymbol{\gamma}^T(\tilde{\mathbf{d}} - \mathbf{H}\tilde{\mathbf{c}})$$

where

$$\boldsymbol{\gamma}^T = (\mathbf{G}^T\mathbf{S}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{S} \tag{C9}$$

The new matrix to be factored is only 2×2 (or 3×3 if the static shift constant is included). The results in this paper set $\mathbf{p} = 0$ to allow more accurate forward modelling of the electric and magnetic fields. Note that this does not mean that there are no horizontal gradients of $\mathbf{m}$ at the top and bottom of the model.

Equations (C7), (C8) and (C5) and the definitions of $\mathbf{p}$ and $\tilde{\mathbf{m}}$ are used in (C1) to obtain

$$\mathbf{e}^T\mathbf{e} = (\mathbf{S}\mathbf{a})^T\mathbf{S}\mathbf{a} \tag{C10}$$

Then using the definition of $\mathbf{S}$ and the singular value decomposition (SVD) $\mathbf{H} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal (unitary)

matrices and $\mathbf{\Lambda}$ is the nonnegative, diagonal matrix of the singular values λ_k of $\mathbf{H}$, we can write

$$\mathbf{e}^T\mathbf{e} = \sum_{k=1}^{n_d} \frac{\tilde{a}_k^2}{(\beta\lambda_k + 1)^2} \tag{C11}$$

where the $\tilde{a}_k$ are the elements of $\tilde{\mathbf{a}} = \mathbf{U}^T\mathbf{a}$. Note that $\mathbf{e}^T\mathbf{e}$ and its derivatives are all smooth monotonic functions of β for β positive, so that for a given choice of $\mathbf{e}^T\mathbf{e}$, (C11) can be quickly solved for β using Newton's method. Note also that $(\mathbf{e}^T\mathbf{e})_{\min} = \sum \tilde{a}_k^2$ for k such that $\lambda_k = 0$. It is important to check that the target value for $\mathbf{e}^T\mathbf{e}$ in (C2) exceeds this minimum. Of course, machine precision usually prevents any of the singular values from being exactly zero, and it is necessary to establish a threshold below which the λ_k are considered zero. There is little harm in making this threshold a little high, because the linearization inherent in (C1) usually breaks down for target values of $\mathbf{e}^T\mathbf{e}$ which are much larger than the minimum.

Once the SVD of $\mathbf{H}$ has been obtained, the matrix

$$\mathbf{S} = (\beta\mathbf{H}\mathbf{H}^T + \mathbf{I})^{-1} = \mathbf{U}(\beta\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{I})^{-1}\mathbf{U}^T$$

used in (C7), (C9) and (C10) is trivially calculated. Finally, the actual model can be retrieved from the roughened model using

$$\mathbf{m} = \tilde{\mathbf{R}}^{-1}\begin{bmatrix} p_1 \\ \tilde{\mathbf{m}} \\ p_{n_z} \end{bmatrix}$$

Acknowledgments. This research was supported by the Office of Basic Energy Sciences of U.S. Department of Energy through grants DE-FG06-86ER13472 and DE-FG06-89ER14064 and by the Earth Sciences Division of the U.S. National Science Foundation though grant EAR-8817107. A Chateaubriand Fellowship from the French Government to one of the authors (J.T.S.) allowed for additional refinement of the algorithm. Comments from Ted Madden and two anonymous reviewers were helpful in revising the paper. Nong Wu and Jeffery Knickerbocker were very helpful in displaying the results and in finding numerous logical errors in the implementation.

REFERENCES

Behie, A., and P. K. W. Vinsome, Block iterative methods for fully implicit reservoir simulation, *Soc. Pet. Eng. J.*, *22*, 658-668, 1982.

Brandt, A., Multi-level adaptive solutions to boundary-value problems, *Math. Comput.*, *138*, 333-390, 1977.

Brewitt-Taylor, C. R. and J. T. Weaver, On the finite difference solution of two-dimensional induction problems, *Geophys. J. R. Astron. Soc.*, *47*, 375-396, 1978.

Constable, S. C., R. L. Parker, and C. G. Constable, Occam's inversion: A practical algorithm for generating smooth models from EM sounding data, *Geophysics*, *92*, 289-300, 1987.

deGroot-Hedlin, C., and S. C. Constable, Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data, *Geophysics*, *55*, 1613-1624, 1990.

Jones, A. G., Static shift of magnetotelluric data and its removal in a sedimentary basin environment, *Geophysics*, *53*, 967-978, 1988.

Jupp, D. L. B., and K. Vozoff, Two-dimensional magnetotelluric inversion, *Geophys. J. R. Astron. Soc.*, *50*, 333-352, 1977.

Kershaw, D. S., The incomplete Cholesky-conjugate gradient method for iterative solution of systems of linear equations, *J. Comput. Phys.*, *26*, 43-65, 1978.

Madden, T. R. Transmission systems and network analogies to geophysical forward and inverse problems, *Tech. Rep. N000-14-67-A-0204-0045, M.I.T. Rep. 72-3*, Dep. of Earth and Planet. Sci., Mass. Inst. of Technol., Cambridge, 1972.

Madden, T. R., and R. L. Mackie, Three-dimensional magnetotelluric modeling and inversion. *Proc. IEEE*, *77*, 318-333, 1989.

Oldenburg, D. W., One dimensional inversion of natural source magnetotelluric observations. *Geophysics*, *44*, 1218-1244, 1978.

Oldenburg, D. W., K. P. Whitall, and R. L. Parker, Inversion of ocean bottom magnetotelluric data revisited. *J. Geophys. Res.*, *89*, 1829-1833, 1984.

Parker, R. L., The inverse problem of electromagnetic induction: Existence and construction of solutions based upon incomplete data. *J. Geophys. Res.*, *85*, 4421-4428, 1980.

Pavlis, G. L., and J. R. Booker, The mixed discrete-continuous inverse problem: Application to determination of earthquake hypocenters and velocity structure, *J. Geophys. Res.*, *85*, 4801-4810, 1980.

Pavlis, G. L., and J. R. Booker, Progressive multiple event location (PMEL), *Bull. Seismol. Soc. Am.*, *73*, 1753-1777, 1983.

Rodi, W. L., A technique for improving the accuracy of finite element solutions for magnetotelluric data, *Geophys. J. R. Astron. Soc.*, *44*, 483-506, 1976.

Shure, L., R. L. Parker, and G. E. Backus, Harmonic splines for geomagnetic modelling, *Phys. Earth Planet. Inter.*, *28*, 215-229, 1982.

Smith, J. T., Rapid inversion of multi-dimensional magnetotelluric data, Ph.D. dissertation, Univ. of Wash., Seattle, 1988.

Smith, J. T., and J. R. Booker, Magnetotelluric inversion for minimum structure, *Geophysics*, *53*, 1565-1576, 1988.

Wannamaker, P. E., G. W. Hohman, and S. H. Ward, Magnetotelluric responses of three-dimensional bodies in layered earths, *Geophysics*, *49*, 1517-1534, 1984.

Wannamaker, P. E., J. A. Stodt, and L. Rijo, Two-dimensional topographic responses in magnetotellurics modeled using finite elements, *Geophysics*, *51*, 2131-2144, 1986.

Weidelt, P., Inversion of two-dimensional conductivity structures, *Phys. Earth Planet. Inter.*, *10*, 282-291, 1975.

J. R. Booker and J. T. Smith, Geophysics Program, AK-50, University of Washington, Seattle, WA 98195.

(Received June 19, 1990;
revised October 15, 1990;
accepted November 2, 1990.)