Efficient stochastic estimation of the model resolution matrix diagonal and generalized cross–validation for large geophysical inverse problems

J. K. MacCarthy,[1] B. Borchers,[2] R. C. Aster,[3]

R. C. Aster, Department of Earth and Environmental Science and Geophysical Research Center, New Mexico Institute of Mining and Technology, 801 Leroy Place, Socorro, NM 87801, USA. (aster@ees.nmt.edu)

B. Borchers, Department of Mathematics, New Mexico Institute of Mining and Technology, 801 Leroy Place, Socorro, NM 87801, USA. (borchers@nmt.edu)

J. K. MacCarthy, Geophysics Group, Los Alamos National Laboratory, EES-17 M.S. D408, Los Alamos, NM 87545, USA. (jkmacc@lanl.gov)

[1]Geophysics Group, Los Alamos National

Abstract. In recent years, larger geophysical datasets and novel model parameterizations have dramatically increased both the data and model space dimensions of many inverse problems. Because of their relatively low computational expense, trade–off curve corner estimation for choosing regularized models and "checkerboard" tests for evaluating model resolution are commonly applied, despite their limitations. We present and demonstrate a low–cost method for accurately estimating the diagonal elements of the model resolution matrix diagonal and for implementing generalized cross–validation (GCV) for optimal regularization parameter selection. The ability to estimate the diagonal of the resolution matrix and GCV function thus facilitates the introduction of additional tools for diagonal resolution analysis and regularization evaluation, even for very large inverse problems, with storage and computational costs comparable to those required for obtaining model so-

Laboratory, Los Alamos, NM, USA.

[2]Department of Mathematics, New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA.

[3]Department of Earth and Environmental Science and Geophysical Research Center, New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA.

lutions. We demonstrate the method using a Tikhonov regularized teleseismic body wave velocity inversion example with approximately 260,000 model parameters, where we validate randomly selected $\mathbf{R_m}$ diagonal elements against explicitly calculated values and compare GCV-estimated regularized model results to those obtained through traditional methods.

1. Introduction

Recent expansion of seismic data availability and innovations in model parameterization motivate the need for computationally tractable, unbiased, and easy to implement resolution estimators. In seismology, for example, continent-scale seismic networks, such as EarthScope USArray Transportable Array and increasingly large IRIS PASSCAL and other deployments, along with increasingly large global inversions are dramatically improving the resolution of tomographic studies of the crust, mantle, and whole Earth. Novel innovations in forward modeling and model parameterization are also emerging, such as using adaptive grids [*Li et al.*, 2008], spherical wavelets [*Chiao and Kuo*, 2001], and finite-frequency kernels [*Marquering et al.*, 1999; *Dahlen et al.*, 2000].

Regularized linear inversions are central to geophysics, due in part to their favorable statistical characteristics [*Berryman*, 2000; *Aster et al.*, 2005], the availability of efficient iterative solvers for large systems, such as LSQR [*Paige and Saunders*, 1982], and the commonly ill–posed nature of inverse problems. Even as the size and complexity of linear or linearized inverse problems grows, iterative solvers are able to produce solutions efficiently. Analyzing the balance between model resolution and regularization, however, becomes considerably more computationally intensive than producing solutions.

For linear systems of equations that are sufficiently small to perform a singular value decomposition (SVD) of the forward operator matrix, resolution, a fundamental measure of solution bias, is quantified by the elements of the model resolution matrix. For larger problems, however, it can easily become memory and CPU prohibitive to estimate solution bias in this way. Consequently, it is a common practice to employ resolution spike, checkerboard, or similar tests using synthetic data generated from canonical test models to estimate the effects of imperfect model parameter resolution. Such tests are efficient in that they only require equivalent effort to that necessary for inverting real data. However they can only recover an approximation to a single column of the resolution matrix, or a specified linear combination of such columns, and may thus provide ambiguous and/or incomplete model resolution characterizations under some circumstances.

The choice of regularization parameters affect solution resolution, which generally degrades as regularization constraints, such as solution bounds or smoothness, are added. An optimal degree of regularization is commonly estimated through the use of trade–off curves between a model norm (or seminorm) and the forward modeled misfit with observed data [*Hansen and O'Leary*, 1993]. When the statistical character of the data noise is unknown or only roughly estimated, as is commonly the case, this choice can be rather arbitrary. Generalized cross–validation (GCV) provides a well–characterized method of selecting a regularization parameter that minimizes the predictive data errors in a least squares solution [*Craven and Wahba*, 1979; *Golub et al.*, 1979]. It is a useful selection criterion in cases where the variance of the data noise is unknown and data errors are uncorrelated [*Wahba*, 1990; *Golub and vonMatt*, 1997], or when a trade–off curve is poorly defined, either through lack of smoothness or poor sampling [*Hansen and O'Leary*, 1993]. However, GCV requires calculating the trace of a large matrix, which, when approached straightforwardly, is commonly computationally prohibitive for large inverse problems.

Recent work by *Bekas et al.* [2007] on the statistical estimation of the large matrix diagonals provides a notable new tool to facilitate both resolution analysis and implementation of GCV for large geophysical inversions. Here, we illustrate the application of this stochastic method to produce unbiased and accurate estimates of the GCV function and the diagonal elements of the model resolution matrix, apply this method to a moderately large teleseismic tomographic inverse problem, and provide associated self-contained MATLAB functions (supplementary materials).

2. Resolution and regularization

Here we define the model resolution matrix for a Tikhonov regularized linear forward problem of the form

$$\mathbf{Gm} = \mathbf{d}, \tag{1}$$

where $\mathbf{G}$ is the forward operator matrix, $\mathbf{m}$ is an n–dimensional model vector, and $\mathbf{d}$ is an m–dimensional data vector. Each constraint equation in this system is assumed to be weighted by an estimate of the respective data error standard deviation.

Because many geophysical inverse problems are ill–conditioned and/or rank deficient, additional constraints are typically needed for solution stability and uniqueness e.g., [*Menke*, 1989; *Parker*, 1994; *Aster et al.*, 2005]. We implement regularization here by incorporating a roughening matrix, $\mathbf{L}$, and its associated weighting parameter, α, into the inverse problem corresponding to (1). The resulting Tikhonov regularized least squares problem is

$$\min \left\| \begin{bmatrix} \mathbf{G} \\ \alpha\mathbf{L} \end{bmatrix} \mathbf{m} - \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix} \right\|_2 . \tag{2}$$

It can be shown using the normal equations that the least squares solution to (2) can be expressed by a linear matrix inverse operator acting on the data vector

$$\mathbf{m}_\alpha = \mathbf{G}^\sharp \mathbf{d} \ , \tag{3}$$

where

$$\mathbf{G}^\sharp = (\mathbf{G}^T\mathbf{G} + \alpha^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T \tag{4}$$

[*Aster et al.*, 2005]. The model resolution matrix characterizes the linear model space mapping between a (typically unknown) true model and that recovered using (3), i.e., for some true model $\hat{\mathbf{m}}$ with noise-free associated data $\hat{\mathbf{d}}$,

$$\mathbf{m}_\alpha = \mathbf{G}^\sharp\hat{\mathbf{d}} = \mathbf{G}^\sharp\mathbf{G}\hat{\mathbf{m}} = \mathbf{R_m}\hat{\mathbf{m}}\ . \tag{5}$$

$\mathbf{R_m}(\alpha) = \mathbf{G}^\sharp\mathbf{G}$ is an n by n square matrix that characterizes the model bias inherent in the regularized inversion. Columns of $\mathbf{R_m}$ are resolution kernels corresponding to point spread (i.e. spike test) functions for each model parameter. Off-diagonal entries represent smearing/trade–off between parameters in the recovered solution, and diagonal entries characterize the independent resolvability of each parameter. The closer $\mathbf{R_m}$ is to the identity matrix, the less bias inherent in the inversion, and the higher the fidelity of the solution will be to the unknown true model that generated the observed data.

3. Motivation for and implementation of stochastic estimation of a matrix diagonal

A significant practical difficulty in calculating $\mathbf{R_m}$ directly is that, although $\mathbf{G}$ may be sparse (as in a typical seismic tomography problem), $(\mathbf{G}^T\mathbf{G}+\alpha^2\mathbf{L}^T\mathbf{L})^{-1}$ in (4) is typically an n by n dense matrix. For problems with n larger than a few tens of thousands of parameters, this can require in excess of many tens of gigabytes of storage and prohibitively time consuming calculations.

Because of the central importance of this problem for large linear or linearized inverse problems, a number of methods have been proposed to estimate or calculate the full resolution matrix (5). Approaches include iterative methods that complement the LSQR algorithm [*Zhang and McMechan*, 1995; *Yao et al.*, 1999; *Zhang and Thurber*, 2007]. These methods, while taking advantage of the computational efficiencies of the LSQR algorithm, produce an "effective resolution matrix," that may not fully represent the model resolution [*Deal and Nolet*, 1996; *Berryman*, 2000; *Zhang and Thurber*, 2007]. *Nolet et al.* [1999] formulated an explicit expression for an approximation to the resolution matrix using a one-step back–projection method. This method, however, makes special assumptions about the structure of the forward operator. Finally, a highly computationally intensive class of methods exploits Choleski factorization and parallel computation to evaluate model resolution [*Boschi*, 2003].

Both the least squares solution and the model resolution in (3) and (5) are dependent on the choice of regularization roughening matrix $\mathbf{L}$ and its weighting parameter, α. Generalized cross–validation (GCV) selects the regularization parameter that minimizes the predictive error for all data points when left out one at a time. This is done by minimizing the GCV function, $V_0(\alpha)$,

$$V_0(\alpha) \approx \frac{m\,||\mathbf{G}\mathbf{m}_\alpha - \mathbf{d}||_2^2}{\mathrm{Tr}(\mathbf{I} - \mathbf{G}\mathbf{G}^\sharp)^2}, \tag{6}$$

where Tr denotes the matrix trace and m is the data space dimension [*Craven and Wahba*, 1979]. Implicit in (6) is the approximation that matrix diagonals $(\mathbf{G}\mathbf{G}^\sharp)_{k,k} \approx \mathrm{Tr}(\mathbf{G}\mathbf{G}^\sharp)/m$, which is shown by *Golub et al.* [1979] to be reasonable for large m. It is favorable to use GCV to choose $\mathbf{m}_\alpha$ because, making certain assumptions about the smoothness and noise of the true model, $\hat{\mathbf{m}}$, it can be shown that $E[\|\hat{\mathbf{m}} - \mathbf{m}_\alpha\|_2]$ goes to 0 as m goes to infinity, for an $\mathbf{m}_\alpha$ chosen through GCV [*Craven and Wahba*, 1979; *Wahba*, 1990]. *Golub and vonMatt* [1997] applied a stochastic trace estimator to estimate (6), but did so by calculating upper and lower bounds through a more complex method than that presented here. The stochastic matrix diagonal estimator presented here is independent of the number of iterations used to find the model solution and makes no assumptions of the structure of the forward operator.

The following stochastic algorithm comes largely from *Bekas et al.* [2007], who initially applied it to atomic density functional theory and noted its broad relevance, and is in turn based upon work by *Hutchinson* [1990] and *Girard* [1987]. Here, we apply the matrix diagonal estimator to the resolution matrix (5) and the calculation of the GCV function (6).

Consider a sequence of s n-length random vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_s$, with independent elements drawn from a standard normal distribution. The s^{th} estimate for the diagonal of an n by n square matrix $\mathbf{A}$ is then

$$\mathbf{D}_s = \left[\sum_{k=1}^{s} \mathbf{v}_k \odot \mathbf{A}\mathbf{v}_k \right] \oslash \left[\sum_{k=1}^{s} \mathbf{v}_k \odot \mathbf{v}_k \right], \tag{7}$$

where $\odot$ signifies element-wise vector multiplication and $\oslash$ signifies element-wise vector division. The algorithm corresponding to (7) is the following:

Stochastic matrix diagonal estimator

1. $\mathbf{t}_0, \mathbf{q}_0 = 0$
2. **for** $k = 1 \ldots s$

   (i) Generate a random vector realization $\mathbf{v}_k$

   (ii) $\mathbf{t}_k = \mathbf{t}_{k-1} + (\mathbf{A}\mathbf{v}_k \odot \mathbf{v}_k)$

   (iii) $\mathbf{q}_k = \mathbf{q}_{k-1} + (\mathbf{v}_k \odot \mathbf{v}_k)$

   (iv) $\mathbf{D}_k = \mathbf{t}_k \oslash \mathbf{q}_k$
3. **end**

In practice, the choice of s will depend on the desired accuracy of the diagonal determination, which can be assessed by statistically examining repeated estimates generated with independent random vectors and by the convergence of the estimates $\mathbf{D}_s$. Equation (7) contains the matrix-vector product $\mathbf{A}\mathbf{v}_k$, which cannot be evaluated directly if $\mathbf{A}$ is incalculable. When $\mathbf{A}$ is the resolution matrix, $\mathbf{R_m}$, this product can be computed by noting that a product $\mathbf{y} = \mathbf{R_m}\mathbf{v}_k$ can be rewritten in terms of the known matrices $\mathbf{G}$ and $\mathbf{L}$ by combining (5) and (4) as

$$\mathbf{y} = (\mathbf{G}^T\mathbf{G} + \alpha^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T\mathbf{G}\mathbf{v}_k \ , \tag{8}$$

which is the normal equations solution for

$$\min \left\| \begin{bmatrix} \mathbf{G} \\ \alpha\mathbf{L} \end{bmatrix} \mathbf{y} - \begin{bmatrix} \mathbf{G}\mathbf{v}_k \\ \mathbf{0} \end{bmatrix} \right\|_2 . \tag{9}$$

In estimating the GCV function (6), let $\mathbf{A}$ be $\mathbf{G}\mathbf{G}^\sharp$. We first evaluate the product $\mathbf{y} = \mathbf{G}^\sharp\mathbf{v_k}$ as

$$\mathbf{y} = (\mathbf{G}^T\mathbf{G} + \alpha^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T\mathbf{v}_k \ , \tag{10}$$

which is the normal equations solution for

$$\min \left\| \begin{bmatrix} \mathbf{G} \\ \alpha\mathbf{L} \end{bmatrix} \mathbf{y} - \begin{bmatrix} \mathbf{v}_k \\ \mathbf{0} \end{bmatrix} \right\|_2 . \tag{11}$$

The least squares solution to (11) is subsequently left–multiplied by $\mathbf{G}$ to obtain the desired matrix–vector product $\mathbf{G}\mathbf{G}^\sharp\mathbf{v_k}$ in (7). Once the diagonal of $\mathbf{G}\mathbf{G}^\sharp$, and hence its trace, are estimated, calculating (6) is trivial. Both (9) and (11) can be readily solved with an iterative solver such as LSQR.

The computational cost of using this algorithm to minimize the GCV function in terms of the number of LSQR calls required, is $s \cdot p$, where p is the number of regularization weighting parameters tested. Estimating the resolution matrix diagonal requires only s calls to LSQR.

4. An example from teleseismic tomography

We apply the method to select the regularization parameter and estimate the resolution matrix diagonal for a moderately large seismic tomographic inversion. The CREST (Colorado Rockies Experiment and Seismic Transects; [*Aster et al.*, 2009; *MacCarthy*, 2010]) teleseismic inversion data subset examined here consists of 19,608 mean-removed teleseismic P-wave travel time residuals and estimated data errors, measured at 167 broadband seismic stations in the region [*MacCarthy*, 2010] (Figure 1). The model space is parameterized by 267,520 constant slowness blocks, each 0.25°by 0.25°by 25 km in size. The forward problem matrix was constructed via infinite frequency raytracing through a one–dimensional reference velocity model (ak135; [*Kennett et al.*, 1995]) with crustal corrections, and solutions are expressed as percent velocity or slowness variation from this model.

Forward problem constraint equations were scaled by respective standard deviations estimated from ensemble P arrival waveform crosscorrelation (using approximately one principal period of the first arrival) across the network [*VanDecar and Crossen*, 1990]. Analysis of data errors suggested that the crosscorrelation methodology underestimates the true measurement errors. We note that other authors have reached similar conclusions, suggesting that a factor of 2–10 typically brings crosscorrelation–derived error estimates in teleseismic inversion data sets closer to those estimated by data analysts [*Waite et al.*, 2006; *Pavlis and Vernon*, 2010]. As error amplification factor increases, error values that are divided into rows of $\mathbf{G}$ and $\mathbf{d}$ will reduce the weight of the data equations relative to the regularization equations for a given α (Equation 2), thus producing smoother solutions. At the same time, the value of $||\mathbf{Gm} - \mathbf{d}||$ decreases with increasing error amplification for the same α, thus bring both branches of the regularization l-curve (Figure 2a) towards zero while maintaining shape and relative data variance reduction. We find that scaling crosscorrelation–determined error estimates by a factor of 4, producing a root mean square estimated error of 0.148, brings the model seminorm versus residual trade–off curve corner and GCV minimum into consistency with the noise level, per the discrepancy principle describing statistically expected data fit [*Hansen and O'Leary*, 1993; *Aster et al.*, 2005] and have adopted this scaling factor in further work with this data set.

Like most geophysical tomographic inversions, this example is rank–deficient. We thus regularize the inversion using superimposed zeroth–order and second–order (Laplacian) smoothing in equal proportion, scaled by the regularization parameter α, and by a constant level of edge–damping [*MacCarthy*, 2010]. Second–order smoothing is used in order to discourage spurious features in the resulting models, and zeroth–order damping is employed to minimize model amplitudes and to aid in convergence. We examine the selection of the regularization parameter using trade–off curves and via GCV, and use the different recovered models to demonstrate the use of the diagonal resolution estimation algorithm in solution bias characterization.

In trade–off curve analysis, α was selected visually from the corner vicinity of the plot of data residual versus model seminorm (Figure 2a). The corner provides a heuristic for estimating an optimal degree of regularization, but its character will be influenced by the plotting range and scale (e.g., linear, linear-log, or log-log plotting are variously used in practice). It is common for preferred models in such studies to be somewhat over–

regularized relative to the mathematically "best" solution in the interest of producing stable, conservative, or geologically reasonable models. We show a model that is slightly towards the smoother side of a linear-linear trade–off curve, corresponding to $\alpha = 0.7$ (Figure 3a–c). This particular model has maximum amplitudes of $\pm 4.5\%$ in V_p and corresponds to a data variance reduction of 78.7% (a root–mean–square data fit of 89%) compared to ak135.

We next determined α to minimize the GCV function (6). The GCV–optimal α for the CREST inversion, selected from its broad minimum, is near 0.1 (Figure 2b, 3d–f). While structurally similar to the model with $\alpha = 0.7$, maximum amplitudes in this model are $\pm 6.8\%$, with a data variance reduction of 91.7%. Note that these high amplitude P-wave variations are believed to be petrologically infeasible, and the high roughness (large seminorm) of the GCV–optimal model likely indicates that this particular solution is unduly rough. This is likely due in part to the flat and broad minimum region in the GCV curve, and/or the presence of correlated data errors [*Wahba*, 1990; *Hansen and O'Leary*, 1993]. Insights into the inverse problem obtained through GCV, such as these, may not otherwise be obtained through traditional regularization methods.

We show both a checkerboard resolution test and estimated model resolution diagonals for the two example regularized solutions discussed above to illustrate the effect of regularization weighting on resolution and to highlight how the two methods of resolution analysis offer different insights. Alternating 3^3-block clusters of $\pm 2\%$ Vp were used to generate synthetic travel time data using the CREST forward problem, and the data were contaminated with noise at the same level as that estimated for the CREST data. The synthetic data were then inverted using the same $\alpha = 0.1$ and 0.7 inversions as previously discussed. The resulting checkerboard recovery models are a rough approximation of a spatial distribution of superimposed respective resolution kernels within the model space (Figures 3b, 3f). The tests highlight regions with high shape and amplitude recovery, versus poorly constrained regions dominated by smearing. A significant shortcoming of this approach, however, is that interpreting amplitude recovery for a given parameter is complicated by smearing/superposition from adjacent parameters. For example, maximum amplitude recovery for the $\alpha = 0.1$ and 0.7 solutions is greater than the input amplitude for both checkerboard inversions. Because of this effect, the recovered models for both inversions look very similar and quantitative distinctions of amplitude recovery between different inversions is difficult. The model resolution matrix diagonal is a more quantitative measure of amplitude recovery that is independent of the geometry of synthetic input models.

The stochastic method of Section 3 was used to estimate the model resolution matrix diagonal for the two regularized inversions, using $s = 256$ random vectors. Stable values were obtained by running $N = 20$ realizations of the diagonal estimation and calculating median values. A random subset of 100 elements were validated against explicitly calculated elements for each of the N estimations. Figures 4a and 4c compare median stochastic estimates of R_m diagonal elements versus their true values, for $\alpha = 0.7$ and α=0.1, respectively. Symmetric sample standard deviations for $N = 20$ realizations are shown as error bars. In all cases, true values are within the one standard deviation of the median estimated value. Figures 4b and 4d depict the frequency of absolute errors in median estimated R_m diagonal elements. The mean and maximum absolute errors of the median estimates was 0.005 and 0.024 for the $\alpha = 0.1$ inversion, and 0.002 and 0.011 for the $\alpha = 0.7$ inversion.

To further illustrate the accuracy of the stochastic method, the diagonal elements of the resolution matrix for a synthetic tomographic problem were estimated and compared to the explicitly calculated values. The problem consisted of an $8 \times 8 \times 8 = 512$ element Cartesian block model of known slowness, through which straight rays were traced. The problem was regularized using smoothing and damping in equal proportion, with $\alpha = 0.5$ (Equation 2). Resolution matrix diagonal elements were estimated using the stochastic method, with $N = 20$ and $s = 256$, and median values were compared to those from the formal resolution matrix, $\mathbf{R_m}(\alpha) = \mathbf{G}^\sharp\mathbf{G}$ (Figure 5). As in the larger example, median values are within one sample standard deviation from the true value. Mean and maximum absolute errors are 0.0003 and 0.022, respectively.

Selection of appropriate values for s and N will vary from problem to problem. Estimated elements across N realizations are derived from independently generated pseudorandom numbers. Estimated elements also appear to be approximately normally distributed, with a mean about the true value. Thus, under the assumptions of independence and normality, the mean value of the N estimates converges to the true value at a rate proportional to $\sqrt{N}$, or $O(1/\sqrt{N})$. Under these assumptions, one can select N such that any estimated parameter's standard error is below some threshold, δ. First, choose a small number of realizations, N_1, compute the sample standard deviation for each diagonal element, s_N, and find the maximum value, s_N^{max}. One can now select a larger number realizations, N_2, such that $s_N^{max}/\sqrt{N_2} < \delta$. The mean of each estimated diagonal element over N_2 realizations will then be less than δ from the true value. Selection of the number of random vectors, s, is more complicated, as a mathematical description of estimate convergence with increasing s is not well characterized. In their application of the stochastic trace estimater, *Bekas et al.* [2007] noted the very few vectors are required to produce somewhat accurate estimates, with steady but slow convergence thereafter. Due to the speed of the calculation, however, we recommend that an s of 256–512 will be adequate for many large geophysical inversions.

While the pattern of well-resolved regions is similar between the two CREST inversions, the amplitude bias due to regularization is notably different (Figures 3c, 3g). The resolution diagonal in the $\alpha = 0.7$ model is nearly half that of the $\alpha = 0.1$ model, with maximum $\mathbf{R_m}$ diagonal values of 0.375 and 0.618 respectively. This implies a much larger degree of smoothing inherent in the $\alpha = 0.7$ inversion that is not apparent through the corresponding traditional multiblock checkerboard analysis. A drawback of looking only at the $\mathbf{R_m}$ diagonal, of course, is not being able to visualize smearing bias in the inversion.

It has been suggested that ray–sampling density is a low–cost qualitative tool to evaluate spatial model resolution in tomographic inversions [e.g. *Zhang and Thurber*, 2007], as more highly sampled parameters tend to exhibit higher resolution. This formulation, however, does not take into account the angular sampling of rays as they traverse model parameters or the regularization employed in the inversion, both of which contribute to parameter resolution. In natural–source studies, such as in teleseismic tomography, the distribution of sources and stations commonly results in similar ray paths sampled multiple times, with little angular diversity across model parameters. Consequently, parameters may have both high ray–density and relatively low resolution. Conversely, in many active–source studies, model elements may be traversed by fewer rays with higher angular diversity, resulting in parameters with relatively low ray density but high resolution.

We compare ray–sampling to estimated model resolution diagonals to further illustrate the utility of the latter in quantitative resolution analysis. Figure 3d (and h) shows log total ray length across the model volume for the sources and stations shown in Figure 1. The large number of events with northwest back azimuths result in total ray length > 500 km along northwest–directed rays, to $\sim$400 km depth beneath the CREST network. From this metric, one may infer a corresponding co–located region of moderately well–resolved model parameters. However, equivalent plots of estimated resolution diagonal for the $\alpha = 0.7$ inversion show a region of approximately equal (diagonal) resolution of 0.1–0.2 along northwest–directed rays to depths of 500–600 km (Figure 3c). The $\alpha = 0.1$ inversion, because it employs less smoothing and damping, has ubiquitous higher resolution and shows diagonal resolution > 0.4 to depths exceeding 600 km along northwest–directed rays (Figure 3g). Because there is not a strict correlation between ray sampling and (diagonal) resolution, particularly in the presence of regularization, estimates of diagonal resolution may be a favorable low–cost alternative to ray–sampling density for resolution analysis.

5. Conclusions

We present a general low–cost stochastic matrix diagonal method to estimate the model resolution matrix diagonal and the generalized cross–validation (GCV) function. The method is demonstrated using a moderately large teleseismic P velocity linear inversion example, and the results are compared against those from trade–off curves, checkerboard resolution tests, and ray–sampling density. The method presented here relies on LSQR and is comparable in computational demand to the effort necessary for obtaining model solutions. The method thus provides easily implemented estimation and assessment of the complete resolution matrix diagonal as well as wider usage of GCV–determined regularization parameter estimation, and is scalable to very large inverse problems.

Acknowledgments. Seismic instruments were provided by the Incorporated Research Institutions for Seismology (IRIS) through the PASSCAL Instrument Center at New Mexico Tech. The CREST project is funded by the National Science Foundation Continental Dynamics Program under award EAR-0607693. The facilities of the IRIS Consortium are supported by the National Science Foundation under Cooperative Agreement EAR-0552316, the NSF Office of Polar Programs and the DOE National Nuclear Security Administration. Data from the EarthScope Transportable Array network were made freely available as part of the EarthScope USArray facility supported by the National Science Foundation Major Research Facility program under Cooperative Agreement EAR-0350030.

References

Aster, R., B. Borchers, and C. Thurber (2005), *Parameter Estimation and Inverse Problems*, Academic Press.

Aster, R., J. MacCarthy, M. Heizler, S. Kelley, K. Karlstrom, L. Crossey, K. Dueker, and the CREST Team (2009), CREST experiment probes the roots and geologic history of the colorado rockies, *Outcrop*, *58*(1), 6–21.

Bekas, C., E. Kokiopoulou, and Y. Saad (2007), An estimator for the diagonal of a matrix, *Applied Num. Math.*, *57*(11-12), 1214–1229.

Berryman, J. (2000), Analysis of approximate inverses in tomography ii. iterative inverses, *Optimization and Engineering*, *1*(4), 437–473.

Boschi, L. (2003), Measures of resolution in global body wave tomography, *Geop. Res. Lett.*, *30*(19), 1978, doi:10.1029/2003GL018222.

Chiao, L., and B. Kuo (2001), Multiscale seismic tomography, *Geophysical Journal International*, *145*(2), 517–527.

Craven, P., and G. Wahba (1979), Smoothing noisy data with spline functions, *Numerische Mathematik*, *31*, 377–403.

Dahlen, F. A., S. H. Hung, and G. Nolet (2000), Frechet kernels for finite-frequency traveltimes - i. theory, *Geophys. J. Int.*, *141*(1), 157–174.

Deal, M., and G. Nolet (1996), Comment on 'estimation of resolution and covariance for large matrix inversions, *Geophysical Journal International*, *127*(1), 245–250.

Girard, D. (1987), Un algorithme simple et rapide pour la validation croisée généralisée sur des problèmes de grande taille, *RR 669-M, Grenoble, France: Informatique et Mathématiques Appliquées de Grenoble*.

Golub, G., M. Heath, and G. Wahba (1979), Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, *21*(2), 215–223.

Golub, G. H., and U. vonMatt (1997), Generalized cross-validation for large-scale problems, *J. Comp. Graph. Stat.*, *6*(1), 1–34.

Hansen, P., and D. O'Leary (1993), The use of the l-curve in the regularization of discrete ill-posed problems, *SIAM Journal on Scientific Computing*, *14*(6), 1503.

Hutchinson, M. F. (1990), A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines, *Communications Stat.-Simulation and Computation*, *19*(2), 433–450.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the earth from travel-times, *Geophys. J. Int.*, *122*(1), 108–124.

Li, C., R. D. van der Hilst, E. R. Engdahl, and S. Burdick (2008), A new global model for p wave speed variations in earth's mantle, *Geoc., Geop., Geosys.*, *9*, Q05,018, doi: 10.1029/2007GC001806.

MacCarthy, J. (2010), The structure of the lithosphere beneath the Colorado Rocky Mountains and support for high elevations, *Ph.D. Thesis, New Mexico Institute of Mining and Technology.*

Marquering, H., F. A. Dahlen, and G. Nolet (1999), Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, *137*(3), 805–815.

Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press.

Nolet, G., R. Montelli, and J. Virieux (1999), Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, *138*(1), 36–44.

Paige, C., and M. Saunders (1982), LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Transactions on Mathematical Software (TOMS)*, *8*(1), 43–71.

Parker, R. (1994), *Geophysical Inverse Theory*, Princeton Univ. Press.

Pavlis, G. L., and F. L. Vernon (2010), Array processing of teleseismic body waves with the USArray, *Computers & Geosciences*, *36*(7), 910 – 920, doi:10.1016/j.cageo.2009.10.008.

VanDecar, J., and R. Crossen (1990), Determination of teleseismic releative phase arrival times using multi-channel cross-correlation and least squares, *Bull. Seism. Soc. Am.*, *80*(1), 150–169.

Wahba, G. (1990), *Spline models for observational data*, Society for Industrial Mathematics.

Waite, G., R. Smith, and R. Allen (2006), VP and VS structure of the Yellowstone hot spot from teleseismic tomography: Evidence for an upper mantle plume, *Journal of Geophysical Research*, *111*(B4), B04,303.

Yao, Z. S., R. G. Roberts, and A. Tryggvason (1999), Calculating resolution and covariance matrices for seismic tomography with the LSQR method, *Geophys. J. Int.*, *138*(3), 886–894.

Zhang, H., and C. H. Thurber (2007), Estimating the model resolution matrix for large seismic tomography problems based on lanczos bidiagonalization with partial reorthogonalization, *Geophys. J. Int.*, *170*(1), 337–345, doi:10.1111/j.1365-246X.2007.03418.x.

Zhang, J., and G. A. McMechan (1995), Estimation of resolution and covariance for large matrix inversions, *Geophys. J. Int.*, *121*(2), 409–426.

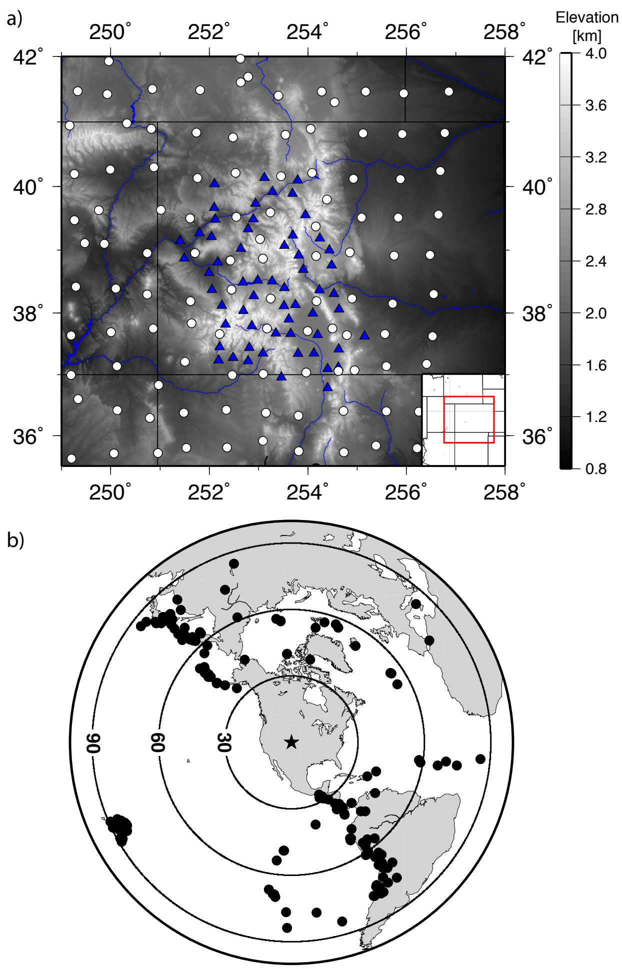

Figure 1. a) Map of stations used in the CREST experiment over elevation. CREST stations are triangles, and USArray stations are circles. b) Distribution of teleseismic earthquake sources (black circles). The center of the CREST network is noted by a star.

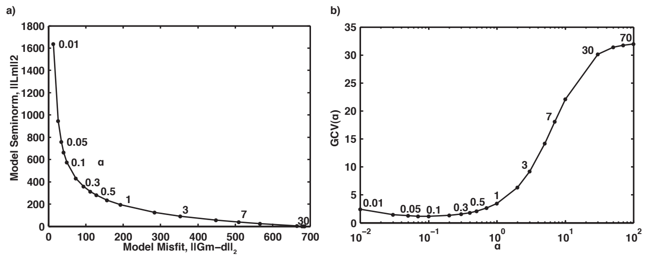

Figure 2. a) Example trade–off curve between model seminorm versus data residual 2-norms as a function of regularization weighting parameter, α (2) for regularization as described in the text. b) Generalized cross–validation (GCV) curve, showing regularization parameter (α) versus GCV function value (6).

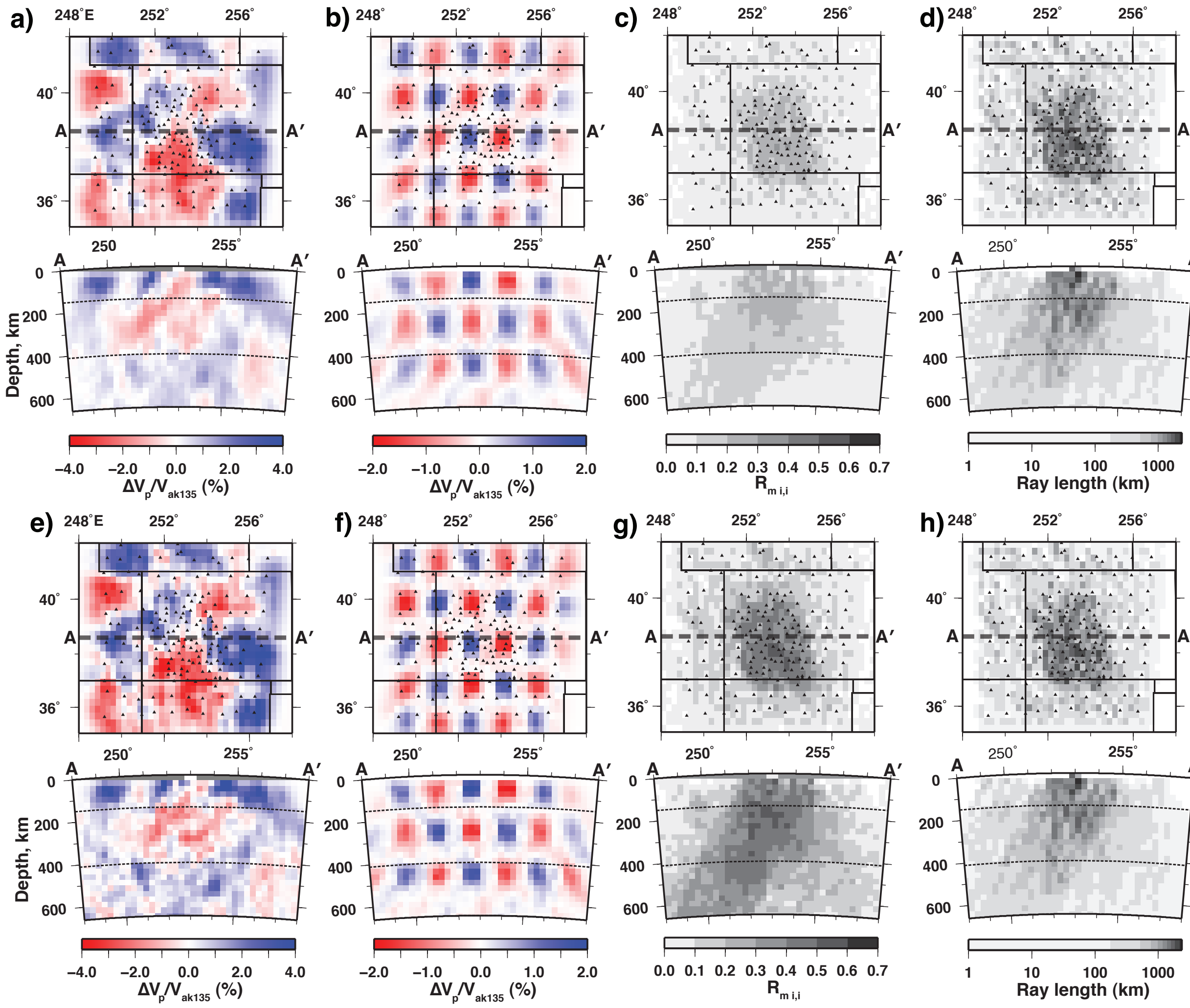

Figure 3. CREST regional model slices and resolution analysis of example P-wave regularized inversions with $\alpha = 0.7$ (a–d) and with $\alpha = 0.1$ (e–h). (a,e): depth slice of velocity model at 90 km depth (top). Seismic stations are small black triangles, and the dashed line **AA′** is the location of the paired cross section (bottom). Depths at 150 km and 440 km are shown as dashed lines in cross section. Velocities are percent of Vp relative to the ak135 reference model. (b,f): Checkerboard recovery at same depth and latitude as previous. Input perturbations were ±2% P velocity relative to background across sets of 3^3 model blocks. (c,g): Stochastic estimate of diagonal elements of $\mathbf{R_m}$. (d,h): Total ray length for all used P rays through each model parameter. Plots d) and h) are identical, repeated to aid visual comparison.

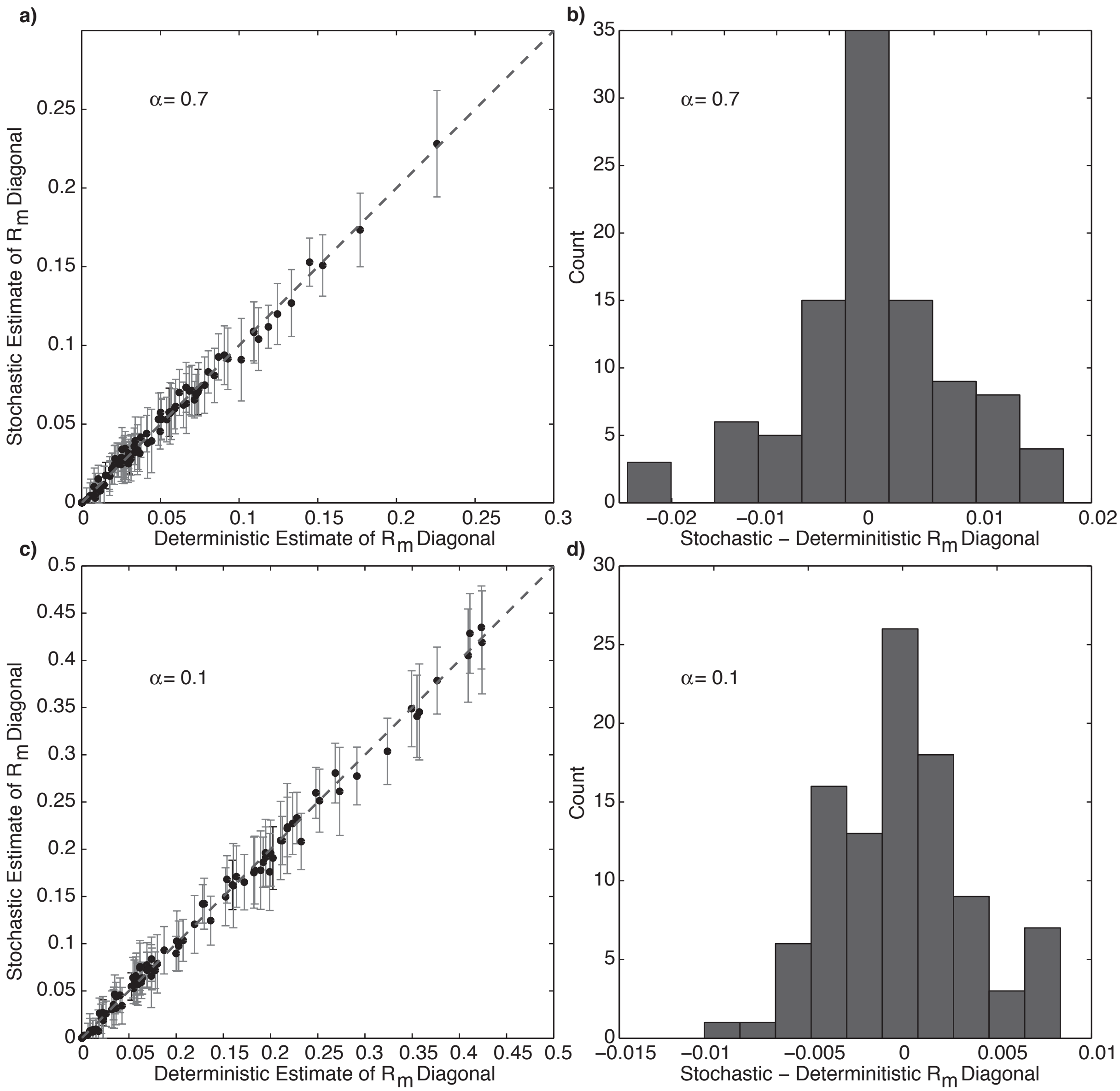

Figure 4. a) Stochastic estimates of the resolution matrix diagonal (y-axis), versus true values (x-axis) for 100 randomly selected values parameter values, $\alpha = 0.7$ case. Points are the median values of N=20 realizations using s=256 random vectors each. Bars are the symmetric sample standard deviations for each parameter. b) Histogram of residuals between median estimated and true R_m diagonal values for the same 100 parameters. (c,d) Same previous plots, but for α=0.1 case. The same 100 parameters are investigated.

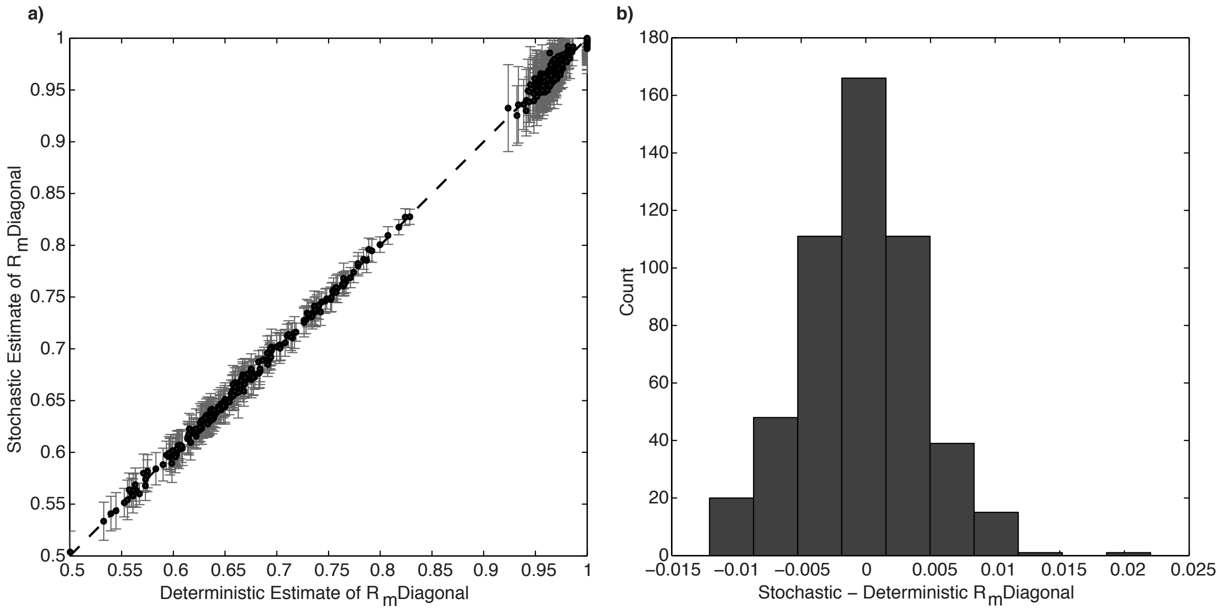

Figure 5. a) Stochastic estimates of the resolution matrix diagonal (y-axis) versus true values (x-axis) for all 512 parameters in a synthetic 3D tomography example, using $\alpha = 0.5$. Points are the median values of N=20 realizations using s=256 random vectors each. Bars are the symmetric sample standard deviations for each parameter. b) Histogram of residuals between median estimated and true R_m diagonal values.