Estimating data uncertainties for least squares optimization

Kasper van Wijk
Department of Geophysics, Utrecht University, The Netherlands

John A. Scales
Department of Geophysics & Center for Wave Phenomena, Colorado School of Mines

William Navidi
Department of Mathematical and Computer Sciences, Colorado School of Mines

K. Roy-Chowdhury
Department of Geophysics, Utrecht University, The Netherlands

ABSTRACT
When fitting model parameters to observed data, as in inverse calculations, it is essential to have some estimate of the data uncertainties. If these estimates are too small, then the fitting procedure will likely put features into the model that are not required by the data. If the estimated data uncertainties are too large, then the fitting will not extract all the available information from the data. In practice, these uncertainties are often asserted *a priori* or estimated globally for the entire data set. We develop several simple and robust optimization procedures that estimate data uncertainty automatically from the data. The simplest approach is to obtain a global estimate of the errors from ordinary least-squares. We then show how binning the data can be used to compute a more detailed, estimate of the error distribution. Finally, we propose and iterative extension of this approach. This uncertainty information can then be incorporated into the least squares optimization procedure by weighting the observations in each bin with the reciprocal of their locally computed variance. This leads to a robust least squares solution since bins with outliers in them will be down-weighted. We apply these methods to a synthetic vertical seismic profile (VSP), and to a real data set involving bathymetry measurements from the Sea of Galilee. These examples illustrate the fact that better estimates of data errors lead to a better distinction between data uncertainty and the effects of local model features.

Introduction

Geophysical inversion problems are fundamentally problems of statistical inference. In this paper we study a key component of such calculations, namely the calculation (by optimization) of finite dimensional vectors of model parameters whose response agrees with the observed data to some tolerance. To solve the complete inverse problem we would need to address many other issues such as the accuracy of the forward modeling, the level of discretization (since earth models are really functions) and *a priori* information. Here, instead, we focus on the problem of determining the data uncertainties directly from the data, so that we can measure the extent to which models fit the data. Thus our problem can be stated mathematically as follows.

Let $d \in R^n$ be a vector of observations, $x \in R^m$ be a vector of model parameters, and $A \in R^{n \times m}$ a linear operator (the forward modeling operator) mapping model vectors into data vectors. We assume that the observed data d are related to the *true* data d_T (the data associated with a perfect noise-free experiment) by $d = d_T + \epsilon$, where ϵ is random. Further, we assume that the components of d are independent and can be adequately described by a Gaussian distribution. Finding models that fit the data can then be cast as a weighted least squares problem. For the classical problem of estimating m pa-

rameters from n observations we use the standard χ^2 measure:

$$\chi^2 = \frac{1}{n}\sum_{i=0}^{n}\left(\frac{\sum_{j=1}^{m} A_{ij}x_j - d_i}{\sigma_i}\right)^2 \quad (1)$$

with σ_i being the standard deviation of the i-th datum.

The difficulty in practice is that we do not know σ_i. Various strategies for assigning or computing the data uncertainties have been proposed. The simplest of course is to fix all the σ_i to some σ *a priori*, based on the assumed precision of the experiment. For example, Oldenburg *et al.* (1997) invert a variety of exploration geophysical data sets, but all are assumed to have a constant error of 5%. Another strategy is to try to estimate a global σ from the residuals of a "good" model (e.g., Rogers & Wahr (1993)). Strictly speaking this is not possible since the data errors must be independent of the model, but it is a reasonable strategy in practice, and we will show below how to extend this idea to estimate a spatially varying data variance. In Bayesian inversion it is necessary to estimate the joint distribution function of the data uncertainties. For example, Gouveia & Scales (1998) use subsets of the data, which are assumed to contain only ambient noise, to estimate the mean and covariance of an N-dimensional normal error distribution. But this is a complicated procedure which may introduce sampling artifacts unless some model of the covariance structure is assumed.

In this paper we will describe three simple methods for estimating the data uncertainty automatically from the data. There are certainly more sophisticated approaches available, but our goal is to develop algorithms that are robust and easy to use.

Let us first state the basic linear estimation problem for known i.i.d. (independent, identically distributed) data uncertainties:

χ^2 estimation *Let x be the m-dimensional model vector, and d the vector with n observations. Let the vector σ contain an estimate of the data standard deviation.*

(i) Let $\chi^2 = \frac{1}{n}\sum_{i=0}^{N}\left(\frac{\sum_{j=1}^{m} A_{ij}x_j - d_i}{\sigma_i}\right)^2$

(ii) Solve $\hat{x} = \min_x |\chi^2 - 1|$

By minimizing $\chi^2 - 1$, algorithms will stop when the model predicts the observations on average within one standard deviation. We can replace 1 by some other number, but whatever we choose we must acknowledge the fact that the least squares model (obtained by replacing the 1 with a 0) will usually over-fit the data. We could also use $\chi^2 = 1$ as a constraint in a more general optimization problem; the point is simply that we fit the data up to some tolerance. In practice we achieve this by adding features to the model until we just manage to fit the data. This sort of algorithm can be implemented with any least squares code; we refer to the appendix for the weighted conjugate gradient algorithm used in this paper. In any case, χ^2 estimation assumes that one knows the data errors σ_i.

Global estimate of data uncertainty from OLS

We will first show that one can obtain an average of the data uncertainty directly from the data. Recall that we have

$$Ax = d = d_T + \varepsilon.$$

A is the forward operator with n rows and m columns, d is the n dimensional observed data, and x is the vector of model parameters of dimension m. The noise vector ε also has dimension n. To get a global estimate of data uncertainty each component of ε is assumed to be random with zero mean and variance σ^2.

The ordinary least squares (OLS) estimate of x is

$$\hat{x} = A^{\dagger}d \quad (2)$$

where $A^{\dagger}$ is the pseudo inverse of the A. When A^TA is nonsingular, the OLS estimate is unique:

$$\hat{x} = (A^TA)^{-1}A^Td.$$

In practice, A^TA is usually not invertible. In our numerical calculations we use the conjugate gradient least squares algorithm (Scales, 1987), which converges to the pseudo inverse solution.

The normal equations are a set of m linear equations in m unknowns. The residual vector $d - A\hat{x}$ is the projection of the error vector ε into the $n-p$ dimensional space orthogonal to the column space of A, where p is the rank of A. We assume that $p < n$. For this reason, the squared length (i.e. l_2 norm) of the residual vector has the expected value $(n-p)/n$ times the expected squared length of the error vector ε. Since the expected squared length of ε is $n\sigma^2$, the expected squared length of the residual vector is $(n-p)\sigma^2$ (Stuart & Ord, 1987). Therefore, the estimate of the global variance σ^2 is

$$\hat{\sigma}^2 = \frac{||d - A\hat{x}||^2}{n - p} \tag{3}$$

We call the resulting algorithm Global χ^2 estimation:

Algorithm 1. Global χ^2 estimation Let $A^\dagger$ be the pseudo inverse of A, let p be the rank of A and let n be the number of data, d.

(i) Compute the OLS estimate $\hat{x} = A^\dagger d$.
(ii) Define the average data variance $\hat{\sigma}^2 = \frac{||d - A\hat{x}||^2}{n-p}$.
(iii) Perform χ^2 estimation with $\sigma_i^2 = \hat{\sigma}^2$.

Assuming locally constant data uncertainty

The Global χ^2 approach leads to a global value for the data uncertainty. This is simple but has the disadvantage that we weight all data equally, regardless of quality. We will now introduce a method to recover an estimate of the data uncertainty that is only locally constant. This will allow us to differentiate between areas of good and bad data quality.

We will assign sub-regions (or bins) where we assume the data uncertainty is constant. For example, bins may represent regions in space or time. Within each bin we compute the variance. The resulting regional data variances are then used as the weights in the basic χ^2 estimation procedure. We call this algorithm Local χ^2 estimation.

Algorithm 2. Local χ^2 estimation

(i) Group n observations in b bins.
(ii) Compute the data variance in each bin i: $\{\sigma_j^{*\,2}\}_{j=1}^b$.
(iii) Perform χ^2 estimation with $\sigma_i^2 = \sigma_{j_i}^{*\,2}$, where j_i is the bin containing the observation d_i.

Local χ^2 estimation down-weights regions of the data which exhibit large local fluctuations. This leads to a more robust procedure, since the influence of large outbursts of noise ("spikes") is reduced. On the other hand, bins in which large local fluctuations are systematic will be down-weighted as well. The models generated by this procedure will tend to smooth out local features in the data.

In applying this method, we must make decisions regarding the number and size of the bins. Our goal here is to discuss the effects of binning the data rather than to recommend optimal methods. For some discussion on bin selection methods, see (Freedman & Diaconis, 1981a) and (Freedman & Diaconis, 1981b).

Extension to an iterative scheme

In our third algorithm we will iteratively up-date the weights σ_i to achieve higher resolution in the model. The initial weights are determined by computing the variance of the binned residuals resulting from Local χ^2 estimation (algorithm 2). With these weights we perform χ^2 estimation. The resulting model gives rise to a new set of residuals from which updated variances are computed. The process repeats until the weights converge. The advantage of this method is that the component of the data uncertainty as computed in algorithm 2 that was due to structure in the model has been reduced. We refer to this algorithm as Iterative Local χ^2 estimation:

Algorithm 3. Iterative Local χ^2 estimation Let the model $\tilde{x}$ be the result of optimization with b bins each with constant data variance (algorithm 2).

(i) Group the n residuals $A\tilde{x} - d$ in b bins.
(ii) Let $\sigma_j'^{\,2}$ be the variance of the residuals in the j-th bin.
(iii) Perform χ^2 estimation with $\sigma_i^2 = \sigma_{j_i}'^{\,2}$, where j_i is the bin containing the observation d_i.

The data variances can be iteratively updated by using the resulting model x in item 1.

Summarizing, we have an optimization procedure (χ^2 estimation) that requires an estimate of the data uncertainty. We have proposed three methods to obtain such an estimate. The first method involves a global data variance estimate, while the other two involve local estimates based on binning the observations. In the next sections we will apply each method to a synthetic example.

A synthetic VSP example with Gaussian noise

This synthetic VSP experiment contains of one nearly zero-offset shot recorded at 78 geophones in a bore-hole (figure 1). The velocity model of the subsurface is represented by 41 constant velocity layers. Using a ray tracing forward operator, we have computed a synthetic data set from a model with increasing velocity with depth, except

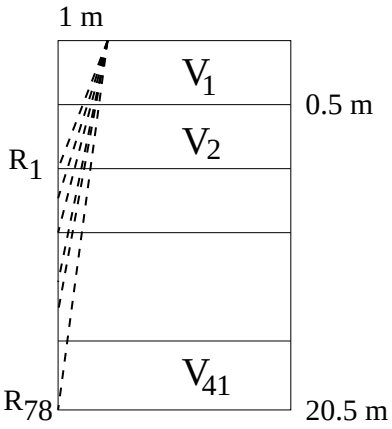

Figure 1. *Configuration of shot and receivers in VSP.*

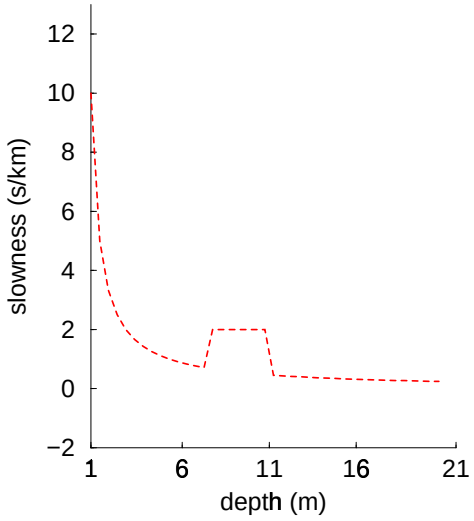

Figure 2. Model to be recovered from VSP travel times. The velocity is increasing with depth except for a low-velocity layer.

that layers 14 through 20 are low velocity layers (LVL) as shown in figure 2.

We added uncorrelated Gaussian noise with zero mean and a variance of 1 millisecond to the noise free data. All results in this section are the average of 100 synthetic experiments with different realizations of this noise. For simplicity, the rays are treated as straight lines. Also, the first two layers contain no receivers. These conditions make it impossible to resolve the slowness of the first two layers, individually. Further, note that a rapid increase in velocity with depth in the shallow layers makes it possible for the arrival time at the second receiver to be shorter than at the first.

Our first experiment on the synthetic data set is an attempt to retrieve the (correct) 1 ms value of the data uncertainty with Global χ^2 estimation. From equation 3 we know that if we have the ordinary least squares estimate of x, we can calculate an estimate of the variance. Using an undamped conjugate gradient algorithm (described in the appendix), the estimate for the variance converges within 40 iterations to 1.01 ms, as shown in figure 3. Figure 4 shows the results of χ^2 estimation with this global data variance estimate. Because the slowness of the first two layers cannot be resolved, the estimate is poor near the surface.

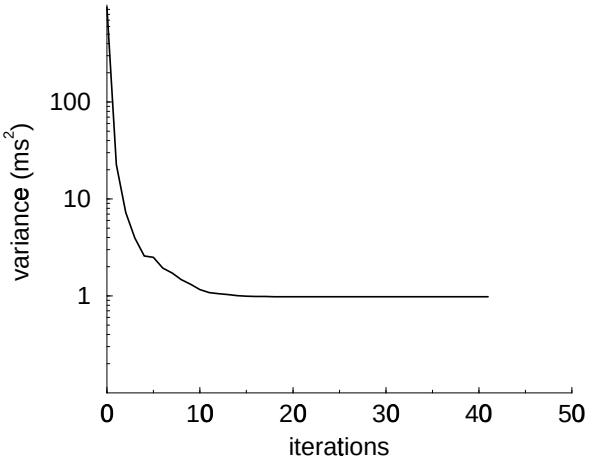

Figure 3. *Global estimate of the variance as a function of the number of iterations using a conjugate gradient algorithm to perform χ^2 estimation. The estimate converges to 1.01 ms.*

Next, we applied Local χ^2 estimation (algorithm 2). We grouped the data in bins of six receivers (13 bins). In each bin the data variance is computed. The variances act as weights in the χ^2 estimation scheme, the results of which are shown in figure 5. Although the computed model still seems to represent the features of the true model, we see a slight loss of resolution in the flanks of the LVL. Finally, we performed five iterations of Iterative Local χ^2 estimation (algorithm 3). These results are shown in 6. We see very similar results to those obtained with Local χ^2 estimation, except for an improvement in the flanks of the LVL.

Inversion with a global estimate of the data uncertainty performed best since the observations do in fact have constant data variance. The Local χ^2 estimation leads to the over-estimate of the data variance (see figure 7) causing loss of resolution in the LVL. The Iterative Local χ^2 estimation manages to diminish the over-estimate. Table 1 shows the true and computed slownesses in depths sur-

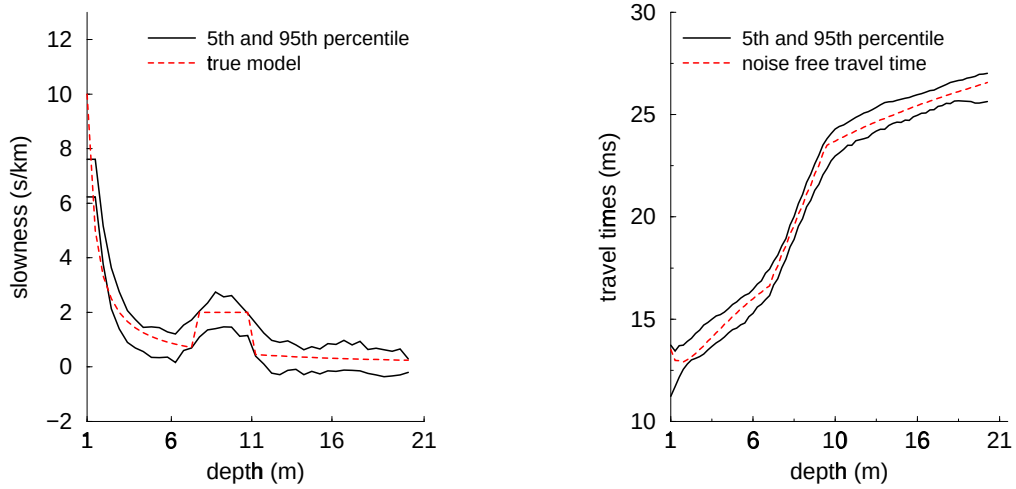

Figure 4. *Predicted model (left) and data (right) from Global χ^2 estimation with an estimated data uncertainty of 1.01 ms. The solid lines represent the the top and bottom 5 percentile estimate of slowness (left) and travel time for 100 experiments, each with a different realization of uncorrelated noise with zero mean and a standard deviation of 1 ms.*

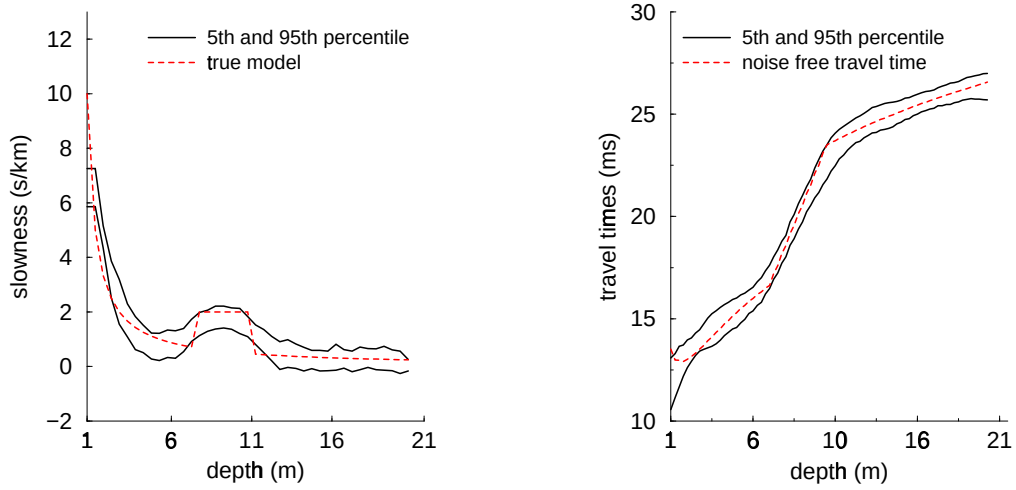

Figure 5. *Same as in figure 3, but for Local χ^2 estimation.*

rounding the LVL. It is clear that iterating the variance estimates improves the resolution.

The VSP example with Gaussian noise and outliers

In geophysical data, noise spikes are relatively common. In this example we will use the same VSP data as in the last section, but in addition to the Gaussian noise two noise spikes of 10 ms amplitude have been added. These spikes are located at depths of 5 m and 19 m. Rather than apply a rejection criterion we will show how locally updating the data variances down-weights these large deviations automatically.

First, we computed the the global estimate of the data variance (equation 3). The value was 3.73 ms, which, due to the noise spikes, is significantly more than the 1 ms Gaussian noise added. (Without the Gaussian noise, the two spikes lead to a global data variance of 2.7 ms.) Using a global estimate of the data variance of 3.73 ms we have computed the Global χ^2 estimate (figure 8). With the global estimate of the data uncertainty, every data point is of equal weight. Therefore, this algorithm resulted in a model that predicts not only the LVL, but also part of the outliers.

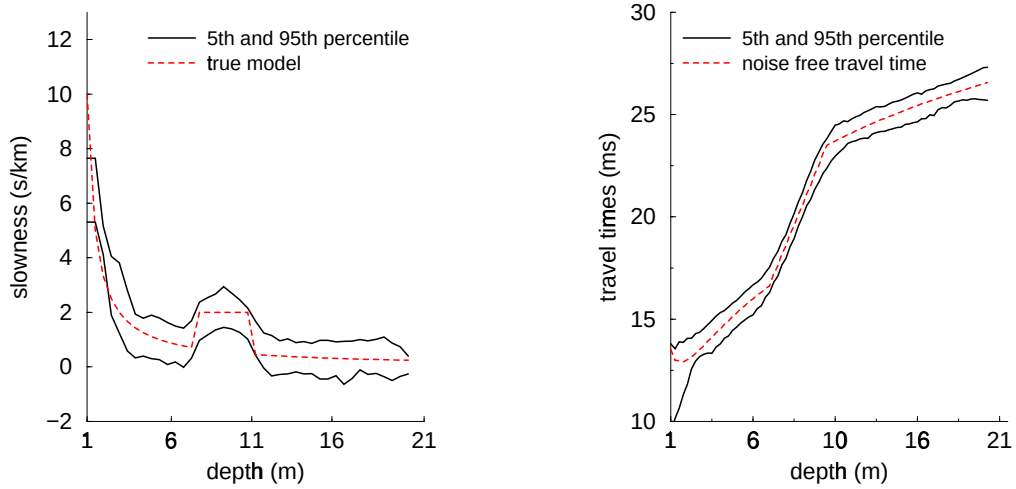

Figure 6. *Same as figure 3, but for Iterative Local χ^2 estimation.*

parameter	*13*	*14*	*15*	*19*	*20*	*21*
true model	0.71	2.0	2.0	2.0	2.0	0.45
Local χ^2	1.32	1.55	1.68	1.72	1.48	1.16
Iterative Local χ^2	1.22	1.66	1.90	1.84	1.48	1.01

Table 1. *Table shows the true and computed slownesses for the VSP problem with Gaussian noise. The low velocity layer (LVL) starts in the true model at model parameter 14 and ends at parameter 20. The results are the average over 1000 experiments.*

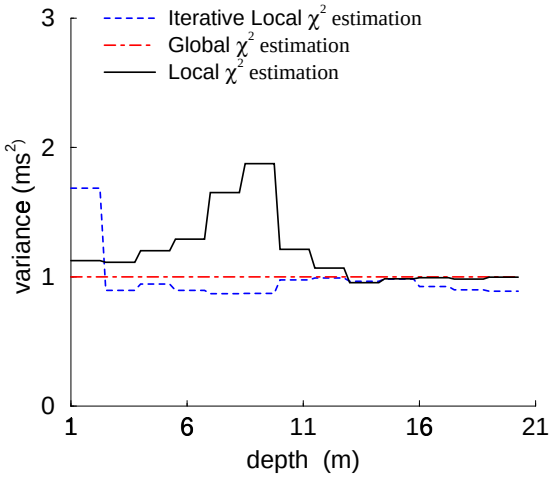

Figure 7. *The variance distribution from the three different different algorithms assuming 1 ms uncorrelated Gaussian noise.*

Second, we applied Local χ^2 estimation (algorithm 2) to the data with outliers. We grouped the receivers in 13 bins again. With the variance per bin as an estimate of the data uncertainty, the resulting model does not follow the spikes (figure 9). The relatively high variance in the bins with outliers (see figure 10) causes down-weighting of these observations.

After grouping the data in 13 bins, algorithm 2 produced the initial model to apply the iterative method of algorithm 3. After 5 iterations, the results also show robust behavior in the presence of outliers (figure 11). Again, we see an improvement in resolution in the flanks of the LVL with respect to Local χ^2 estimation, because of a locally more accurate estimate of the data uncertainty (see figure 10).

Estimating the data variances locally provided robustness in the optimization, while the physical features as the low velocity zone were still well resolved. Although the bins were large due to the limited number of data, inversion with the data variance estimate from binning the data still showed encouraging results. With a more dense data set we could choose a smaller bin size. This could improve resolution. The next section provides an example with a more dense distribution of observations.

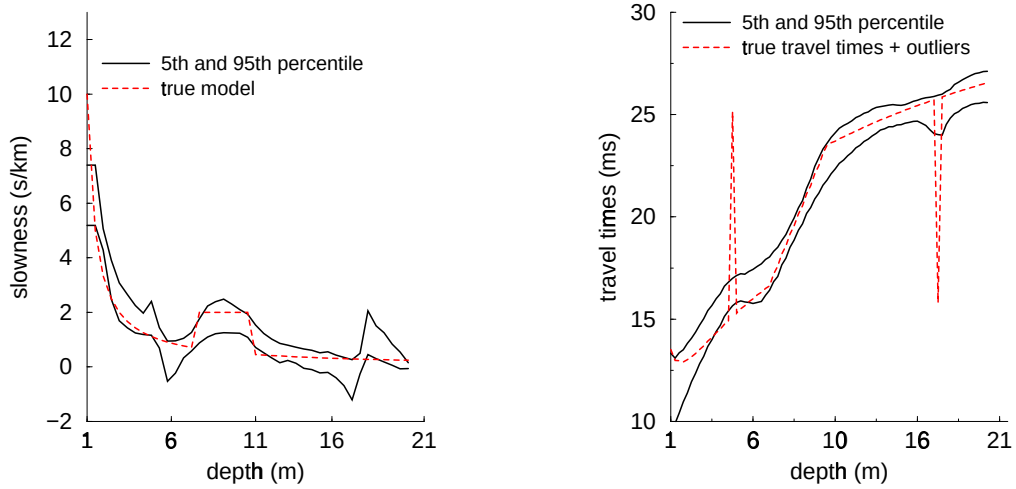

Figure 8. *Predicted model (left) and data (right) using Global χ^2 estimation in the presence of outliers. Solid lines represent the the top and bottom 5 percentile estimate of slowness and travel time for 100 experiments, each with a different realization of uncorrelated noise with zero mean and a standard deviation of 1 ms.*

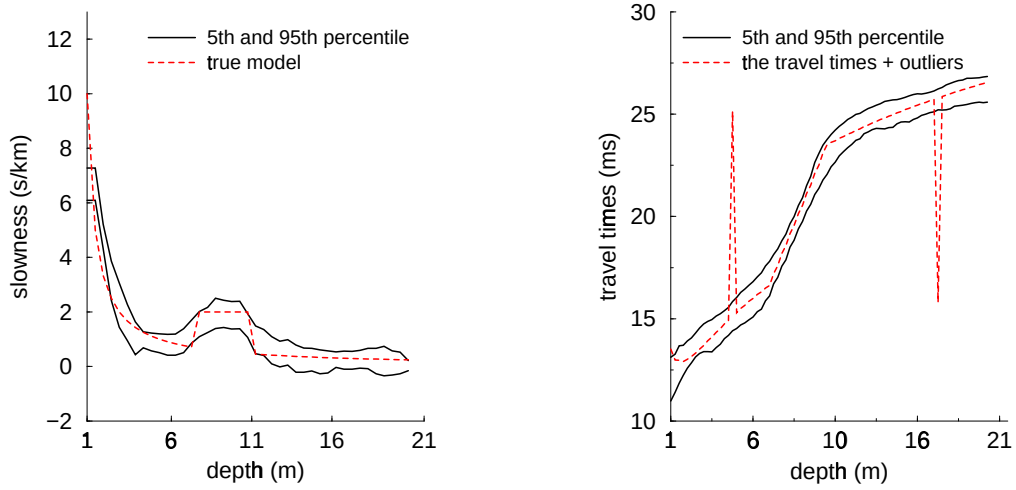

Figure 9. *Same as in figure 8, but for Local χ^2 estimation.*

Example: Sea of Galilee bathymetry

Ben-Avrahim *et al.*(1990) have published depth measurements of the Sea of Galilee. The system they used combined estimates of the boat's position with echo-sounding data. Tidal and other corrections were applied in real-time. The echo-soundings themselves were calibrated every few hours using a bar that was lowered to the bottom. The depth values were transformed to values below mean sea level using measured values of the water level on the surface. The errors associated with each of these steps combine. A lower bound on the errors in the depth values is .01%, which is the quoted accuracy of the echo-sounder.

The soundings were made on an irregular grid. If we define linear interpolation as extracting values between the known values on a regular grid, then inferring the depths on a regular grid from the irregularly spaced data can be considered a linear inverse problem (Claerbout, 1997). We will now solve this problem by the χ^2 estimation procedure using estimates of the data uncertainties obtained from our three algorithms. We use a four point linear interpolator as the forward operator. To speed up the calculations, we worked on a subset of the data, highlighted in figure 12. To view the data, we have plotted the mean depth in square bins of 50 meter sides (figure 13).

Computing the least squares estimate of the depths on a

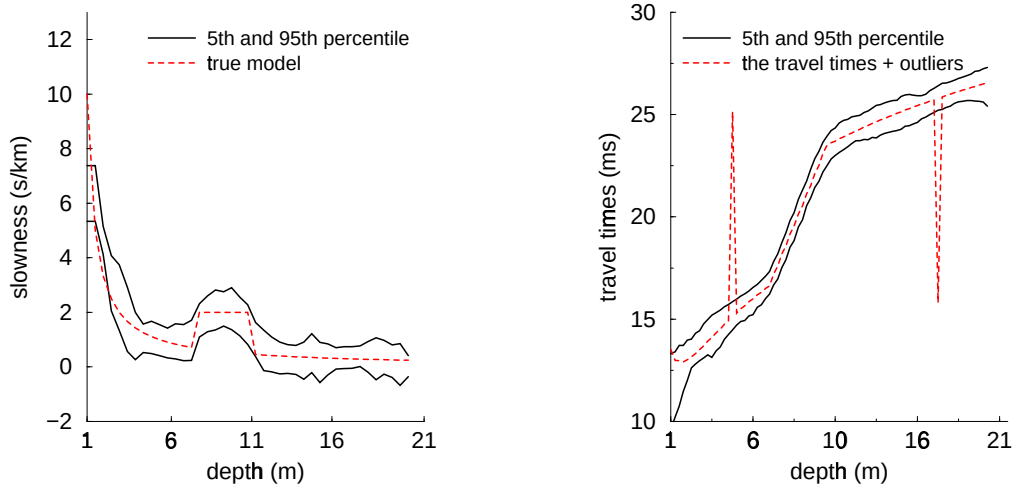

Figure 11. *Same as in figure 8, but for Iterative Local χ^2 estimation.*

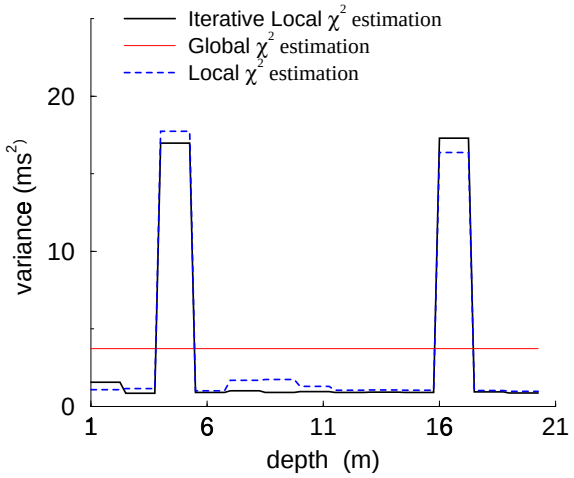

Figure 10. *The variance distribution from the different data variance estimating methods assuming 1 ms uncorrelated Gaussian noise and two 10 ms noise spikes.*

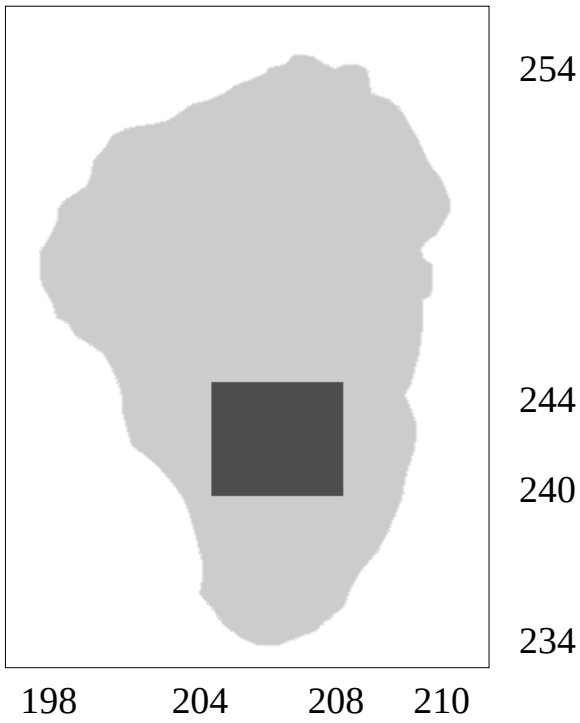

Figure 12. *Location of the sub-set in the Sea of Galilee. Dimensions in km.*

regular grid of 125 by 125 meters, resulted in a estimate of the data variance of 0.41 m^2 (figure 14). This means the standard deviation is 0.64 m or 0.3 percent of the average depth. Inversion with the global estimate of a data variance of 0.41 m^2 resulted in a overall smooth model with some local structures at (205 km, 243 km) and (207 km, 241 km) (figure 15).

Next we performed the Local χ^2 estimation procedure assuming the data variance to be constant over square areas of 125 by 125 meter. We defined our interpolation grid to match the bin distribution, i.e., the grid points are 125 meters apart. With this estimate of the data variance, we have obtained a model of the subsurface (shown figure 16), that achieves a normalized χ^2 of slightly less than 1. As can be seen in this figure, typical variances are on the order of .075 m^2, with isolated bins of higher variance. (A variance of .075 m^2 corresponds to depth errors on the order of .1%.) This means that in most of the model we will be able to extract more information from the data (put more features into the model) than we could with the relatively large global variance esti-

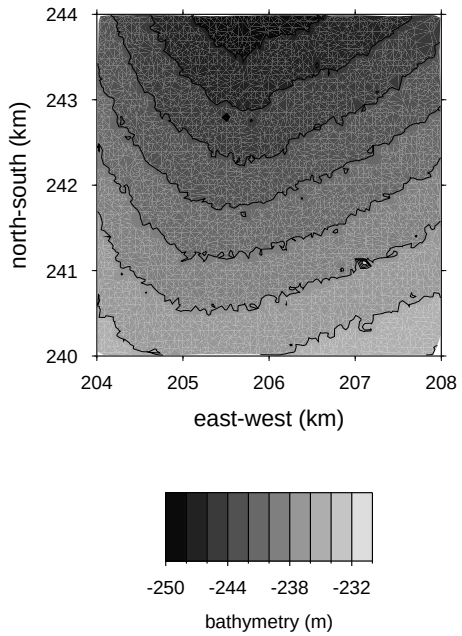

Figure 13. *Contour plot of the mean of the depth measurements in bins of 50 by 50 m.*

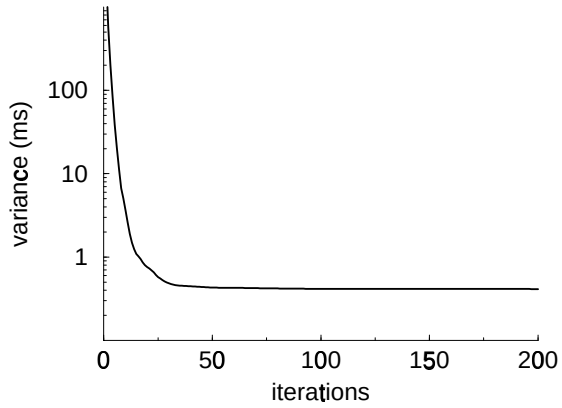

Figure 14. *Global data variance estimate as a function of the number of iterations. The estimate converges to 0.41* m^2.

mate of .41 m^2. But those features associated with large residuals in the data will be down-weighted.

Figure 17 is the result of Iterative Local χ^2 estimation from the initial Local χ^2 estimate (algorithm 2). The weights were up-dated five times. The character of the model has not changed much with respect to the result from 2 (figure 16). This implies the seafloor is generally

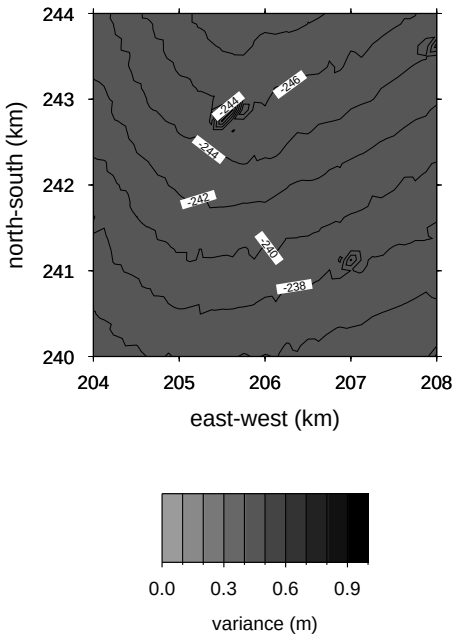

Figure 15. *Contour plot of the sea floor after χ^2 estimation with a global data variance estimate of 0.41* m^2. *The optimization is stopped when $\chi^2 = 1$. The grey levels refer to variance of the bins(in this case a constant), while the contours are labeled according to depths. Notice that with such a large global data variance the model is quite smooth, but with a few isolated topographic features.*

quite smooth, so that the weights do not change much when updated.

Conclusion

The character of least squares estimated models depend on the data variance estimate. If these estimates are too small, then the fitting procedure will likely put features into the model that are not required by the data. If the estimated data uncertainties are too large, then the fitting will not extract all the available information from the data. We have shown three simple, efficient algorithms for automatically estimating the data uncertainty for least squares optimization. One of the algorithms estimates a global value for the data uncertainty (Global χ^2 estimation). The other two algorithms allow regionally varying estimates of the data uncertainty. Local χ^2 estimation computes the variance of the observations in bins. In case of model structure on a scale smaller than

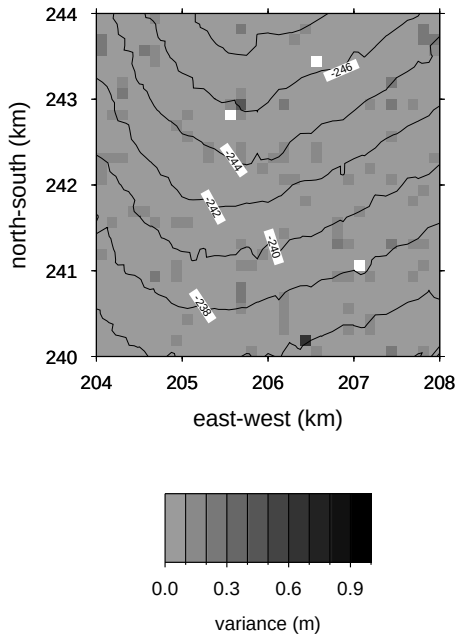

Figure 16. *Contoured image of the sea floor after local χ^2estimation. The shades of grey reflect the data variance in each bin. The prevailing variances are of the order of 0.075* m^2. *White bins have a data variance larger than 1* m^2. *Thus most of the data is considerably lower variance than we estimated with Global χ^2, but there a number of bins of higher variance. These get down-weighted by the local algorithms.*

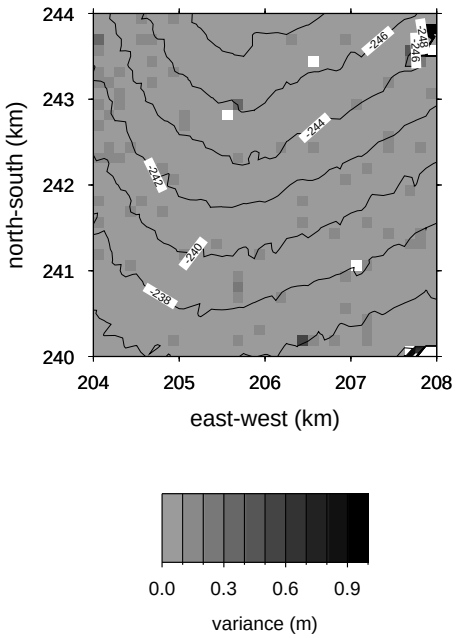

Figure 17. *Contoured image of the sea floor after five iterations of Iterative Local χ^2 estimation. The shades of grey reflect the data variance in each bin. The prevailing variances are of the order of 0.075* m^2. *White bins have a data variance larger than 1* m^2. *This result Iterative Local χ^2 result is very similar that obtained using Local χ^2.*

the bin size, such a procedure would lead to an overestimate of the data variance. An iterative up-dating of the weights (Iterative Local χ^2 estimation) can improve the accuracy of the data variance estimate. We have shown that in cases where the data uncertainty varies regionally, these local estimates of the data uncertainty introduce robustness into the optimization scheme by down-weighting regions of large variance. In regions of small data variance we can extract more information about the model parameters. We have illustrated the application of these algorithms on both synthetic examples as well as a bathymetry study of the Sea of Galilee.

Acknowledgment

KVW acknowledges the support of the Vening Meinsez Research School of Geodynamics, Utrecht University. The authors would like to thank Professor Zvi Ben-Avraham of The Dead Sea Research Center for making the data of the Sea of Galilee available to us.

References

Ben-Avraham, Z., Amit, G., Golan, A., & Begin, Z. 1990. The bathymetry of Lake Kinneret. *Israel J. Earth Sci*, **39**, 77–84.

Claerbout, J. F. 1997. *GEOPHYSICAL EXPLORATION MAPPING (GEM): Environmental soundings image enhancement.* At http://sepwww.stanford.edu/sep/prof.

Freedman, D., & Diaconis, P. 1981a. On the histogram as a density estimator. *Zeitschrift fur Wahrscheinlichkeits Theorie und verwandte Gebiete*, **57**, 453–476.

Freedman, D., & Diaconis, P. 1981b. On the maximum deviation between the histogram and the underlying density: L_2 theory. *Zeitschrift fur Wahrschein-*

lichkeits Theorie und verwandte Gebiete, **58**, 139–167.

Gouveia, W., & Scales, J. 1998. Bayesian Seismic Waveform Inversion: Estimation and Uncertainty Analysis. *Journal of Geophysical Research*, **103**(B2), 2759–2779.

Hestenes, M., & Stiefel, E. 1952. Methods of conjugate gradients for solving linear systems. *NBS J. Research*, **49**, 409–436.

Oldenburg, D.W., Yaoguo, L., & Ellis, R.G. 1997. Inversion of geophysical data over a copper gold porphyry deposit: a case history for Mt. Milligan. *Geophysics*, **62**, 1419–1431.

Rodger, K., & Wahr, J. 1993. Inference of core-mantle boundary topography from ISC PcP and PKP travel times. *Geophysical Journal International*, **115**, 991–1011.

Scales, J. A. 1987. Tomographic inversion via the conjugate gradient method. *Geophysics*, **52**, 179–185.

Stuart, A., & Ord, J. 1987. *Kendall's Theory of Advanced Statistics, 5th Edition, Volume I.* N.Y.: Oxford University Press.

APPENDIX A: Conjugate Gradient Method for Weighted Least-Squares

Conjugate gradient can be extended to the least squares solution of arbitrary linear systems. When you weight the data the normal equations are:

$$A^T R A \mathbf{x} = A^T R \mathbf{h} \tag{A1}$$

Now, the algorithm for weighted least squares will be based on the Conjugate Gradient Least Squares algorithm (CGLS) of Hestenes & Stiefel (1952). Note that R is a diagonal matrix, with non-negative values. Therefore, $R = R^{1/2}R^{1/2}$ and $R^T = R$. Now we decompose R:

$$\begin{array}{lll} A^T R^{1/2} R^{1/2} A\mathbf{x} & = A^T R^{1/2} R^{1/2} \mathbf{h} & \Leftrightarrow \\ (R^{1/2}A)^T R^{1/2} A\mathbf{x} & = (R^{1/2}A)^T R^{1/2}\mathbf{h} & \Leftrightarrow \\ B^T B\mathbf{x} & = B^T \mathbf{y} & \end{array}$$

where $B = R^{1/2}A$, and $\mathbf{y} = R^{1/2}\mathbf{h}$.

The algorithm for weighted least squares will be: Choose $\mathbf{x}_0$. Put

- $\mathbf{s}_0 = \mathbf{y} - B\mathbf{x}_0 = R^{1/2}(\mathbf{h} - A\mathbf{x}_0)$
- $\mathbf{r}_0 = \mathbf{p}_0 = B^T\mathbf{s}_0 = A^T R^{1/2}\mathbf{s}_0$

Then for $i = 0, 1, 2, \ldots$

$$\begin{array}{ll} \mathbf{q}_i & = B\mathbf{p}_i = R^{1/2}A\mathbf{p}_i \\ \alpha_{i+1} & = \frac{(\mathbf{r}_i, \mathbf{r}_i)}{(\mathbf{q}_i, \mathbf{q}_i)} \\ \mathbf{x}_{i+1} & = \mathbf{x}_i + \alpha_{i+1}\mathbf{p}_i \\ \mathbf{s}_{i+1} & = \mathbf{s}_i - \alpha_{i+1}\mathbf{q}_i \\ \mathbf{r}_{i+1} & = B^T\mathbf{s}_{i+1} = A^T R^{1/2}\mathbf{s}_{i+1} \\ \beta_{i+1} & = \frac{(\mathbf{r}_{i+1}, \mathbf{r}_{i+1})}{(\mathbf{r}_i, \mathbf{r}_i)} \\ \mathbf{p}_{i+1} & = \mathbf{r}_{i+1} + \beta_{i+1}\mathbf{p}_i \end{array}$$

By moving the calculation of q_i to the top of the loop, it is not necessary to define a starting value for q.