Simultaneous least squares deconvolution and kriging using conjugate gradients

Jonathan Kane
William Rodi
M. Nafi Toksöz
ERL, MIT

Abstract

Least squares deconvolution is a method used to sharpen tomographic images of the earth by undoing the bandlimiting effects imposed by a seismic wavelet. Kriging is a method used by geoscientists to extrapolate and interpolate sparse data sets. These two methodologies have traditionally been kept separate and viewed as unrelated fields of research. We demonstrate the connection between these methods by deriving them both as examples of linear inversion. By posing the methods in this way we can define a joint inverse problem in which observed values of reflectivity in wells are used to improve deconvolution, and, conversely, seismic data is used to help extrapolate well data.

Solving this joint problem involves the solution of large sparse sets of linear equations. Due to the structure of the problem, the conjugate gradients method is ideal to perform the solution. Preliminary results show that convergence to a solution for a 3-D problem is fast and accurate, requiring only a few iterations.

This methodology can be of great use to interpreters by sharpening the post stack image as well as helping to tie seismic data to wells.

Introduction

In petroleum exploration knowledge of the subsurface is usually derived from two sources: well data and seismic data. The former gives very accurate but very sparse information about petrophysical parameters at a few locations. The latter provides a larger scale image of the subsurface with much lower resolution. A proper combination of these two data sets should result in a more optimal image of the subsurface.

In this work the seismic data is in post-stack form. Ideally this would mean that the convolutional seismic model holds. Thus the seismic data is the result of a linear operation (a convolution) of a seismic wavelet on a reflectivity sequence. If the seismic wavelet is known a reasonable objective might be to deconvolve the wavelet out of the seismic data to obtain the reflectivity, thus obtaining a better image of subsurface material discontinuities. To do this we use the method known as least squares deconvolution (Yilmaz, 1987).

Given velocity and density logs, we can convert these from depth into reflectivity as a function of two-way travel time. With these reflectivity logs we can perform *kriging* (Cressie, 1993; Isaaks and Srivastava, 1989). This is a geostatistical method that provides statistically optimal estimates of parameters of interest given only sparse samples and probabilistic knowledge of the spatial fluctuations in the parameter field.

We define a statistically optimal method for integrating these two data sets in the form of linear inverse theory. Reflectivity logs and post-stack seismic data are simultaneously kriged and deconvolved to obtain an optimal image of reflectivity.

We use the conjugate gradients algorithm to construct the image, avoiding the need to solve a large system of linear equations and requiring only the iterative application of a forward operator.

The Joint Linear Inverse Problem

Let $\mathbf{x}$ be a random vector representing the 3-D reflectivity of the subsurface. 3-D seismic data will be represented by $\mathbf{s}$ and is related to the reflectivity by the following equation:

$$\mathbf{s} = \mathbf{Hx} + \mathbf{n_H}.$$

$\mathbf{H}$ is a block diagonal matrix where each diagonal block is a Toeplitz matrix performing convolution of each vertical reflectivity profile with a seismic wavelet. $\mathbf{n_H}$ is a random vector independent of $\mathbf{x}$. It represents the noise contaminating the seismic data, as well as the inherent inaccuracy of the convolutional operator $\mathbf{H}$. Due to lack of any information to its correlation structure, we assume the noise to be stationary white (uncorrelated with equal variance, $\sigma^2_{\mathbf{H}}$, at all points).

The method of kriging presented here is derived somewhat differently than that typically presented in the geostatistical literature. We derive the kriging equations from linear inversion theory where some level of noise is assumed in the data to be kriged. We let $\mathbf{d}$ represent the observed reflectivity in wells. These data are related to the reflectivity field via the following equation:

$$\mathbf{d} = \mathbf{Px} + \mathbf{n_P}.$$

P is a rectangular matrix made up of a subset of the rows of the identity matrix. It "picks" the observed data out of the reflectivity field $\mathbf{x}$. Again, $\mathbf{n_P}$ is a random vector that we assume to be uncorrelated with $\mathbf{x}$ and white with variance $\sigma^2_{\mathbf{P}}$.

In kriging we also assume a-priori a certain amount of smoothness in the field to be estimated. This can be represented as attempting to minimize the result of a difference operator acting on the reflectivity field. This can be understood by seeing that, by differencing a field, one effectively roughens it. Thus by minimizing the energy of the differenced field, one smooths the original field. We represent this difference operator by $\mathbf{L}$. We then have

$$\mathbf{Lx} = \mathbf{n_L}.$$

The noise $\mathbf{n_L}$ is white and has variance $\sigma^2_{\mathbf{L}}$.

Since $\mathbf{n_H}$, $\mathbf{n_P}$, and $\mathbf{n_L}$ are all random vectors, we need to assign them probability distributions. The most tractable and convenient distribution to use is the multi-dimensional Gaussian, thus we give our noise vectors this distribution:

$$f_{\mathbf{H}}(\mathbf{n_H}|\mathbf{x},\mathbf{s}) = \frac{1}{(2\pi\ det(\mathbf{C_H}))^{\frac{N}{2}}}\ exp[-\frac{1}{2}(\mathbf{s}-\mathbf{Hx})^T\mathbf{C_H^{-1}}(\mathbf{s}-\mathbf{Hx})] \tag{1}$$

$$f_{\mathbf{P}}(\mathbf{n_P}|\mathbf{x},\mathbf{d}) = \frac{1}{(2\pi\ det(\mathbf{C_P}))^{\frac{N}{2}}}\ exp[-\frac{1}{2}(\mathbf{d}-\mathbf{Px})^T\mathbf{C_P^{-1}}(\mathbf{d}-\mathbf{Px})] \tag{2}$$

and

$$\begin{aligned} f_{\mathbf{L}}(\mathbf{n_L}|\mathbf{x}) &= \frac{1}{(2\pi\ det((\frac{1}{\sigma^2_{\mathbf{L}}}\mathbf{L}^T\mathbf{L})^{-1}))^{\frac{N}{2}}}\ exp[-\frac{1}{2\sigma^2_{\mathbf{L}}}(\mathbf{x}^T\mathbf{L}^T\mathbf{Lx})] \\ &= f_{prior}(\mathbf{x}) = \frac{1}{(2\pi\ det(\mathbf{C_L}))^{\frac{N}{2}}}\ exp[-\frac{1}{2}\mathbf{x}^T\mathbf{C_L^{-1}}\mathbf{x}] \end{aligned} \tag{3}$$

The notation $f(\mathbf{a}|\mathbf{b})$ represents the probability of a vector $\mathbf{a}$ *conditioned* on an observed vector $\mathbf{b}$. As mentioned above, we assume $\mathbf{n_H}$ and $\mathbf{n_P}$ to be identically distributed white noise. Thus $\mathbf{C_H} = \sigma^2_{\mathbf{H}}\mathbf{I}$ and $\mathbf{C_P} = \sigma^2_{\mathbf{P}}\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

Our decision to describe the noise in equations 1, 2, and 3 as multivariate Gaussian should not be ad-hoc, but rather based on prior observations of the noise in the system. If the true noise departs significantly from the multi-Gaussian distribution, the results of our estimation will suffer by making the multi-Gaussian assumption. In this work we did not have prior knowledge of the distribution of the noise vectors and thus fell back to the multi-Gaussian hypothesis because it also represents the least informative state of our knowledge about a system. Equation 3 is what is known in the Bayesian inversion community as the *prior*. It describes our state of knowledge of the quantity we are estimating before the experiment occurs.

To find the optimal reflectivity field $\mathbf{x}$ we must jointly maximize the probability in equations 1, 2, and 3. If we visualize these three functions as three hyper-surfaces over the possible values of $\mathbf{x}$, then we can multiply the three functions together and renormalize (keeping its integral equal to unity) to obtain the *posterior* distribution:

$$\begin{aligned} f_{post}(\mathbf{x}|\mathbf{s},\mathbf{d}) &= f_{\mathbf{H}}(\mathbf{n_H}|\mathbf{x},\mathbf{s}) f_{\mathbf{P}}(\mathbf{n_P}|\mathbf{x},\mathbf{d}) f_{\mathbf{prior}}(\mathbf{x}) \\ &= k\ exp\left[-\epsilon(\mathbf{x}|\mathbf{s},\mathbf{d})\right] \end{aligned} \tag{4}$$

where

$$\epsilon(\mathbf{x}|\mathbf{s},\mathbf{d}) = \frac{1}{2}\left(\frac{1}{\sigma^2_{\mathbf{H}}}(\mathbf{s}-\mathbf{Hx})^T(\mathbf{s}-\mathbf{Hx}) + \frac{1}{\sigma^2_{\mathbf{P}}}(\mathbf{d}-\mathbf{Px})^T(\mathbf{d}-\mathbf{Px}) + \frac{1}{\sigma^2_{\mathbf{L}}}\mathbf{x}^T\mathbf{L}^T\mathbf{Lx}\right) \tag{5}$$

and k is an appropriate normalizing constant.

By finding the maximum of $f_{post}(\mathbf{x}|\mathbf{s},\mathbf{d})$ we have found the most probable model of $\mathbf{x}$ and solved the inverse problem. Maximizing equation 4 corresponds to minimizing equation 5. If the operators $\mathbf{H}$ and $\mathbf{P}$ are linear, then we also know that equation 5 is a convex function, i.e. it only has one minimum. To find this minimum we first take the derivative of equation 5 with respect to $\mathbf{x}$ and set the result equal to zero:

$$\frac{\partial\epsilon(\mathbf{x}|\mathbf{s},\mathbf{d})}{\partial\mathbf{x}} = (\frac{1}{\sigma^2_{\mathbf{H}}}\mathbf{H}^T\mathbf{H} + \frac{1}{\sigma^2_{\mathbf{P}}}\mathbf{P}^T\mathbf{P} + \frac{1}{\sigma^2_{\mathbf{L}}}\mathbf{L}^T\mathbf{L})\mathbf{x} - \frac{1}{\sigma^2_{\mathbf{H}}}\mathbf{H}^T\mathbf{s} - \frac{1}{\sigma^2_{\mathbf{P}}}\mathbf{P}^T\mathbf{d} = 0$$

Rearranging, we have

$$(\frac{1}{\sigma^2_{\mathbf{H}}}\mathbf{H}^T\mathbf{H} + \frac{1}{\sigma^2_{\mathbf{P}}}\mathbf{P}^T\mathbf{P} + \frac{1}{\sigma^2_{\mathbf{L}}}\mathbf{L}^T\mathbf{L})\mathbf{x} = \frac{1}{\sigma^2_{\mathbf{H}}}\mathbf{H}^T\mathbf{s} + \frac{1}{\sigma^2_{\mathbf{P}}}\mathbf{P}^T\mathbf{d}.$$

We can multiply both sides of this equation by $\sigma^2_{\mathbf{P}}$ without changing the solution. This gives

$$(\lambda_1\mathbf{H}^T\mathbf{H} + \mathbf{P}^T\mathbf{Px} + \lambda_2\mathbf{L}^T\mathbf{L})\mathbf{x} = \lambda_1\mathbf{H}^T\mathbf{s} + \mathbf{P}^T\mathbf{d}, \tag{6}$$

where $\lambda_1 = \frac{\sigma^2_{\mathbf{P}}}{\sigma^2_{\mathbf{H}}}$ and $\lambda_2 = \frac{\sigma^2_{\mathbf{P}}}{\sigma^2_{\mathbf{L}}}$. λ_1 expresses the relative importance of fitting the seismic data to honoring the well data, λ_2 expresses the relative importance of smoothness to honoring well data. Put another way, low λ_1 means that when we perform deconvolution at a well location, we weight the observed reflectivity data over the deconvolved result at that location. The limit of $\lambda_2 \mapsto 0$ forces interpolation to go through the observed well data values (i.e. true kriging). We group the three variance terms into two λ terms because it lessens the amount of computations in the conjugate gradients algorithm below. Readers familiar with inverse theory will recognize equation 6 as the *normal* equations of least squares inversion theory.

By setting values for λ_1 and λ_2 and then solving the preceding system of linear equations one obtains the jointly kriged and least squares deconvolved reflectivity section.

Structure and Numerical Operation of A

Let us pose the estimation of $\mathbf{x}$ as a matrix inversion by setting $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A} = (\lambda_1\mathbf{H}^T\mathbf{H} + \mathbf{P}^T\mathbf{Px} + \lambda_2\mathbf{L}^T\mathbf{L})$ and $\mathbf{b} = (\lambda_1\mathbf{H}^T\mathbf{s} + \mathbf{P}^T\mathbf{d})$. Then we must numerically perform the equivalent of $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ to obtain our solution. There are many ways to perform this solution (Strang, 1986; Trefethen and III, 1997). However the sheer size of $\mathbf{A}$ for a problem of realistic size rules out all but a few methods. In this work we invert a small 3-D cube of seismic data which consists of 1836 seismic traces and 100 samples per traces ($100 \times 51 \times 36$). The $\mathbf{A}$ matrix for such a problem would then be 183,600 by 183,600. That is 33,708,960,000 entries and at 8 bytes per entry we would require 269,671,680,000 bytes just to store the matrix! Most modern computers don't have 270 Gb of RAM so direct creation and inversion of this hypothetical matrix is impossible. There is, however, an excellent property of this matrix that we should and do exploit. It is exceedingly sparse. Most of its entries are 0. This prompts us to turn to iterative methods of solving a system of linear equations. The method we use is conjugate gradients (CG) (Strang, 1986; Trefethen and III, 1997).

Iterative methods require the repeated application of the $\mathbf{A}$ matrix. We show how this can be quickly executed by looking individually at each component of $\mathbf{A}$. First, however we mention how the vector $\mathbf{x}$ is stored on a computer. $\mathbf{x}$ represents a 3-D cube of reflectivity profiles with the dimensions listed above. $\mathbf{x}$ is stored as a 1-D vector in a profile by profile fashion. Thus we put 1836 reflectivity profiles tip-to-tail into $\mathbf{x}$ to make a 183,600 entry long vector that takes 1,468,800 bytes to store in double precision.

The first component of $\mathbf{A}$ is $\mathbf{H}^T\mathbf{H}$. This has the effect of going through each profile and convolving it with the autocorrelation function of the seismic wavelet. In looking at $\mathbf{H}^T\mathbf{H}$ as a matrix, it is a block-diagonal matrix with each diagonal block an identical toeplitz matrix. Since each diagonal block is identical, we need to store only one of them. However, since convolution is done more efficiently in the Fourier domain for long seismic wavelets, we simply store the Fourier transform of the autocorrelation function (i.e. the power spectrum) in memory. We have thus compressed a matrix with 33,708,960,000 entries into a single vector with only 183,600 entries! When operating $\mathbf{H}^T\mathbf{H}$ on $\mathbf{x}$ we need only cycle through $\mathbf{x}$ each 100 entries at a time, transform those 100 values to the Fourier domain, multiply by the power spectrum of $\mathbf{H}^T\mathbf{H}$, and then transform back. That becomes approximately $O(1836 \times 100 \log_2 100)$ operations for one application of $\mathbf{H}^T\mathbf{H}$.

As mentioned above, we model $\mathbf{L}$ as a multidimensional differencing operator. This essentially says that we believe our discrete field of values is correlated in such a way that, by applying our differencing operator, we would obtain discrete white noise. The effect of this operator is to impose some smoothness on our estimated $\mathbf{x}$. Storage of $\mathbf{L}$ on a computer amounts to storing the indices and values of the differencing operator (i.e. one row of the non-zero elements of the $\mathbf{L}$ matrix). For low order differencing operators, this amounts to just a few entries in 2 vectors. The action of $\mathbf{Lx}$ then amounts to a fast explicit convolution. The $\mathbf{L}^T$ operator is then applied by reversing the order of the values in the vector of $\mathbf{L}$ coefficients and applying another convolution to the output of $\mathbf{Lx}$. For our purposes $\mathbf{L}^T\mathbf{Lx}$ amounts to $O(2 \times (\nu \times 183,600))$ operations, where ν is the number of elements in the differencing operator (in our case $\nu = 11$).

Finally, we come to $\mathbf{P}^T\mathbf{P}$, the simplest of the operators to apply. If viewed as a large square matrix it would have all entries equal to zero except at a few sequences of 100 values along the diagonal, which would equal one. This operator has the effect of selecting reflectivity profiles at a few locations and adding them to the output of $\mathbf{H}^T\mathbf{H}$ and $\mathbf{L}^T\mathbf{L}$ at those locations. This can be done explicitly and requires no computation.

Conjugate Gradients

Rather than derive the conjugate gradients algorithm, we will instead briefly describe it and its major properties and then show the algorithm. (Press et al., 1995) provide good introductions to the subject.

The conjugate gradients method, like many iterative methods, constructs what is known as a *Krylov subspace* of vectors that has the property that, at the k^{th} iteration the quantity $(\hat{\mathbf{x}} - \mathbf{x}_k)^T\mathbf{A}(\hat{\mathbf{x}} - \mathbf{x}_k)$ is minimum, where $\hat{\mathbf{x}}$ is the true solution to the equation $\mathbf{Ax} = \mathbf{b}$.

Furthermore, the residual $\mathbf{r}_k = \mathbf{b} - \mathbf{Ax}_k$ is orthogonal to all previous residuals: $\mathbf{r}_j^T\mathbf{r}_k^T = 0$, $j < k$. Also, the search direction $\mathbf{p}_k$ is $\mathbf{A}$ orthogonal to all previous search directions $\mathbf{p}_j^T\mathbf{A}\mathbf{p}_k^T = 0$, $j < k$.

When applied to a positive definite matrix, the conjugate gradients algorithm can be viewed as minimizing the objective function $\Phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{b}$. Each iteration lowers the objective function and brings the vector $\mathbf{x}$ closer to the minimizing solution. This is one of the attractions of the method. It can be stopped any time before the final solution is obtained and still give a useful answer. This is not the case with other standard matrix inversion algorithms, such as Gaussian elimination. Another attraction has already been mentioned: the ability to invert a huge matrix (assuming it has a nice structure).

The algorithm is as follows (Press et al., 1995; Trefethen and III, 1997):

1. Set $\mathbf{x}_0 = \textit{arbitrary}$, and $\mathbf{r}_0 = \mathbf{b} - \mathbf{Ax}_0$

2. $\beta_j = \mathbf{r}_{j-1}^T\mathbf{r}_{j-1}/\mathbf{r}_{j-2}^T\mathbf{r}_{j-2}$ (except $\beta_1 = 0$)

3. $\mathbf{p}_j = \mathbf{r}_{j-1} + \beta_j\mathbf{p}_{j-1}$ (except $\mathbf{p}_1 = \mathbf{r}_0$)

4. $\alpha_j = \mathbf{r}_{j-1}^T \mathbf{r}_{j-1} / \mathbf{p}_j^T A \mathbf{p}_j$

5. $\mathbf{x}_j = \mathbf{x}_{j-1} - \alpha_j \mathbf{p}_j$

6. $\mathbf{r}_j = \mathbf{r}_{j-1} - \alpha_j A \mathbf{p}_j$

7. Repeat step 2.

The conjugate gradients algorithm is applicable only to symmetric positive definite matrices. The **A** matrix defined above fits into this category, but in the case where we do not have this structure (such as in the next section), we must turn to an alternative technique called the *bi-conjugate gradients* algorithm (BCG). Rather than describe it in detail, we reference (Trefethen and III, 1997). It essentially generalizes the CG algorithm to non-symmetric, non-positive definite matrices by constructing two bi-orthogonal gradients vectors rather than one. This doubles the number of computations in each loop of the algorithm, but it can be shown that in the least squares problem the BCG method actually converges to a solution much faster. It was not obvious to us which method was optimal, so we chose the CG algorithm for the inverse problem and reserved the BCG method for the generation of a random field (described below).

Synthetic Data Set Simulation

It would be useful to validate our inversion method before applying it to a real data set. To do this we must create a three-dimensional synthetic model of reflectivity, convert it to seismic data, and extract some vertical profiles from the reflectivity cube to use as well logs. We will also add white noise with known variance to the synthetic well and seismic data.

We would like to specify the correlation in our synthetic field in the form of the **L** matrix defined above. The question is: How do we generate a random field that generates white noise when operated on by **L**? Typical methods to construct a realization of such a random field include Fourier methods (Cressie, 1993), wavelet methods (Daniel, 1997), turning bands (Cressie, 1993), **LU** decomposition (Cressie, 1993), mid-point displacement (Daniel, 1997) (for certain kinds of random fields), and random walk methods (Fishman, 1996). We propose here a new method that, to our knowledge, has never been used for this purpose. We apply the BCG method to the following problem:

$$\mathbf{L}\mathbf{x} = \mathbf{w}, \tag{7}$$

where **L** is the differencing operator defined above and **w** is a vector of discrete white noise. We need to use the BCG method because the **L** operator is quite possibly not symmetric or positive definite.

A realization of a 3-D random field via this method is shown in Figure 1. The **L** operator used to generate this field consists of a first difference vertically and a fourth difference horizontally. We choose it this way to make a field that is more strongly correlated in the horizontal direction than vertically, the way we might expect geological sequences to be. One might say at this point "Wait a minute, reflectivity sequences are actually vertically *anti*-correlated!" This is true but our approximation is actually not that bad. A random field whose 1^{st} order difference is white noise is correlated at large scales (low frequencies) but actually becomes anti-correlated at small scales (high frequencies). This can be intuitively understood by remembering that a *differencing* operator is not the same as a *differential operator*. A differential operator multiplies by $j\omega$ in the frequency domain, while the first difference operator multiplies by $sin(\frac{\omega}{2})$ in the frequency domain. These two operators coincide only for low frequencies. We also mention that the field in Figure 1 is created with $\sigma_{\mathbf{L}}^2 = 0.1$.

We use our synthesized random field as our synthetic model of reflectivity for the purpose of illustrating our inversion method in the next section. We first convert the reflectivity to seismic and extract well logs. Then we add white noise with $\sigma_{\mathbf{H}}^2 = 10$ and $\sigma_{\mathbf{P}}^2 = 0.01$ and show our resulting noisy seismic and well data in Figure 2.

Inversion of a Synthetic Data Set

Starting with the noisy data sets we perform inversion twice, once assuming we know the true correlation structure of our data set, and a second time where we assume it to be white noise (the traditional deconvolution assumption) by letting $\sigma_L^2 \mapsto \infty$. Figure 3 shows the inversion with correct assumptions. We notice first the smoothness of the inverted reflectivity field. A geophysicist might complain about this and want more detail, but this is actually the best that can be done. The smoothing of the inverted field essentially minimizes the inversion error introduced by the noise in the data domain. On the positive side we see a slight sharpening of the features globally, but a dramatic sharpening in the vicinity of the well logs (as we would hope).

If we increase the assumed variance ($\sigma_{\mathbf{L}}^2$) of the reflectivity field, we approach the limit where the result of the inversion is not smoothed at all. Thus the seismic data solely determines the result of the inversion. We see the result of this in Figure 4. Notice that the noise in the seismic data has caused erroneous high frequency features in the estimated field. Also, there is no longer any constraint enforcing horizontal continuity. This illustrates why anyone doing any inverse problem should consider smoothing, whether they correctly know the correlation structure of the true field or not.

As a side note we show the decrease of the objective function vs. iteration in Figure 5. This plot is typical of all runs of the CG algorithm in 3-D and thus only shown once. It illustrates the excellent convergence of the method in few iterations (≈ 50, which takes about 50 seconds).

Inversion of a Real Data Set

We now move on to the real seismic data and reflectivity logs. The original seismic data cube consisted of undulating events. To facilitate the use of our methods, we needed to first flatten the seismic data at a particular prominent reflector and adjust all well logs accordingly. In a real life application, the idea would be to perform the inversion on the flattened data, and then reverse the flattening process on the resulting estimated reflectivity field. As is obvious in looking at the seismic data (Figure 6), we see that the flattening was not done well. This accounts for the ratty appearance of the seismic data, where some seismic traces are shifted disproportionately to others. Despite the obvious shortcomings of the data set, we proceed with the inversion.

To perform the inversion on the real data set we need to set values for $\sigma_{\mathbf{H}}^2$, $\sigma_{\mathbf{P}}^2$, and $\sigma_{\mathbf{L}}^2$. Unfortunately we do not have this information a-priori. The best we can do is vary these parameters and draw conclusions from the different results. We start off by setting $\sigma_{\mathbf{H}}^2 = 10$, $\sigma_{\mathbf{P}}^2 = 0.01$, and $\sigma_{\mathbf{L}}^2 = 0.1$ for the first of the inversion runs. We see the results in Figure 7. Comparing this to the original seismic data, we see a slight improvment in terms of the deconvolution, but not as much as we hoped for. This may be due to the poor data quality and high noise level. Thus the model that best fits the seismic data and has horizontal continuity is a necessarily smooth one. On the positive side we see increased horizontal continuity and suppression of flattening artifacts.

In the next run we keep $\sigma_{\mathbf{H}}^2$ and $\sigma_{\mathbf{P}}^2$ the same but set $\sigma_{\mathbf{L}}^2 = 10,000$. These parameters are the same as the second run of the synthetic inversion. Again they correspond to assuming no correlation in the underlying field and relying on the seismic data for inversion. This does a trace by trace deconvolution with no horizontal continuity constraint on the inverted field. We see the result in Figure 8. Notice how the artifacts (vertically shifted seismograms) from the seismic data are preserved in the inverted field.

For the final run we set $\sigma_{\mathbf{H}}^2 = 10,000$, $\sigma_{\mathbf{P}}^2 = 0.01$ and $\sigma_{\mathbf{L}}^2 = 0.1$. This practically ignores the seismic data in doing the inversion and relies only on the well data and the smoothness constraints. This is equivalent to kriging with measurement noise in the well data. We see in Figure 9 that away from the wells the value of reflectivity drops to zero. We expected this but not quite so dramatically. Having $\mathbf{L}$ be a 4^{th} difference horizontally, we expected to see long distance correlations in the kriging. We *do* see them in Figure 7. More thought and investigation needs to be done to understand what is happening numerically to give such results.

Conclusions

This work represents a statistically optimal method for integrating well and seismic data into an image of the subsurface. Computationally it is exceedingly fast. The results obtained, however, were only marginal improvements on the seismic. This may be a result of the poor data quality or a lack of understanding into the numerical method we are using. In particular, the **L** operator is not invertible and has zero eigenvalues. Thus when we weight $\sigma_{\mathbf{H}}^2$ large, we may be running into numerical problems with the conjugate gradients algorithm. Future work will probably involve pre-conditioning or otherwise modifying the **L** matrix into a more stable form.

In addition to representing the optimal way to integrate two data sets, the method also has the practical application of helping a seismic interpreter tie wells to seismic data. This could be done by increasing $\sigma_{\mathbf{H}}^2$ until the well data appears more prominently.

Finally, we have also presented a new method for generating random fields in multi-dimensions via the bi-conjugate gradients algorithm. We are unaware of anyone in any discipline making use of this algorithm for this purpose. Future work may lead to using the bi-conjugate gradients algorithm to synthesize posterior realizations of random fields or perhaps even non-Gaussian random fields.

Acknowledgements

We would like to thank Christie Callendar, Texaco, and Geco-Prakla for the seismic data used in this work.

References

Noel A. C. Cressie. *Statistics for Spatial Data.* John Wiley and Sons, Inc., 1993.

Mike M. Daniel. *Multiresolution Statistical Modeling with Application to Modelling Groundwater Flow.* PhD thesis, Massachusetts Institute of Technology, 1997.

George S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications.* Springer, 1996.

Edward H. Isaaks and R. Mohan Srivastava. *Introduction to Applied Geostatistics.* Oxford University Press, 198 Madison Avenue, New York, New York 10016-4314, first edition, 1989.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C.* Cambridge University Press, second edition, 1995.

Gilbert Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley, Massachusetts 02181, first edition, 1986.

Lloyd N. Trefethen and David Bau III. *Numerical Linear Algebra.* Society for Industrial and Applied Mathematics, Philadelphia, 1997.

Özdoğan Yilmaz. *Seismic Data Processing.* Society of Exploration Geophysicists, P.O. Box 702740 Tulsa, OK, 74170, first edition, 1987.

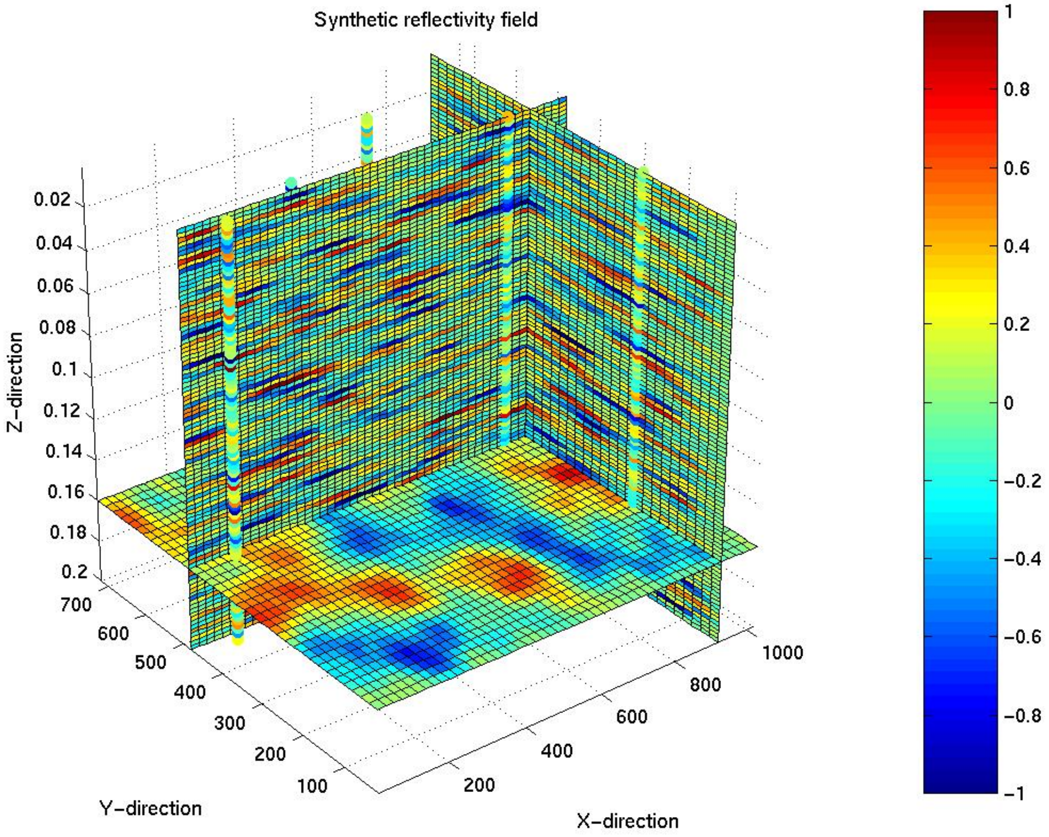

Figure 1: Synthetic reflectivity field

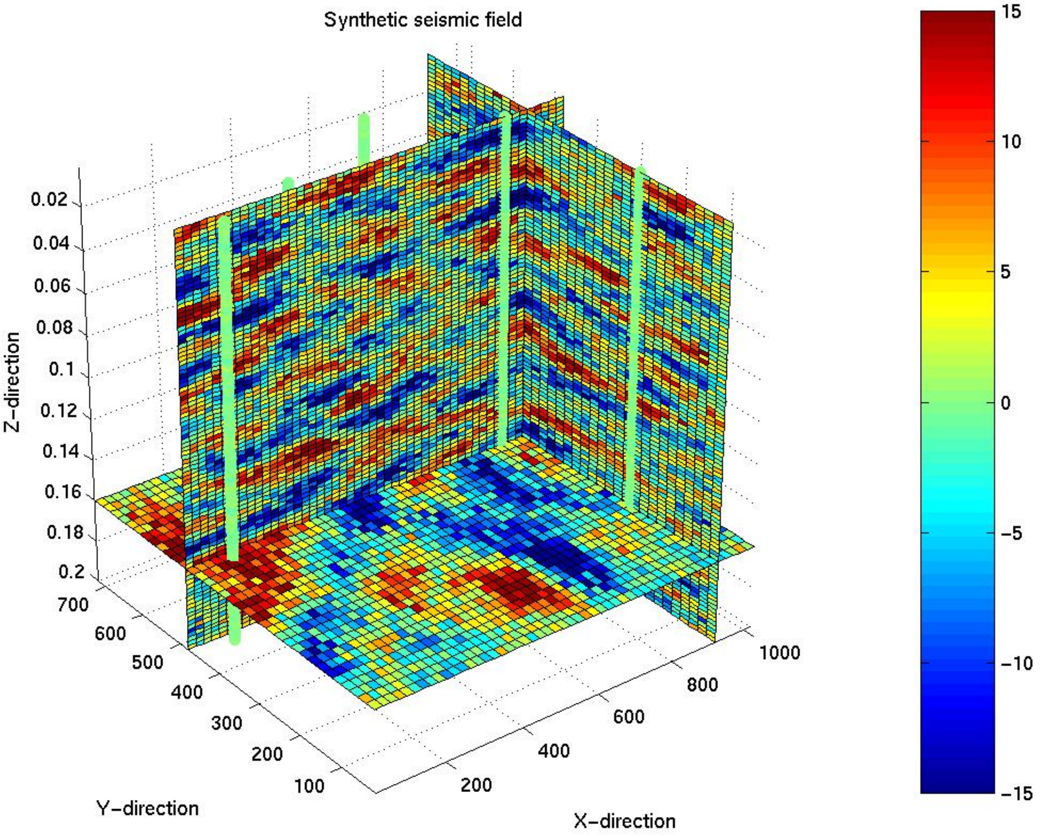

Figure 2: Synthetic seismic data

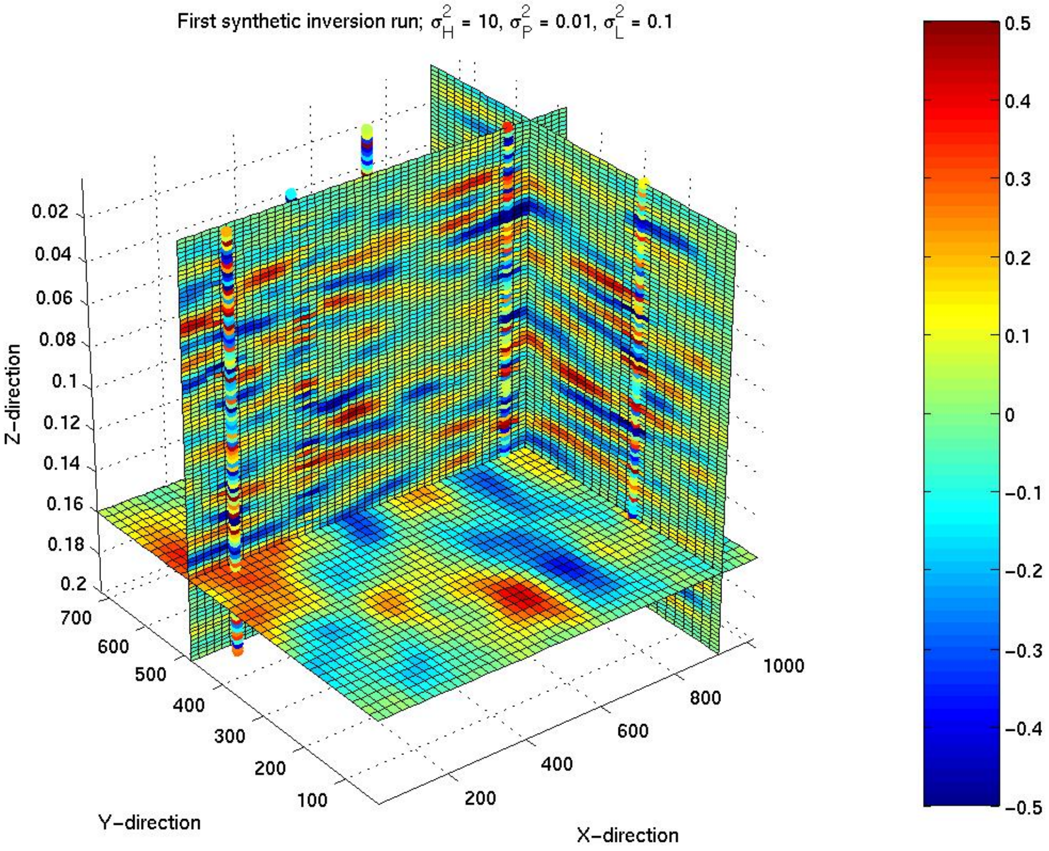

Figure 3: Synthetic Inversion 1

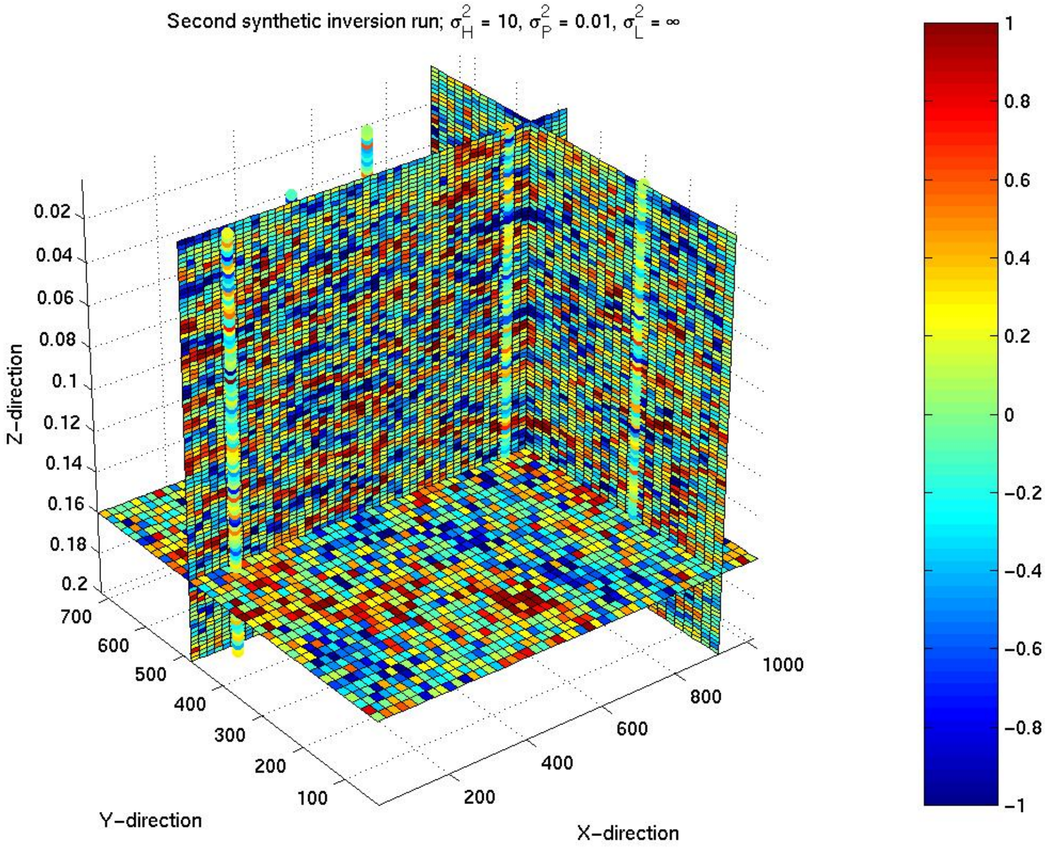

Figure 4: Synthetic Inversion 2

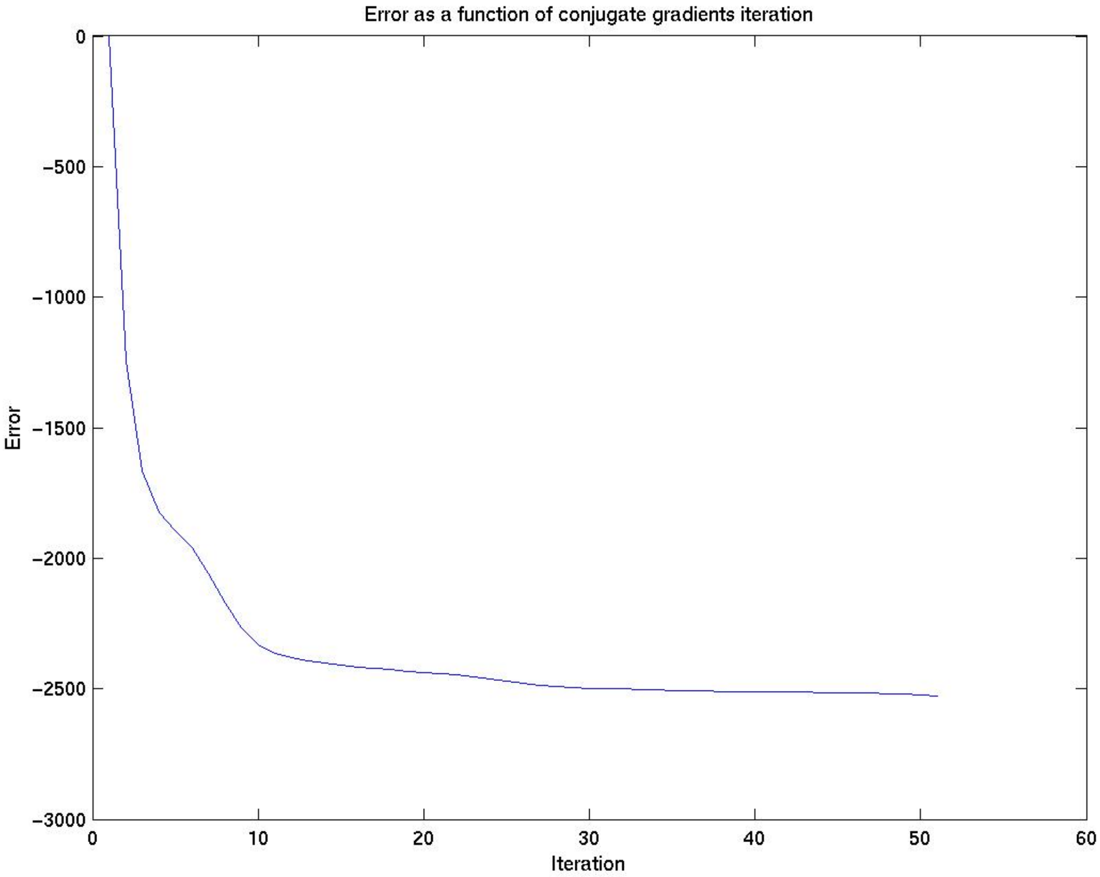

Figure 5: Error vs. iteration

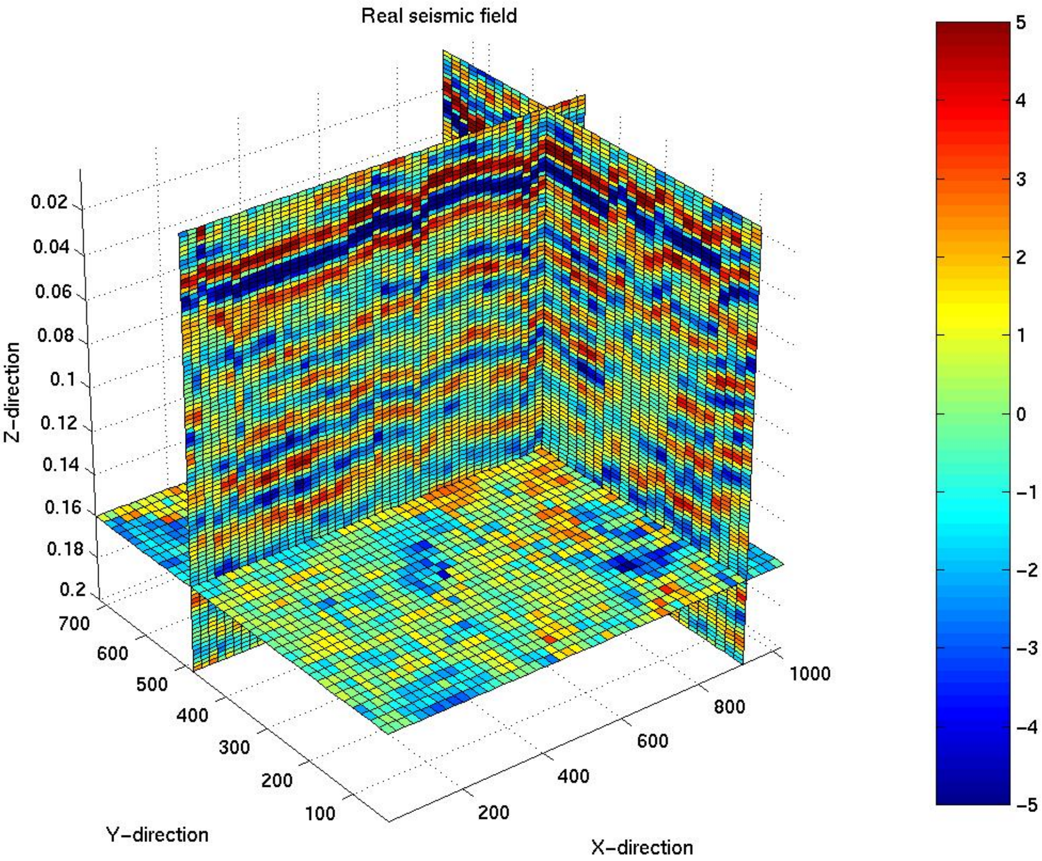

Figure 6: Real seismic data

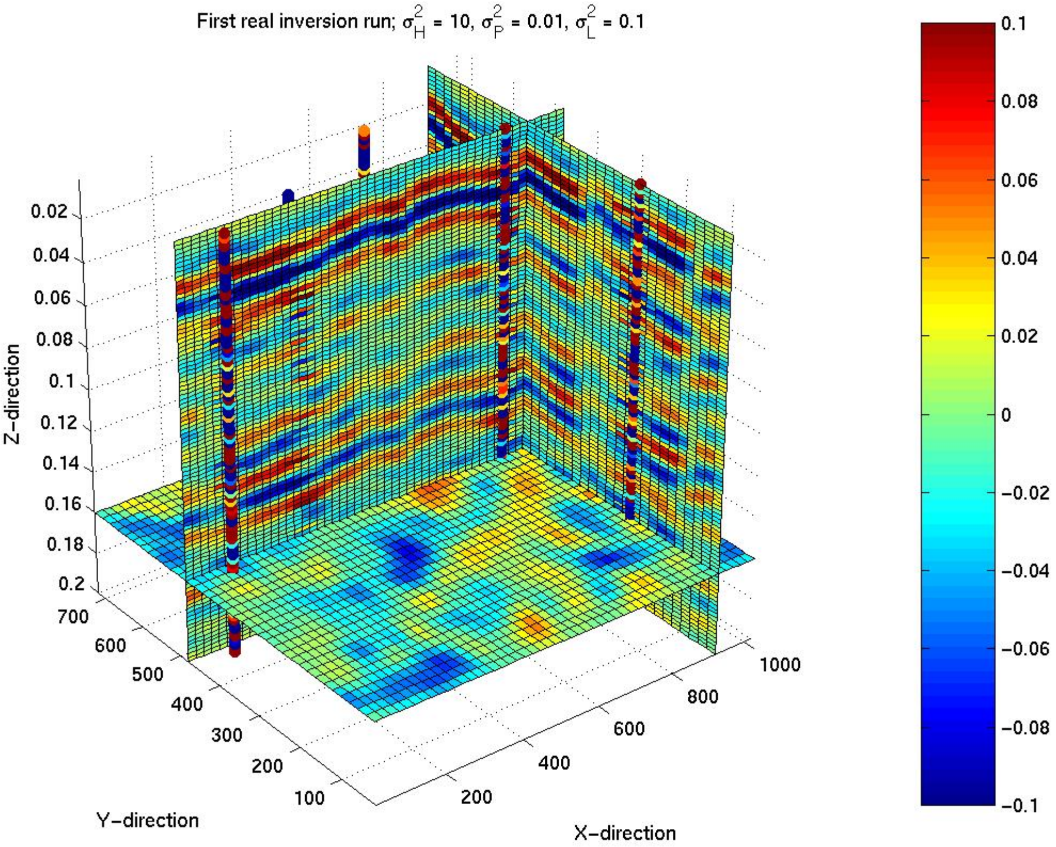

Figure 7: Real Inversion 1

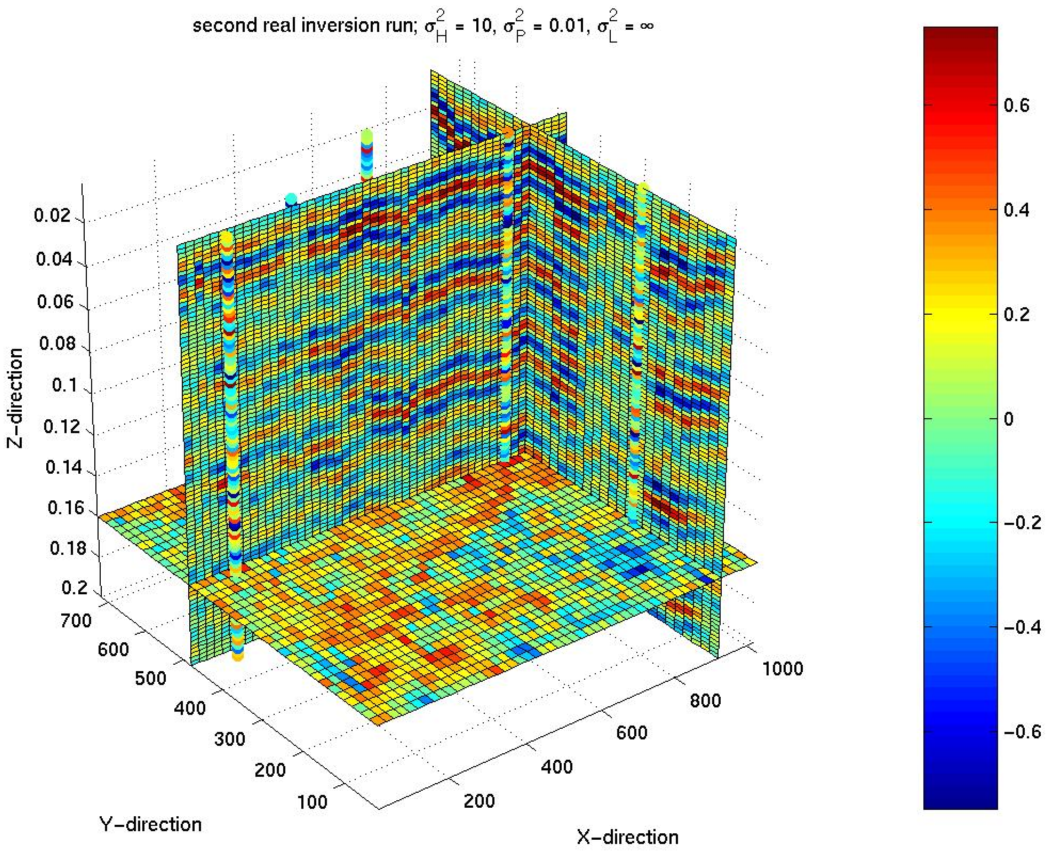

Figure 8: Real Inversion 2

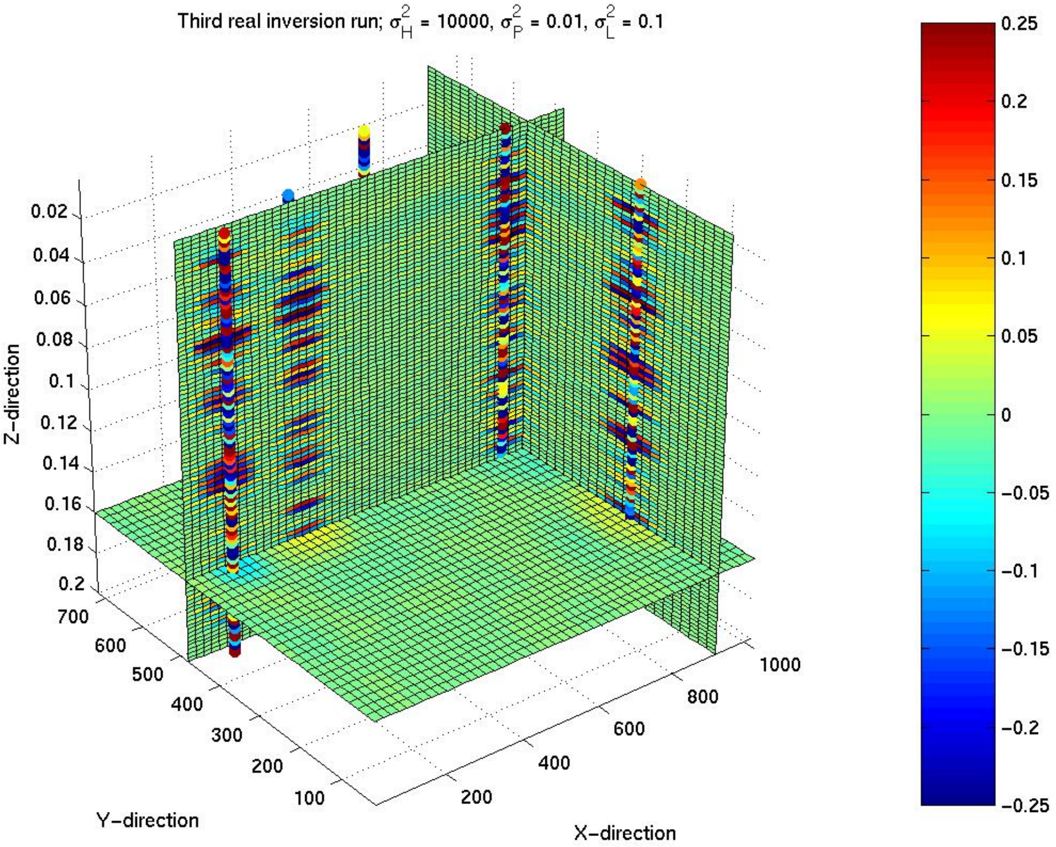

Figure 9: Real Inversion 3