A simple method to incorporate correlation information into the inversion process of Full-wave tomography

Terumitsu Tsuchiya  (Dia Consultants Co., Ltd.)

Abstract

Inversion process in the Full-wave tomography (FWT) is delicate for noise. This is because the process cannot distinguish between signal and noise, and hence it works to fit even noise to model. A partial solution of this problem is to take care of the data uncertainty. While Tarantola's probabilistic theory, which is a basis of the standard waveform inversion, advises a formal methodology for this direction, his formula is not easy to realize in the practical calculation. A simplified expression, however, suggests that incorporation of correlation information of medium can be an approximation of the method.

To seek a method to manipulate a noisy dataset, I did a numerical study based on the simplified expression. Since one of the efficacies of FWT is its high resolving power, I set up a discontinuous structure to see if it maintains the resolution at the discontinuity. Then I observed effects of random noise and tested a procedure for incorporation of correlation information. Consequently I found, 1) the effect of random noise appears as velocity fluctuation, 2) the incorporation of correlation information is effective, and 3) if appropriate correlation lines and reasonable candidates for correlation length are established, improvement of tomograms with maintaining the resolution can be achieved via an AIC-based method.

1 Introduction

Real data is always contaminated by noise. Besides the inversion process of Full-wave tomography (FWT) cannot distinguish between signal and noise, and hence it works to fit even noise to model. Thus the FWT inversion is delicate for noise. However we can find only a few studies in the literature for manipulation of noisy data. Watanabe et al.(1997) is an example from which we can learn something about the noise problem in FWT. In that paper they observed the effect of noise by means of numerical study, and indicated that the pattern of velocity fluctuations due to noise depends little on the amount of noise.

In this paper, we first follow their work also by means of numerical study. Next we seek a method to manipulate a noisy dataset. For this purpose, we will test a procedure for incorporation of correlation information. Since some objective criterion to determine the correlation length is necessary, we will take a way to introduce the Akaike(1974)'s information criterion.
2 Summary of the experiments

The model structure used is shown in Figure 1. Squares and circles indicate source and receiver positions respectively. In the source (left) hole, 25 sources were arranged with 4m intervals. In the receiver (right) hole, 49 receivers were arranged with 2m intervals. Velocities are depicted by the color bar. A fault structure was set up between the source and receiver holes. The ratio of Vp to Vs is a constant of 2, hence the colored structure is common both for Vp and Vs. Density is a constant of 2000kg/m³.

The source wavelet was a 350Hz Ricker type, hence a typical P-wave wavelength was around 7.2m and exceeded one layer thickness, whereas the S-wave wavelength was around 3.6m and less. The ratio of the amplitude of vertical stress component to the horizontal one was defined as 0.44:1.00.

The goal of inversion was to infer Vp, Vs in a range from 2m to 38m from the left borehole. This limitation was due to the difficulty of inversion near the source and receiver positions.

3 FWT on noisy data

The noiseless waveforms were first inverted. The results are shown in Figure 2. The left and right parts are the Vp and Vs tomograms respectively. Inspection of these tomograms shows that both structures were well restored, especially the Vs tomogram is almost perfect.

Next some random noise was added to the simulated waveforms. The frequency band of noise was limited to 220Hz - 450Hz so that it would have approximately the same amplitude spectra as the source wavelet. Then the S/N peak ratio was decreased to around 10. Comparison of the results with those on the noiseless data showed that weak fluctuations of velocity have appeared on both tomograms, however these tomograms resemble for the noiseless case.

Then we see the case with the dataset containing more noise. For this study, random noise was added more to the noiseless dataset so that the S/N peak ratio would decreased to around 2. The same noise generator as in the previous case was used to generate the noise. Hence the generated noise is identical to the previous case, but the amplitude is different. An example of these waveforms is shown in Figure 3. We must keep in mind that such low S/N ratio is not rare in the real field since we often have to apply a narrow band frequency filter to eliminate strong tube waves. Figure 4 shows the inverted results on this lower S/N dataset. We see the remarkable velocity fluctuations and find a important issue; the apparent wavelengths of the fluctuations of both tomograms are roughly equivalent to the corresponding actual wavelengths. A scale around a wavelength is the most sensitive to FWT, hence the result shown in Figure 4 indicates that the random noise related velocity fluctuation tends to appear at such a scale. This is a big problem. In fact, we would have to say at least that fault interpretation would be impossible if we acquired such low S/N records.
4 Incorporation of correlation information

Next we try to improve the quality of the tomograms shown in Figure 4.

In the nonlinear inversion, the model which minimizes the misfit function

$$S(m) = \frac{1}{2} \left[ (f(m) - d_0)^T C_d^{-1} (f(m) - d_0) + (m - m_0)^T C_m^{-1} (m - m_0) \right]$$  \hspace{1cm} (1)

is sought based on Tarantola's probabilistic theory. In this formulation, the data uncertainty is taken care of through the tradeoff between data covariance $C_d$ and model covariance $C_m$. A realization of this argument in the local optimization manner is the following formula

$$m_{n+1} = m_n - \left[ F_n^T C_d^{-1} F_n + C_m^{-1} \right]^{-1} \times \left[ F_n^T C_d^{-1} (f(m_n) - d_0) + C_m^{-1} (m_n - m_0) \right].$$  \hspace{1cm} (2)

by which we can iteratively approach to the desired model (Tarantola, 1987). Hence, if this formula is applied faithfully, the care of data uncertainty is naturally implemented. However it is not easy task to realize eq. (2), and a generalized expression

$$m_{n+1} = m_n - \Omega_n P_n \hat{g}_n$$  \hspace{1cm} (3)

is often adopted in the practical calculation where $P$ denotes positive transform matrix to approximate the inverse Hessian and $\hat{g}$ denotes gradient (Tsuchiya, 1997a). Eq. (3) does not show a practical care of data uncertainty. Thus we wish to seek a substitutive expression of eq. (3) so as to be applicable to the noise problem.

First we must note that the data noise forms a data null space and the data null space components are naturally ignored if the least squares works well, i.e. if the inverse problem is sufficiently over-determined. However if the over-determination is insufficient, the least squares tends to fit its solution also to the noise existing in the data null space. To restrain this action we have to reduce the size of model space.

Suppose the following condition

$$C_d = \sigma_d^2 I_d, \quad C_m = \sigma_m^2 R_m$$  \hspace{1cm} (4)

where $\sigma_d$ and $\sigma_m$ are standard deviations of data and model, $I_d$ is a unit matrix, and $R_m$ is a correlation matrix. If we define a correlation line which connects correlated medium points successively and correlation coefficients along the correlation line by a distance $\Delta$ and correlation length $l$ as for instance $r = \exp(-\Delta/l)$. $R_m$ is developed reasonably (Tsuchiya, 1997b).

Now we remember that to apply a non-diagonal $R_m$ in the assessment of eq. (3) let the degree of freedom decrease (Charara et al., 1996), where the number of free parameters (let us write it as NFP hereafter) is determined by the correlation length. That is to say, the incorporation of correlation information via $R_m$ may lead to a solution of the noise problem.

Since we are discussing about the problem of data uncertainty, we must assume non small value for $\sigma_d$. However to assume that we have no a priori information of actual values of model is rather natural. Thus we can assume
\[ \sigma_d^2 \ll \sigma_n^2. \]  

Substituting eqs. (4) and (5) into eq. (3), we obtain

\[
m_{n+1} = m_n - \Omega_n P_n \left( \sigma_d^2 F_n'(f(m_n) - d_0) + \sigma_n^2 R_n^{-1}(m_n - m_0) \right)
\]

\[
= m_n - \Omega_n P_n \left( \frac{\sigma_d^2}{\sigma_n^2} R_n F_n'(f(m_n) - d_0) + (m_n - m_0) \right)
\]

\[
= m_n - \Omega_n P_n R_n F_n'(f(m_n) - d_0). \tag{6}
\]

Inpection of eq.(6) shows that the incorporation of correlation information is more or less equivalent to apply local averaging.

To verify the efficacy of this simple method, we then return to the numerical study. The procedure we will adopt is as follows.

i. Establish practically realizable correlation lines.

ii. Establish reasonable candidates of the adequate correlation matrix so that the property of each candidate varies monotonously between two extremes of those.

iii. Evaluate the NFP(number of free parameters) for each candidate.

iv. Find the existence of the minimum AIC within the candidates, and study the quality of corresponding result.

4.1 Correlation lines

To establish the correlation lines, we assume that we know the existence of a steep fault between two boreholes and have measured layers' dips on both sides of the fault.

It is natural to establish correlation lines as shown in Figure 5. In the figure, sparsely selected lines are plotted on the model. In this example, the location of the fault is assumed to be in the center of both boreholes since the exact location is normally unknown.

4.2 Correlation matrix

We consider that the correlation between model parameters exists only along the correlation line. In other words, parameters on different correlation lines have no correlation each other. Thus we regard a parameter set on a correlation line as one set and deal with each parameter set independently. If we ignore the small error due to the slight dip on the right half of the display, the total NFP can be expressed as

\[ NFP_{total} = NTP \times N_z \times 2, \tag{7} \]

where NTP denotes the number of total parameters along a correlation line, \( N_z \) denotes the number of grids along the vertical axis, and the last factor 2 reflects two kinds of unknown; Vp and Vs. As the total NFP is in proportion to the NFP along a correlation line, we will mainly consider the latter and call it NFP hereafter.
As the inversion range is from 2m to 38m in distance from the left borehole and the grid interval is 0.5m, we have 73 grids along a correlation line, namely NTP=73. So our current problem is to create a 73x73 correlation matrix. We assume the center point, namely 37th grid along a correlation line, has no correlation with any other point, and any point at one side with respect to the center point is assumed to have no correlation with any point on any other side.

Bearing in mind such conditions, we consider the following correlation matrix for the points along a correlation line:

\[
\begin{pmatrix}
1 & a & a^2 & a^3 & \ldots & a^N & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
0 & 1 & a^{N/2} & a^{N/2} & \ldots & a^N & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & a & a^2 & \ldots & a^{N/2} & \ldots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & a & a^2 & \ldots & a^{N/2} & \ldots & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & a & a^2 & \ldots & a^{N/2} & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & a & a^2 & \ldots & a^{N/2} & \ldots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & a & a^2 & \ldots & a^{N/2} & \ldots & 0 \\
\end{pmatrix}
\]

where
\[
a = \exp\left(-\frac{3}{N^p}\right),
\]
andN satisfies
\[
2(N + 1) + 1 = NTP.
\]

thus N=35. The variable \( p \) is a parameter. We will term it a correlation parameter and write it as CP hereafter. Next we assume that \( r^3 \) is very small. If we wish to take the weakest correlation as one extreme, we let the largest coefficient \( a \) except 1 in the matrix be given \( e^{-3} \). This is realized by CP=0.0. If we wish to take the strongest correlation on condition that some smoothness at the center point will be preserved, we let the smallest coefficient \( a^N \) be given \( e^{-3} \). We assume that the change from \( e^{-3} \) to 0 is smooth, and this corresponds CP=1.0. The strength of correlation varies monotonously between these two extremes. Then we will manipulate the parameter CP in a range from 0.0 to 1.0.

### 4.3 Number of free parameters and AIC

To evaluate NFP we decompose the correlation matrix \( \hat{R}_m \) as

\[
\hat{R}_m = \hat{M} \hat{M}^T.
\]

Charara et al. (1996) proved that the NFP can be evaluated as

\[
NFP = \text{trace}(\hat{M}).
\]
We will follow their work by applying the LU decomposition method to $\tilde{R}_m$.

To determine an adequate NFP which relates to correlation length, we introduce the Akaike Information Criterion (AIC):

$$AIC = -2l(\hat{\theta}_m) + 2 \times NFP.$$  

(13)

where $l(\hat{\theta}_m)$ denotes the maximized log-likelihood of the model.

4.4 Existence of the AIC minimum, and the result

The experiment to verify the existence of the minimum AIC was implemented on the noisy dataset of S/N=2. In this experiment, three values 0.8, 0.9 and 1.0 were selected for the CP in addition to 0.0 which is the case of Figure 4.

Change of AIC with the procession of inversion is shown in Figure 6. The results show that there exists a minimum AIC between CP=0.8 and CP=1.0, and CP=0.9 gives the minimum value among the tested cases. Then we see the result of the case CP=0.9 in Figure 7. Inspection of the figure shows that the effect of the procedure is remarkable, especially on the Vs tomogram. Moreover, in spite of the cutting of the correlation at the center point, the continuity of the layers at that point is almost recovered.

Thus we can conclude as follows. The incorporation of correlation information is effective to manipulate a noisy dataset even with a faulted structure. If we establish appropriate correlation lines and reasonable candidates for correlation length, we can find the minimum AIC among the candidates. The result corresponding to the AIC minimum is expected to give a better result.

5 Conclusion

In this article, we studied a simple method to incorporate correlation information into FWT. As a result, we found;

i. The effect of random noise appears as velocity fluctuations, and the wavelength of the velocity fluctuation is similar to the actual seismic wavelength.

ii. As the S/N ratio decreases, velocity fluctuations increase.

iii. The incorporation of correlation information is effective in manipulating a noisy dataset.

iv. If we establish appropriate correlation lines and reasonable candidates for correlation length, we can find the minimum AIC among the candidates.
References


**Fig. 1** Model structure and acquisition geometry

**Fig. 2** Inverted tomograms on noise free data

**Fig. 3** Simulated input waveforms with S/N = 2. Source depth is 50m.

**Fig. 4** Inverted tomograms on noisy data (S/N=2) without taking care of data uncertainty
Fig. 5 Correlation lines

Fig. 6 Change of AIC

Fig. 7 Inverted tomograms on noisy data (S/N=2) by incorporating correlation information with CP=0.9