Tomography Imaging
Through the Monte Carlo Sampling

Wojciech DĘBSKI
Institute of Geophysics, Polish Academy of Sciences
ul. Księcia Janusza 64, 01-452 Warszawa, Poland
debski@igf.edu.pl

Abstract

This paper illustrates advantages of the Bayesian approach to seismic tomography imaging enhanced by the Monte Carlo sampling technique. The theoretical considerations are illustrated by an analysis of seismic data from Rudna (Poland) copper mine. Contrary to the classical LSQR techniques, the Bayesian approach allows to construct not only the “best fitting” model of the sought velocity distribution but also other estimators like, for example, the average model which is often expected to be a more robust estimator than the maximum likelihood solution. We demonstrate that using the Markov Chain Monte Carlo sampling technique within the Bayesian approach opens a possibility of an analysis of tomography imaging uncertainties with a minimal additional computational effort compared to the robust optimization approach.

1. Introduction

Seismic velocity tomography is the inversion technique aimed at imaging spatial distribution of velocity heterogeneities in global, regional, local or even laboratory scale problems. It relies on the high frequency approximation according to which seismic energy may be assumed to propagate along a thin “tube” between a source and a receiver, called the ray path. Thus, the energy travel time between two points provides information on the average seismic velocity along the ray path. If the travel time data are available for a number of ray paths probing different parts of the studied area, it becomes possible to obtain a spatial map of the local heterogeneities of the velocity distribution – the velocity image called also a tomogram.

The basic tomography forward problem formula relates the observable travel time data $d^\text{obs}$ with the slowness distribution $s(r)$ and reads (Aki and Richards 1985)
\[ d^{th}(s) = \int_{\text{ray}[s]} s(r) \, dl \]  \quad (1)

where \( s(r) \) denotes slowness (inverse of velocity), \( d^{th} \) is the predicted travel time, and the integral is taken along a seismic ray path. The ray path depends on the slowness distribution; consequently, the above relation between \( d^{th} \) and \( s \) is highly nonlinear what manifests itself in a bending of the ray path. However, this effect may often be neglected in case of local seismic tomography if the slowness distribution does not have large gradients (Cardarelli and Cerro 2002). Such a simplification makes the tomography problem easier to solve at the cost of a slight lost of sharpness of tomograms (Maxwell and Young 1993). We follow this assumption through the analysis presented here.

Having a set of travel times data \( d = (t_1, t_2, \cdots, t_N) \) recorded for different source-receiver pairs and having the discretized slowness distribution (for example, assuming that the images may be described as a set of homogeneous cells) the linearized tomography forward problem turns to a set of linear equations (Nolet 1987)

\[ d^{th}_i = \sum_j G^{ij} s_j \quad i = 1 \cdots N \]  \quad (2)

where \( s = (s_1, s_2, \cdots, s_M) \) is the vector of discretized slowness field and \( G \) is a matrix whose elements \( G^{ij} \) are equal fractions of \( i \)-th ray path in a \( j \)-th cell.

An estimation of the slowness parameters \( s \) can be carried out in various ways, among which a version of the direct algebraic solution and the optimization approach are the most often used (Iyer and Hirahara 1993). Another, Bayesian approach based on the sampling of the \textit{a posteriori} probability density is the method which, following the pioneer work of Tarantola and Vallete (1982) (see also Sambridge and Mosegaard (2002) for a review of the history of application of the Monte Carlo technique in geophysics) is recently gaining greater and greater popularity (Curtis and Lomax 2001, Bosch et al. 2000, Mosegaard and Tarantola 2002, Dębski 2004). Both approaches are shortly described below.

2. Tomography Imaging – Theory

2.1 Classical approach

If the forward modeling formula given by Eq. 1 can be linearized, the relation between the data \( d \) and model parameters \( s \) (slowness) takes the form of a set of linear equations
and the problem of estimation of \( s \) from \( d^{\text{obs}} \) can be viewed as the task of solving the set of linear equations after the substitution \( d = d^{\text{obs}} \). This task can be accomplished within a simple algebraic approach (Menke 1989, Tarantola 1987, Parker 1994) by matrix manipulation, as follows.

In the first step, the so-called normal equation is formed by multiplying both sides of Eq. 3 by the transposed \( G^T \) operator:

\[
G^T d^{\text{obs}} = (G^T \cdot G) s.
\]  

The matrix \( G^T G \) is a square matrix which usually cannot be inverted yet because of possible singularity. To fix this problem, the \( G^T G \) matrix is regularized which, in the simplest case, is achieved by adding a small diagonal term:

\[
G^T G \rightarrow G^T G + \gamma I
\]  

where \( I \) is a diagonal matrix. Finally, the analytical formula for the model \( s^{\text{lsqr}} \) regarded as the solution can be cast into the form with the explicitly introduced initial (\textit{a priori}) model \( s^{\text{apr}} \) (Menke 1989, Tarantola 1987)

\[
s^{\text{lsqr}} = s^{\text{apr}} + (G^T \cdot G + \gamma I)^{-1} G^T (d^{\text{obs}} - G s^{\text{apr}}) + O(\gamma^2).
\]  

This form is particularly convenient when \( s^{\text{lsqr}} \) is estimated iteratively (Tarantola 1987) and is de facto the starting point of various linear inversion techniques (Parker 1994, Limes and Treitel 1983).

The above formula for \( s^{\text{lsqr}} \) in fact coincides with the solution obtained by the optimization approach, when the least squares difference between observed and modeled travel times is minimized, subject to an additional “smoothness” condition. For this reason, this method is often called the \textit{LSQR} (Damped Least Squares) solution (Parker 1994, Menke 1989).

### 2.2 Bayesian point of view

The method shortly described above is very popular and most often used in everyday practice (Zhao 2001, Iyer and Hirahara 1993, Nolet 1987). The drawback of this approach is the lack of a reliable estimation of the quality of the image found, because the solution is just a single (optimum) model. The method provides no information on how large is a region in the model space around the best solution such the models (slowness distribution) taken from it lead to similar predictions as the optimum model.
The above drawbacks of the classical method are conveniently overcome when the Bayesian inverse theory is applied to the tomography problem in hand (Bosch 1999, Dębski 2004).

The solution of the tomography inverse problem according to the Bayesian inverse theory consists in building the \textit{a posteriori} probability distribution $\sigma(s)$ over the model space which describes the probability of a given model (slowness distribution $s$) being the true one (Tarantola 1987, Dębski 2004). It has been shown by Tarantola and Vallete (1982), Tarantola (1987), Jackson and Matsu’ura (1985) that $\sigma(s)$ is the product of the distribution $f(s)$ describing \textit{a priori} information by the \textit{likelihood function} $L(s)$ which measures to what extent theoretical predictions fit the observed data:

$$
\sigma(s) = \text{const}. f(s) L(s)
$$

where the constant represents normalization of the probability density and the \textit{likelihood function} $L(s)$ reads (see, for instance, Tarantola 1987, Duijndam 1988, Jackson and Matsu’ura 1985)

$$
L(s) = \exp \left\{ -\|d_{\text{obs}} - d_{\text{th}}(s)\| \right\}
$$

and the symbol $\| \|$ stands for a norm used to measure the “distance” between two vectors.

Knowledge of the $\sigma(s)$ distribution allows not only to find the most likelihood model $s_{ml}$ for which $\sigma(s_{ml}) = \max$ but also other characteristics like, for example, the average model

$$
s_{\text{avr}} = \int_M s \sigma(s) ds,
$$

and the covariance matrix

$$
C_{ij}^p = \int_M (s_i - s_{\text{avr}}^i)(s_j - s_{\text{avr}}^j) \sigma(s) ds.
$$

These two basic characteristics of the \textit{a posteriori} distribution $\sigma(s)$ are indeed very important. The importance of the average model $s_{\text{avr}}$ comes from the fact that it provides not only information on the best fitting model but also includes information about other plausible models from the neighborhood of the “best” model $s_{ml}$. If sub-optimum models defined as those for which $\sigma(s) \sim \sigma(s_{ml})$ are similar to $s_{ml}$, then $s_{\text{avr}} \sim s_{ml}$. The diagonal elements of the \textit{a posteriori} covariance matrix $C^p$ are convenient estimators of the inversion uncertainties for each component of $s$ while the non-diagonal elements measure the degree of correlation between pairs of parameters.
(Menke 1989, Jeffreys 1983). In fact, \( C^p \) given by Eq. 10 is a generalization of the \textit{LSQR} covariance matrix to the case of an arbitrary statistics \( \sigma(s) \) including possibly nonlinear forward problems. As in the case of the average model, the posterior covariance matrix is meaningful only if the \( \sigma(s) \) distribution is unimodal. In cases of multi-modality, the existence of non-resolved directions in the model space, or other “pathologies”, a more exhaustive error analysis is necessary by a full inspection of the \textit{a posteriori} distribution (Tarantola 1987, Wiejacz and Dębski 2001). This can be achieved, for example, by calculating the higher-order moments of the \textit{a posteriori} distribution (Jeffreys 1983).

\subsection{Sampling a posteriori PDF}

To calculate the point estimators like the average model and the covariance matrix, we face the problem of sampling of \( \sigma(s) \), usually in the context of calculation of multidimensional integrals as in Eqs. 9, 10.

More generally, we may need to calculate the average integrals represented in a general form as

\[
< F > = \int_{\mathcal{M}} F(s)\sigma(s)ds
\]

where \( F(s) \) is an arbitrary function of \( s \).

If the number of parameters is very small (smaller than, say, 10), the integral in Eq. 11 can be calculated by sampling of \( \sigma(s) \) over a predefined regular grid like, for example, in the case of the seismic source location problem (Wiejacz and Dębski 2001, Lomax \textit{et al}. 2000, Sambridge and Kennett 2001). Otherwise, the stochastic Monte Carlo technique (\textit{MC}) is to be used (Mosegaard and Tarantola 1995, Bosch \textit{et al}. 2000, Dębski 2004).

Generally speaking, the \textit{MC} technique allows to generate an ensemble of models which can be regarded as a set of samples drawn from the \( \sigma(s) \) distribution. Then, the integral \( < F > \) can be approximated by (Robert and Casella 1999)

\[
< F > \approx \frac{1}{N} \sum_{\alpha} F(s^\alpha)
\]

where the sum is taken over an ensemble of all \( N \) generated models \( s^\alpha \).

One of the greatest advantages of the possibility of calculating the statistical averages from the ensemble of the \textit{MC} generated models is that inversion errors can be easily estimated with no additional cost. If they can be approximated by variance of the \textit{a posteriori PDF} then the choice

\[
F = (s - s^{av})^2
\]
leads to the extremely simple, easy to calculate estimator of the a posteriori errors

\[ \epsilon_a^2 = \frac{1}{N} \sum_{s^\alpha} (s^\alpha - s^{avr})^2. \]  

(14)

The choice of a sampling scheme of \( \sigma(s) \) (generating samples drawn from the \( \sigma(s) \)) is by no means a trivial task especially when the number of estimated parameters is large (Mosegaard and Tarantola 2002, Bosch et al. 2000) like in the case of seismic tomography. This is caused by the fact that with an increasing number of parameters, the model space becomes extremely large but only its very small part contributes to the a posteriori distribution (Curtis and Lomax 2001). In cases like seismic tomography, only the Monte Carlo technique can be applied to solve practical problems. Among various MC algorithms, the most flexible and the most efficient for tomography problems seems to be the class of the Markov Chain Monte Carlo algorithms (MCMC).

These algorithms originally simulate Markovian stochastic processes (Gillespie 1992, Robert and Casella 1999) which “samples” a given stationary PDF (Tierney 1994, Mosegaard and Tarantola 2002, Robert and Casella 1999). The simplest form of such an algorithm, the Metropolis algorithm, is shown in Fig. 1.

- Initialize \( s^0 \)
- Repeat
  - generate uniform random number \( u \sim U(0, 1) \)
  - generate new test sample \( s^\beta = s^\alpha + \delta s \)
  - where \( \delta s \) is a small perturbation to the current sample
  - if \( u < P(s^\beta, s^\alpha) = \min \left[ 1, \frac{\sigma(s^\beta)}{\sigma(s^\alpha)} \right] \)
    \( s^{\alpha+1} = s^\beta \)
  - otherwise \( s^{\alpha+1} = s^\alpha \)
- Continue until sufficient number of samples \( \{s^\alpha\} \) is generated

Fig. 1: The Metropolis algorithm for sampling a posteriori PDF \( \sigma(s) \).
3. Rudna Copper Mine Case

3.1 Data, parametrization and inversion setup

The Bayesian algorithm was used to image the velocity distribution in a part of the Rudna copper mine in south-western Poland. The data set was collected during July, August and September 2004. The Rudna mine runs a digital seismic network composed of 32 vertical sensors located underground at depths from 550 m to 1150 m. Willmore seismometers MK-II and MK-III are used. The seismic signals are transmitted to a central site at the surface where they are digitized by a 14-bit converter. The frequency band of digitized data is from 0.5 to 150 Hz, and the final dynamics is about 70 dB. The sampling frequency which determines the absolute accuracy of travel time onset reading is $F_s = 500$ Hz ($C_d = 2$ ms). The uncertainty of event location is better than 100 m, typically around 50 m.

Fig. 2: Ray paths coverage of the studied region. Seismometer positions are marked by circles and the locations of seismic sources used in the current study are depicted by stars.
For the current study I selected a set of 36 events which occurred in the Rudna mine and were recorded by at least 4 out of the 8 nearby stations located almost at the same depth as the estimated hypocenters. This gave a total of 177 travel times used for P-wave velocity inversion. The distribution of selected events, employed stations and considered ray paths are shown in Fig. 2. The events were located assuming a constant background velocity, $V_a = 5800$ m/s. This value was also taken as the initial $a$ priori velocity model for the tomography inversion.

The studied area of around $8 \times 8$ km was parameterized by division into 400 m by 400 m square cells, which meant 400 parameters to be estimated. However, most of these cells were probed by no ray, as shown in Fig. 2. In fact, only 107 cells were touched by at least one ray path. Only such cells were effectively considered during the inversion, and the $a$ priori velocity $V_a = 5800$ m/s was assigned to the remaining, untouched cells.

The Gaussian distributions with constant diagonal covariance matrices $C_{appr} = C_s I$ and $C^p = C_d I$ were used to model the $a$ priori $f(s)$ and the likelihood $L(s)$ functions. The corresponding values of $C_s$ and $C_d$ were estimated in the following way. Firstly, I assumed ad hoc but on the basis of mining practice that velocity changes cannot exceed 10% of background velocity with a 95% confidence level. This $a$ priori assumption corresponds approximately to the choice $C_s^{-1} = 200$ m/s. The value of $C_d$ is limited by the sampling period (2 ms) on the one hand and by the location errors on the other. Location errors of about 100 m lead to travel time inaccuracies of about 17 ms. I choose the intermediate value $C_d = 10$ ms.

3.2 Results

The main results of the tomography imaging are shown in Figs. 3 and 4. Figure 3 shows the average velocity image obtained on the base of 500,000 models generated by the Metropolis algorithm sampling the $a$ posteriori PDF. Spatial distribution of the imaging errors estimated by the diagonal elements of $a$ posteriori covariance matrix according to Eq. 14 is shown in Fig. 4. In addition, Fig. 5 shows the travel time residua calculated for the $a$ posteriori average model.

The very first step in evaluating tomography results relies on inspecting the residua for the model found. Figure 5 shows that $a$ posteriori residua are well bounded within the $a$ priori assumed value of data errors of 10 ms. In fact, the root mean square value of the residua is about 5.5 ms. Only for a few ray paths the differences between observed and modeled travel times are larger than $2C_d$ amounting to 20 ms. It means that the $a$ posteriori velocity model explains the travel time data quite well.
Fig. 3: Average velocity image obtained by the Monte Carlo sampling technique. The seismometers and epicenters of used seismic events are depicted by squares and white stars, respectively.

A more comprehensive insight into the imaging accuracy is obtained by an inspection of the \textit{a posteriori} covariance matrix which is shown in Fig. 4.

Note that the \textit{a priori} value $C_s = 200 \text{ m/s}$ is attached to cells probed by no rays. The obtained distribution shows that within the applied velocity parameterization (cell size) the imaging errors were diminished with respect to $C_s$ in the area well covered by ray paths. In the central part of this area, the \textit{a posteriori} errors get values lower than 150 m/s. Let us note, however, that outside of it the imaging errors were generally larger than $C_s$ which means that the velocity distribution in this part is poorly resolved by tomography. Finally, let us note the irregular shape of the area where the \textit{a posteriori} errors are smaller than $C_s$. 
Fig. 4: Spatial distribution of imaging errors. The seismometers and epicenters of used seismic events are depicted by squares and stars, respectively. Notice a decrease of the a posteriori errors with respect to the a priori value $C_s = 200$ m/s in the central, well resolved part of the imaging area, and their increase outside of it.

4. Discussion

Two types of conclusions can be drawn on the basis of the studied case. The first one concerns the performance of the Bayesian approach enhanced by the MCMC sampling technique. Secondly, a very preliminary and quantitative physical interpretation of the obtained image can be attempted.

As follows from the presented case study, the MCMC sampling technique seems to be able to provide a quite robust estimation of the velocity distribution by an efficient generation of an ensemble of velocity models which follows the a posteriori PDF.

Following these comments, one can state that the Monte Carlo sampling for the solution of tomography imaging is a very promising technique which is able to provide
robust and more reliable images than any other currently used techniques. Its application is only limited by the size of the problem in hand. In the studied case, the simple Metropolis sampling algorithm was efficient enough to sample the \textit{a posteriori} PDF due to the very small scale of the problem. In the case of larger-scale problems, when significantly more parameters have to be estimated (sampled), more complicated sampling techniques like, for example, the multi-step Metropolis algorithm, have to be employed (Robert and Casella 1999, Mosegaard and Tarantola 2002).

Finally, having obtained velocity distribution images, it is natural to attempt to try to infer information on the correlation of the velocity field with the observed seismicity. Although the current investigation was carried out to check numerical algorithms only, the obtained results seem to be robust enough to make some preliminary comments.

Firstly, note that the events used in the current studies formed a few spatial clusters, three of which are located in the well resolved region as shown in Fig. 4. For those
clusters, the tomography method has a sufficient resolution to map the velocity heterogeneities with the sufficient precision. The obtained tomogram suggests that two clusters located around \(x = 30500 - 31000, y = 6500 - 7000\) occurred in the area where the velocity was smaller than background but high spatial velocity gradients are mapped. On the other hand, the cluster located at around \(x = 31800, y = 6200\) appeared in the area where the velocity was higher than background but the mapped spatial velocity gradient is smaller than in the case of the previous clusters. No much can be said about other events, as they are located in regions poorly resolved by tomography.

Can the occurrence of the seismic clusters in the areas of the different velocity and velocity gradient characteristics suggest different mechanisms of induced seismicity? Unfortunately, it is impossible to answer this question within the current study. The limited spatial resolution due to the small number of seismic events, the non-homogeneous ray path coverage, and the finite accuracy of forward modeling prevents any more detailed attempt at a physical interpretation of the obtained results and prompts for the further, more detailed studies.

### References


Chapman and Hall.


in 3D and layered models: Introduction of a Metropolis-Gibbs method and com-
parison with linear location.* In: C. Thurber, E. Kissling and N. Rabinovitz (eds.),
"Advances in Seismic Event Location", Amsterdam: Kluver.

Maxwell, S.C., and R.P. Young, 1993, *A comparison between controlled source and

Menke, W., 1989, *Geophysical Data Analysis: Discrete Inverse Theory,* International

Mosegaard, K. and A. Tarantola (1995). Monte Carlo sampling of solutions to inverse

Mosegaard, K., and A. Tarantola, 2002, *International Handbook of Earthquake En-
gineering Seismology,* Chapter Probabilistic Approach to Inverse Problems, pp. 237–
265, Academic Press.


Press.


*Accepted 13 June 2006*