Comparison of deterministic and stochastic approaches to crosshole seismic travel-time inversions

YanZhe Zhao, and YanBin Wang*
Department of Geophysics, School of Earth and Space Sciences, Peking University, Beijing 100871, China

Abstract: The Bayesian inversion method is a stochastic approach based on the Bayesian theory. With the development of sampling algorithms and computer technologies, the Bayesian inversion method has been widely used in geophysical inversion problems. In this study, we conduct inversion experiments using crosshole seismic travel-time data to examine the characteristics and performance of the stochastic Bayesian inversion based on the Markov chain Monte Carlo sampling scheme and the traditional deterministic inversion with Tikhonov regularization. Velocity structures with two different spatial variations are considered, one with a chessboard pattern and the other with an interface mimicking the Mohorovičić discontinuity (Moho). Inversions are carried out with different scenarios of model discretization and source–receiver configurations. Results show that the Bayesian method yields more robust single-model estimations than the deterministic method, with smaller model errors. In addition, the Bayesian method provides the posterior probabilistic distribution function of the model space, which can help us evaluate the quality of the inversion result.

Keywords: stochastic approach; deterministic approach; Bayesian inversion; Markov Chain Monte Carlo; Tikhonov regularization

Citation: Zhao, Y. Z. and Wang, Y. B. (2019). Comparison of deterministic and stochastic approaches to crosshole seismic travel-time inversions. Earth Planet. Phys., 3(6), 547–559. http://doi.org/10.26464/epp2019056

1. Introduction

Because of limitations in observations and the existence of noise, solutions to geophysical inverse problems are often nonunique. Numerous inversion methods have been developed to construct and evaluate geophysical models based on observations, which fall into two different categories: deterministic and stochastic approaches.

From the deterministic point of view, observations and model parameters are considered deterministic, and the inversion process is aimed at finding the model satisfying the observations according to certain criteria. The essence of the deterministic approach is to seek the optimal single-model solution. In general, linear problems lead to a class of linear equations,

\[ Gm = d, \]

whereas nonlinear problems are often cast into iteratively linear problems to obtain approximate solutions. In Equation (1), \( m \) should be determined with \( G \) and \( d \) representing the physical law and observation data, respectively. A solution to the inverse problem in Equation (1) often does not exist because practical problems are always ill-posed. Therefore, Equation (1) is usually solved with regularization, such as damping and smoothing. In this study, we choose the widely used Tikhonov regularization method (Tikhonov and Arsenin, 1977) as a representative approach to regularizing geophysical inverse problems.

In contrast to the traditional deterministic methods, the stochastic inversion, which is often called Bayesian theory, considers both the data and the model parameters as random variables. Therefore, the essence of stochastic inversion is to find the probability distributions of model parameters instead of their fixed values. Bayesian probability theory was first put forward by Bayes in the 1700s, but it was not adopted in geophysical inversions until recently because of the complexity of geophysical problems and the limitations in computational power. Early studies on the Bayesian methods were limited to theoretical studies or relatively simple inversion experiments. Matsu'ura (1984) proposed a new earthquake location method from the Bayesian viewpoint and showed through numerical experiments that the Bayesian approach to earthquake location yielded more stable results. Jackson and Matsu’ura (1985) developed the general theory of the Bayesian solution to linear and nonlinear problems. They suggested that the Bayesian method would be more advantageous than other methods in quantifying prior information and could provide more stable and meaningful error estimates. With the development of sampling algorithms, such as the Monte Carlo (MC) method, it became possible to perform an extensive exploration of the model space (Tarantola, 2005) so that Bayesian inversion could be applied in many geophysical fields. Rothman (1985) first used the MC method to sample the posterior distribution of seismic reflection parameters, whereas Tarits et al. (1994) introduced the MC method into magnetotellurics, with applications to one-dimensional inversions of synthetic and real data. Mosegaard and Tarantola (1995) inverted gravity data by using the MC method. Numerous similar studies have been conducted in a variety of dis-
ciples, and a well-developed sampling algorithm, the Markov chain Monte Carlo (MCMC) method, has gradually become the mainstream approach. In addition to the traditional MCMC method, which samples the model space of fixed dimensions, the reversible-jump MCMC method (rj-MCMC; Green, 1995, 2003) was developed, which can be used adaptively to search the model space of variable dimensions. At the same time, the Bayesian information criterion (Schwarz, 1978) was proposed to infer the optimal number of model parameters, and its combination with the MCMC method renders the Bayesian inversion efficient and practical. Dettmer and Dosso (2012) combined this method with an autoregressive error model to achieve the transdimensional matched-field acoustic inversion of data obtained from a vertical array in the Mediterranean Sea. Pachhai et al. (2014) applied it in a seismic waveform inversion to infer the ultralow velocity zones in the lowermost mantle under the eastern Philippines.

In this study, we compare these two kinds of inverse theory through a crosshole seismic travel-time tomography experiment, an inversion technique aimed at imaging the spatial distribution of velocity heterogeneities. In Section 2, we introduce the theories and algorithms of both the Bayesian and Tikhonov regularization methods. In Section 3, we describe the numerical models and synthetic data used for the inversion experiments. We analyze and compare the inversion results from the two methods in Section 4. Through the analysis, we find that the Bayesian method may be more robust and reliable than deterministic methods, and the posterior probability distribution function can help us better evaluate the quality of the inversion results. On the other hand, deterministic methods are relatively easy to implement and require much less computational effort than stochastic methods. From these characteristics, some insights into the applications of the two inversion approaches can be obtained.

2. Inverse Algorithms

2.1 Tikhonov Regularization Method

The least-squares solution to Equation (1) can be obtained by minimizing misfit between the observed and predicted data, i.e., by solving the following optimization problem:

$$\min \|Gm - d\|_2.$$  \hspace{1cm} (2)

For ill-posed problems, the Tikhonov regularization is applied, and the objective function takes the form

$$\min \left\{ \|Gm - d\|_2 + \lambda \|Lm\|_2 \right\},$$  \hspace{1cm} (3)

where $\lambda$ is a positive constant chosen to control the property of the solution vector and $L$ is a matrix that defines a (semi)norm of the solution by which the property is quantified. Often, $L$ can be the identity matrix or a differential operator. Here, we choose $L$ as the identity matrix, for which the Tikhonov problem is said to be in its standard form. Therefore, the problem in Equation (2) is transformed into

$$\min \left\{ \|Gm - d\|_2 + \lambda \|m\|_2 \right\}.$$  \hspace{1cm} (4)

The parameter $\lambda$ can be determined with several approaches, and here we use the popular $L$-curve criterion (Hansen and O’Leary, 1993). For linear problems, the curve of $\|m\|_2$ versus $\|Gm - d\|_2$ often takes on a characteristic $L$ shape when plotted on a log-log scale, and the optimal value for $\lambda$ is chosen as the point at which the curve has the greatest curvature.

2.2 Bayesian Method

In the Bayes theory, the model parameters are constrained by both the prior information and the observation. Therefore, the solution will be given as a probability distribution, called the posterior probability distribution function (PPDF), whose purpose is to provide an estimation of the model parameters as well as their associated uncertainties.

For a data vector $d$ composed of $N$ measurements and a model vector $m$ involving $M$ model parameters, the Bayes theory can be expressed as (Aster et al., 2019)

$$P(m|d) = \frac{P(d|m)P(m)}{c},$$  \hspace{1cm} (5)

where $c = \int_{\text{all models}} P(d|m)P(m) \, dm$ is a normalization constant. $P(d|m)$ and $P(m)$ are the posterior and prior distributions, respectively. Given the data vector $d$, $P(d|m)$ can be interpreted as the likelihood function $L(m)$:

$$L(m) = P(d|m) = P(d_1|m)P(d_2|m)\cdots P(d_N|m).$$  \hspace{1cm} (6)

The likelihood function measures how well a model with a given set of model parameters $m$ fits the observed data $d$.

In the case of multivariate normal distribution, the correlations among the model parameters and among the observed data can be measured by their covariance matrices, $C_m$ and $C_d$ respectively. Hence, the prior distribution can be written as

$$P(m) \propto e^{-\frac{1}{2}(m-m_{\text{prior}})^T C_m^{-1}(m-m_{\text{prior}})},$$  \hspace{1cm} (7)

where $m_{\text{prior}}$ is the mean of the prior distribution. Thus, the posterior distribution can be expressed as

$$P(m|d) \propto e^{-\frac{1}{2}[(m-m_{\text{prior}})^T C_m^{-1}(m-m_{\text{prior}}) + (Gm-d)^T C_d^{-1}(Gm-d)]} = e^{-\frac{1}{2}(m-m_{\text{max}})^T C_m^{-1}(m-m_{\text{max}})},$$  \hspace{1cm} (8)

where $C_m = (G^T C_d^{-1} G + C_m^{-1})^{-1}$, and $m_{\text{max}}$ is the solution that maximizes the posterior distribution.

2.3 Markov Chain Monte Carlo Method

The MCMC method is a stochastic simulation algorithm. For a complex probability distribution, the MCMC method draws a large number of samples for the distribution. It is based on the Markov chain theorem, which guides the process of generating a series of model samples $x_1, x_2, \ldots, x_n, x_{n+1}$ consecutively. The probability of generating a sample $x_{n+1}$ depends only on the previous sample $x_n$:

$$P(x_{n+1}|x_1, x_2, \ldots, x_n) = P(x_{n+1}|x_n),$$  \hspace{1cm} (9)

where $P(x_{n+1}|x_n)$ is the proposal probability distribution and can be written as $P(x_n, x_{n+1})$. For most chains, the distribution will converge to a limiting probability density distribution function $\pi(x)$ at the same time as $n$ tends to infinity. In other words, $x_i$ $(i \geq n)$ will be the samples of distribution $\pi(x)$ after a sufficiently large number $n$. In contrast, a function $\pi(x)$ satisfying the detailed balance condi-
will definitely be the limiting distribution of a Markov chain with the proposal distribution \( P(x) \). Considering the detailed balance condition, we can construct a Markov chain with a given proposal distribution to generate samples to obtain the PPDF. The Metropolis–Hastings algorithm is a classic form (Metropolis et al., 1953; Hastings, 1970). It involves the following steps (Mosegaard and Tarantola, 1995; Aster et al., 2019):

1. Generate a starting model, \( x_0 \).
2. Repeat and continue until a sufficient number of samples \( \{x_n\} \) is generated:
   a. Generate a candidate model \( c \) from the previous model \( x \) by using the proposal distribution \( P(x) \);
   b. Compute \( q(x, c) = \min \left\{ \frac{P(c)}{P(x)} \right\} \);
   c. Generate a random number \( u \sim U(0, 1) \);
   d. If \( u < q \), let \( x_{n+1} = c \), otherwise \( x_{n+1} = x_n \).

Two key points need special consideration in the implementation of the algorithm. First, as shown above, \( q(x, c) \), the so-called acceptance rate often determines the probability that we accept the newly generated model. Thus, it should not be too large or too small to prevent the samples from being too concentrated or too scattered, respectively. Usually, it is maintained at ~50% by adjusting the proposal distribution. Second, all the samples need to be independent of one another, but in the beginning, they will be concentrated near the starting model. Therefore, we should discard the first few samples, which is called the “burn-in” process.

3. Data and Model

Based on the high-frequency approximation according to which seismic energy propagates along an infinitesimally thin tube, called the ray path, the basic tomographic forward problem relates the travel time between sources and receivers to the slowness by an integral along the ray path (Aki and Richards, 2002):

\[
d_{\text{pre}}^\text{src}(s) = \int_{ray[x]} s(r) \, dl.
\]

where \( s(r) \) denotes the slowness (reciprocal of velocity), \( d_{\text{pre}}^\text{src} \) is the predicted travel time, and the integral is taken along a seismic ray path. In the actual situation, Equation (11) is highly nonlinear because of the uneven distribution of slowness, but for convenience, this can be neglected in our numerical experiments. For this, the paths will be set to straight lines. Given a set of travel-time data \( d = (t_1, t_2, \cdots t_N) \) recorded for different source–receiver pairs and the discretized slowness distribution (for example, assuming that the imaged volume may be divided into a set of homogeneous cells), the tomographic forward problem can be rewritten as a set of linear equations:

\[
d_{\text{pre}}^i = \sum_{j=1}^M G_{ij} s_j.
\]

where \( s = (s_1, s_2, \cdots s_M) \) is the vector of the discretized slowness distribution and \( G_{ij} \) represents the length of the \( j \)th ray path in the \( i \)th cell.

3.1 Experimental Groups

In this study, we use two 2-D velocity structures in our experiments. Both structures have the same square shape, with 5 km in each dimension. Both are evenly divided into 100 uniform cells. Structure I has a chessboard pattern, with velocity values of 7.0 km/s and 4.6 km/s in alternating cells. Structure II is designed to mimic the Mohorovičić discontinuity (Moho) interface, with velocities of 5.8 km/s and 8.0 km/s in the top and bottom layers, which are 1.5 km thick, respectively, and a 2 km thick gradient zone in between, with velocities in the range of 5.8 to 8.0 km/s. The two velocity structures are shown in Figure 1.

We designed three different inversion settings for each of the two structures. The three inversion settings involve two types of model discretization and two types of source–receiver configurations. For the two types of model discretization D1 and D2, we divided the model evenly into 100 larger and 400 smaller cells, with max-

![Figure 1](image-url)
Inversion Setting D1-C1

Inversion Setting D1-C2

Inversion Setting D2-C2

(a)

(b)

(c)

Figure 2. Model discretization scenarios and source–receiver configurations for the three inversion settings in this study. Black lines show the divisions of the model. Circles and triangles indicate sources and receivers, respectively. Blue lines show the ray paths considered in the experiments. (a) Inversion Setting D1-C1. (b) Inversion Setting D1-C2. (c) Inversion Setting D2-C2.

Figure 3 shows the density distribution of ray paths obtained by counting the lengths of rays passing through each cell. Obviously, the cells containing higher ray densities would yield a better resolution in the inversion because more information about them can be extracted from the travel-time data.

3.2 Generation of the Synthetic Data

In this study, we generate the synthetic travel times in the two velocity structures for the two types of source–receiver configurations in the same way for all the inversion experiments. A synthetic travel-time datum \( t \), the components of which represent the travel times for the rays, is obtained by adding random noise \( \delta t \) to the model prediction \( t_0 \). We assume that \( \delta t \) follows a multivariate normal distribution, of which the standard deviation is 1% of the root mean square (RMS) of \( t_0 \). We run a total of 12 inversion experiments with two inversion methods for all six combinations between the two types of structures and three types of inversion settings. The synthetic data generated in Structure I for Inversion Setting D1-C1 are shown in Figure 4.
3.3 Inversion Parameters

For the MCMC method, we assume that the prior distribution of the velocity follows the multivariate normal distribution. Considering that velocity anomalies generally do not exceed 20% of the reference value at a 95% confidence level, the standard deviation of the velocity is set at $C_r = 0.59$ km/s. We set a burn-in length of 10,000, i.e., we discard the burn-in part of the first 10,000 samples in a total of 410,000 model samples, and select 400 samples with a relatively large sampling interval of 1,000 to ensure the independence of samples. We assume that the proposal probability satisfies the normal distribution, whose standard deviation can control the closeness of adjacent samples. We select a proper standard deviation to ensure that the acceptance rate ranges from 40% to 50%. The sampling parameters and acceptance rates are shown in Table 1. As for the Tikhonov method, only the damping parameter $\lambda$ needs to be determined. We show only one experiment as an example in Figure 5 and list all the parameters in Table 1.

4. Results and Discussion

4.1 Inversion Results

Results from the 12 inversion experiments are shown in Figures 6 and 7, with Figure 6 showing the results for Structure I and Figure 7 showing results for Structure II. It should be noted that we obtain three types of solutions from the inversions: one by the Tikhonov method, $\nu_{\text{Tik}}$ and two by the MCMC method, $\nu_{\text{max}}$, which maximizes the PPDF, and $\nu_{\text{ave}}$, which is the average of all sampled models.

Although all the inversion results display the main characteristics of the velocity structures, there are obvious differences. First, solutions for Inversion Setting D1-C2 are closer to the true velocity structures than those for Inversion Setting D1-C1 (Figures 6a, b and 7a, b). Considering that the model discretization is the same for the two inversion settings, this difference in solutions can be attributed to the difference in the ray path distributions. Because the ray path density is higher in Inversion Setting D1-C2, more information on the velocity structures can be extracted from the data. Second, the difference in results among the three types of solutions is the greatest for Inversion Setting D2-C2 (Figures 6c and 7c). In the Tikhonov solutions, many extreme values occur near boundaries of the structure, indicating that the Tikhonov in-

### Table 1. Parameters used in the inversion experiments in this study

| Velocity structure | Inversion setting | MCMC parameters | Damping parameter |
|--------------------|------------------|-----------------|------------------|
|                    |                  | Standard deviation | Acceptance rate | for regularization method |
| I                  | D1-C1            | 0.015            | 0.498            | 0.363 |
|                    | D1-C2            | 0.008            | 0.491            | 0.363 |
|                    | D2-C2            | 0.012            | 0.461            | 0.363 |
|                    | D1-C1            | 0.022            | 0.499            | 0.062 |
| II                 | D1-C2            | 0.012            | 0.506            | 0.062 |
|                    | D2-C2            | 0.016            | 0.482            | 0.176 |

Figure 4. Synthetic data generated in Structure I (Figure 1a) for source-receiver configuration C1 (Figure 2a). (a) Model-predicted travel times (dots) for all source-receiver pairs plotted by the order of source locations in Figure 2a from top to bottom and from left to right. The travel times are connected by the dotted line as a visual aid. (b) Same as (a) but for synthetic travel-time data. (c) Random noise in travel times. The synthetic data in (b) are obtained by adding the model predictions in (a) and the corresponding random noise in (c).

Figure 5. The $L$-curve for choosing the damping parameter $\lambda$ in the inversion experiment for Structure II with Inversion Setting D2-C2. The x dots mark the positions on the curve for some $\lambda$ values, and the dotted-dashed lines mark the best damping parameter we chose.
Figure 6. Velocity images of Structure I obtained by the Tikhonov regularization method ($v_{\text{Tik}}$) and the MCMC method ($v_{\text{max}}$ and $v_{\text{ave}}$). (a), (b), and (c) are results for Inversion Settings D1-C1, D1-C2, and D2-C2, respectively.

Figure 7. Same as Figure 6 but for Structure II.

Zhao YZ and Wang YB.: Comparing deterministic and stochastic approaches to travel-time inversions
version becomes unstable. On the contrary, fewer extreme values occur in the two MCMC solutions, and the images are close to the real structures and similar to the results for the other two inversion settings. In other words, the MCMC solution is more robust when the model is discretized with smaller cells and the ray path density is increased. In addition, the amplitudes of spatial velocity variations in the MCMC solutions are slightly reduced, resulting in a smoother structure (Figure 6a). This may be because the “accept/reject” criterion in the sampling process prevents large changes between the consecutive samples, which limits the range of velocity variation. Moreover, solution $v_{ave}$ is obtained by averaging over a number of nonoptimal models, by which the extreme values in individual samples may be suppressed.

One advantage of the Bayesian method is that the PPDFs of the model samples are already obtained during the sampling process, which can be used to evaluate the solutions. Figures 8 and 9 display the 95% confidence zones of the MCMC solutions for the depth interval of 2.5–3.0 km. As we can see, although the maximum PPDF and average MCMC solutions are not very close to the true structure at some locations, the 95% confidence zones mostly cover the true velocity. This means that the true velocity structure is mostly distributed in the high-probability region. In addition, the 95% confidence zone is much narrower for Inversion Setting D1-C2 than those for the other two settings, which can be attributed to the fact that the total length of rays crossing each cell is much longer, as discussed before. This also demonstrates that the accuracy of the MCMC inversion can be improved significantly with better data coverage. Another point worth noting is that values with large errors exist in the Tikhonov solution for Inversion Setting D2-C2 (Figures 8c and 9c), and their deviations from the 95% confidence zones are significant. In short, the MCMC solutions and confidence zones reliably recover the true velocity structure, and the accuracy of the single-model estimation by the MCMC method can be improved with better data coverage.

We also analyzed the PPDFs by the histograms of samples in the Markov chains. The results for the cells in the selected areas 1 and 2 in Figure 3c are shown in Figure 10. First, because the ray paths are more concentrated in area 2 than in area 1, the distributions are narrower for area 2. Second, it is obvious that the two areas are both homogeneous in the original true structures, so the velocities in the four cells of each area should be the same. Yet the results from the Tikhonov method deviate greatly from one another (Figure 10a). Nevertheless, the regions with high probabilities in

![Figure 8](image-url)

**Figure 8.** Solutions and 95% confidence zones for the depth interval of 2.5–3 km for Structure I. The solid, dotted, dotted-dashed, and dashed black lines represent the velocities in the true structure, the Tikhonov solutions, and the maximum PPDF and average MCMC solutions, respectively. The solid red lines denote the upper and lower bounds of the 95% confidence zones. (a), (b), and (c) are the results for Inversion Settings D1-C1, D1-C2, and D2-C2, respectively.

Zhao YZ and Wang YB.: Comparing deterministic and stochastic approaches to travel-time inversions
the four cells of each area largely overlap one another, even when the two single-model estimations ($v_{\text{max}}$ and $v_{\text{ave}}$) may be very inaccurate. This suggests that the results from the Tikhonov method become unstable when the model is discretized into small cells. But the PPDFs from the Bayesian method can help us achieve more reliable and meaningful conclusions about the structure.

In addition to the above discussion, we point out some other observations in the inversion experiments. As can be seen, some differences always exist between the two kinds of MCMC solutions in all three types of inversion settings. Intuitively, the average solution may be more robust and closer to the true velocity structure than the maximum PPDF solution. In particular, in all the results for Inversion Setting D2-C2, the images of the average solutions are smoother than those of the maximum PPDF solutions. In fact, in this Bayesian inversion implementation, which is based on a multivariable normal distribution, the average solution is the true maximum PPDF solution in the statistical sense (according to the central limit theorem), whereas $v_{\text{max}}$ in the MCMC process is simply one of the samples. Therefore, $v_{\text{max}}$ may not always be the most reliable solution. The degree of smoothing in the average solution with respect to the maximum PPDF solution may be a relative measure of the accuracy and reliability of the maximum PPDF solution. The smoother the average solution is with respect to $v_{\text{max}}$, the less reliable $v_{\text{max}}$ is.

4.2 Model Errors and Data Fitting Residuals

An important step in evaluating the inversion results involves inspecting the errors in the models and the residuals of data fitting. Table 2 displays the RM5s of the velocity errors and data residuals. Although the differences among the residuals of the three types of solutions are not very large, the Tikhonov solutions seem to be slightly better in terms of data fitting. Only for Inversion Setting D1-C2 can the average solutions by the MCMC method achieve lower data residuals than the Tikhonov solutions. Regarding the velocity errors, the MCMC solutions, especially the average solutions, appear to be much better. Therefore, we may conclude that the MCMC method yields results that are the closest to the true velocity structure.

To illustrate the characteristics of the data residuals more clearly, Figures 11 and 12 display the histograms of the residuals of data fitting by the three types of inversion solutions. All the residual histograms look very similar, with variances of ~30 ms, but some noteworthy differences still exist. As can be seen, the residuals of the Tikhonov solutions are more concentrated near zero than those of the MCMC solutions for Inversion Settings D1-C1 and D2-C2, which means that the Tikhonov solutions fit the data better. However, the difference between the histograms of the Tikhonov and MCMC methods decreases when the data coverage (i.e., the distribution of rays) is improved, which is the case in Inversion Setting D1-C2. These characteristics reveal that the MCMC method...
Figure 10. The posterior probability distribution functions of the velocities in the selected cells labeled as areas 1 and 2 in Figure 3c. The solid, dotted, dotted-dashed, and dashed black lines represent the velocities in the true structure, the Tikhonov solutions, and the maximum PPDF and average MCMC solutions, respectively. UL, UR, LL, and LR represent the upper left, upper right, lower left, and lower right cells, respectively. (a) and (b) are the results of Structures I and II, respectively.

Table 2. Velocity errors and data residuals

| Velocity structure | Inversion setting | Velocity errors (km/s) | Data residuals (s) |
|--------------------|------------------|------------------------|-------------------|
|                    |                  | $v_{\text{max}}$ | $v_{\text{ave}}$ | $v_{\text{tik}}$ | $v_{\text{max}}$ | $v_{\text{ave}}$ | $v_{\text{tik}}$ |
| I                  | D1-C1            | 0.541                 | 0.556                | 0.448             | 0.0100                | 0.0093                | 0.0082                |
|                    | D1-C2            | 0.214                 | 0.191                | 0.432             | 0.0098                | 0.0095                | 0.0096                |
|                    | D2-C2            | 0.648                 | 0.595                | 1.592             | 0.0111                | 0.0101                | 0.0074                |
| II                 | D1-C1            | 0.266                 | 0.212                | 0.299             | 0.0099                | 0.0090                | 0.0084                |
|                    | D1-C2            | 0.186                 | 0.135                | 0.146             | 0.0101                | 0.0098                | 0.0117                |
|                    | D2-C2            | 0.381                 | 0.254                | 0.790             | 0.0101                | 0.0090                | 0.0075                |

Zhao YZ and Wang YB.: Comparing deterministic and stochastic approaches to travel-time inversions
may not be the best approach to fitting the data, but its data-fitting quality improves with an improvement in data coverage. To further investigate the model accuracy, we plot the relative errors in velocity values in the three types of inversion solutions in Figures 13 and 14. The influence of the data coverage in the MCMC method is still obvious. In addition, we find two other features. First, the error distributions of the MCMC solutions for Structure II are slightly better than those of the Tikhonov solutions. The overall error levels are lower, and the error distributions are relatively more uniform. However, the results for Structure I are the opposite. This may be attributable to the intrinsic pattern of the structure. Clearly, the spatial variations in velocity in the chessboard structure are rougher than those in the interface structure, and the effect of smoothing in the Bayesian method may have limited its ability to resolve structures with abrupt changes. Second, the error distributions of the MCMC solutions in Figures 13c and 14c are much more uniform, especially in Figure 14c. For the Tikhonov solutions, there is apparently a strong anti-correlation between the distributions of error and ray density, i.e., larger errors exist near boundaries of the structures where there are fewer ray paths. Comparing the error distributions for the Tikhonov solutions in Figures 13c and 14c with their corresponding data residuals in Figures 11c and 12c, we may conclude that the uniformity of error distributions may have been sacrificed to achieve smaller data residuals.

5. Conclusions
In this study, we conducted inversion experiments using two types of structures with three types of inversion settings to examine the performance of two types of inversion methods, the probabilistic method and the deterministic method. Three types of single-model solutions, the Tikhonov regularization solution and the maximum PPDF and average MCMC solutions are compared. The inversion results are also analyzed by visualizing the PPDFs, and the data residuals from the inversion experiments for Structure I in results $v_{Tik}$, $v_{\text{max}}$ and $v_{\text{ave}}$ (a), (b), and (c) are results for Inversion Settings D1-C1, D1-C2, and D2-C2, respectively.

Figure 11. Data residuals from the inversion experiments for Structure I in results $v_{Tik}$, $v_{\text{max}}$ and $v_{\text{ave}}$ (a), (b), and (c) are results for Inversion Settings D1-C1, D1-C2, and D2-C2, respectively.
distributions of model errors, and residuals in data fitting.

Among the three types of single-model solutions, the stochastic Bayesian solutions are smoother than that obtained by the deterministic Tikhonov method. In practical inversion studies, this may prevent the results from having physically unrealistic features, such as exceedingly large velocity gradients. However, this may also lead to difficulty in resolving sharp features in the structure. The deterministic solutions fit the data better but at the same time may also cause uneven distributions in model errors. Because of this, the inversion results may be unreliable in regions with poor data coverage.

Apart from single-model solutions, Bayesian inversion based on the MCMC process provides the PPDF of the model space, from which we can estimate the confidence zones and evaluate the accuracy of the inversion result. In other words, the Bayesian method can provide more information about the inversion process.

Finally, the maximum PPDF solution obtained from the samples in the MCMC process may sometimes be unreliable. For a linear problem with a normal distribution as the prior, the average solution is theoretically the best choice as a single-model estimate. The reliability of the maximum PPDF solution needs to be carefully evaluated by comparison with the average solution. The smoother the average solution is with respect to the maximum PPDF solution, the less reliable the maximum PPDF model is.

The inversion experiments conducted in this study involve relatively simple but intuitive applications of the deterministic and stochastic inversion theories in order to understand the basic characteristics of these approaches. In general, Bayesian methods may be more robust and reliable than deterministic methods, and the PPDFs can help us make better judgments regarding the quality of the inversion results. On the other hand, deterministic methods are relatively easy to implement and require much less computational effort than stochastic methods. For most problems, deterministic inversion remains the mainstream choice for its convenience, but the stochastic approaches will become more com-

Figure 12. Same as Figure 11 but for Structure II.

Zhao YZ and Wang YB.: Comparing deterministic and stochastic approaches to travel-time inversions
Figure 13. Relative errors in velocity from the inversion experiments for Structure I in results $v_{TiK}$, $v_{max}$, and $v_{ave}$. (a), (b), and (c) are the results for Inversion Settings D1-C1, D1-C2, and D2-C2, respectively.

Figure 14. Same as Figure 13 but for Structure II.
mon with improvements in computer technologies and algorithms.

Acknowledgments
This study was supported by the National Natural Science Foundation of China (grant nos. 41930103 and 41674052). We acknowledge the course English Composition for Geophysical Research of Peking University for the help in improving this manuscript.

References
Aki, K., and Richards, P. G. (2002). Quantitative Seismology (2nd ed). Herndon: University Science Books.
Aster, R. C., Borchers, B., and Thurber, C. H. (2019). Parameter Estimation and Inverse Problems (3rd ed). Amsterdam: Elsevier.
Dettmer, J., and Dosso, S. E. (2012). Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains. J. Acoust. Soc. Am., 132(4), 2239–2250. https://doi.org/10.1121/1.4746016
Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82(4), 711–732. https://doi.org/10.1093/biomet/82.4.711
Green, P. J. (2003). Trans-dimensional Markov chain Monte Carlo. In P. J. Green, et al. (Eds.), Highly Structured Stochastic Systems. (pp. 179-206), London: Oxford University Press.
Hansen, P. C., and O'Leary, D. P. (1993). The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J. Sci. Comput., 14(6), 1487–1503. https://doi.org/10.1137/0914086
Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57(1), 97–109. https://doi.org/10.1093/biomet/57.1.97
Jackson, D. D., and Matsu'ura, M. (1985). A Bayesian approach to nonlinear inversion. J. Geophys. Res. Solid Earth, 90(B1), 581–591. https://doi.org/10.1029/JB090iB01p00581
Matsu'ura, M. (1984). Bayesian estimation of hypocenter with origin time eliminated. J. Phys. Earth, 32(6), 469–483. https://doi.org/10.4294/jpe1952.32.469
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. J. Chem. Phys., 21(6), 1087–1092. https://doi.org/10.1063/1.1699114
Mosegaard, K., and Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. J. Geophys. Res. Solid Earth, 100(B7), 12431–12447. https://doi.org/10.1029/94JB03097
Pachhai, S., Tkalčić, H., and Dettmer, J. (2014). Bayesian inference for ultralow velocity zones in the earth’s lowermost mantle: Complex ULVZ beneath the east of the Philippines. J. Geophys. Res. Solid Earth, 119(11), 8346–8365. https://doi.org/10.1002/2014JB011067
Rothman, D. H. (1985). Nonlinear inversion, statistical mechanics, and residual statics estimation. Geophysics, 50(12), 2784–2796. https://doi.org/10.1190/1.1441899
Schwarz, G. (1978). Estimating the dimension of a model. Ann. Stat., 6(2), 461–464. https://doi.org/10.1214/aos/1176344136
Tarantola, A. (2005). Inverse Problem Theory and Methods for Model Parameter Estimation. Philadelphia: SIAM.
Tarits, P., Jouanne, V., Menvielle, M., and Roussignol, M. (1994). Bayesian statistics of non-linear inverse problems: Example of the magnetotelluric 1-D inverse problem. Geophys. J. Int., 119(2), 353–368. https://doi.org/10.1111/j.1365-246X.1994.tb00128.x
Tikhonov, A. N., and Arsenin, V. Y. (1977). Solutions of Ill-Posed Problems. Washington: V.H. Winston.