University of Tehran Press

Journal of the Earth and Space Physics

Homepage: https://jesphys.ut.ac.ir

Vol. 48, No. 4, Winter 2023, P. 107-124
(Research Article)

Markov Chain Monte Carlo Non-linear Geophysical Inversion with an Improved Proposal Distribution: Application to Geo-electrical Data

Tafaghod Khabaz, Z.[1] | Ghanati, R.[2]

1. Department of Earth Physics, Institute of Geophysics, University of Tehran, Tehran, Iran. E-mail: zahratafaghod73@gmail.com
2. **Corresponding Author**, Department of Earth Physics, Institute of Geophysics, University of Tehran, Tehran, Iran. E-mail: rghanati@ut.ac.ir

(Received: 6 March 2022, Revised: 7 May 2022, Accepted: 4 Oct 2022, Published online: 5 March 2023)

Abstract

Geophysical inverse problems seek to provide quantitative information about geophysical characteristics of the Earth's subsurface for indirectly related data and measurements. It is generally formulated as an ill-posed non-linear optimization problem commonly solved through deterministic gradient-based approaches. Using these methods, despite fast convergence properties, may lead to local minima as well as impend accurate uncertainty analysis. On the contrary, formulating a geophysical inverse problem in a probabilistic framework and solving it by constructing the multi-dimensional posterior probability density (PPD) allow for complete sampling of the parameter space and the uncertainty quantification. The PPD is numerically characterized using Markov Chain Monte Carlo (MCMC) approaches. However, the convergence of the MCMC algorithm (i.e. sampling efficiency) toward the target stationary distribution highly depends upon the choice of the proposal distribution. In this paper, we develop an efficient proposal distribution based on perturbing the model parameters through an eigenvalue decomposition of the model covariance matrix in a principal component space. The covariance matrix is retrieved from an initial burn-in sampling, which is itself initiated using a linearized covariance estimate. The proposed strategy is first illustrated for inversion of hydrogeological parameters and then applied to synthetic and real geo-electrical data sets. The numerical experiments demonstrate that the presented proposal distribution takes advantage of the benefits from an accelerated convergence and mixing rate compared to the conventional Gaussian proposal distribution.

Keywords: Markov Chain Monte Carlo, Non-linear inverse problem, Perturbation models, Principal component analysis (PCA), Proposal distribution.

1. Introduction

Observations of the altered data consist of information regarding the Earth's interior physical properties, which are not directly available to the surface or borehole measurements but are inferred by solving an inverse problem or an inductive reasoning process in a sense of logic. The data-earth interaction is described by a model that states the physical theory, an appropriate subsurface parameterization, and a statistical representation of the data error, which may also be characterized by parameters in the model. Geophysical inverse problems are generally formulated as an ill-posed non-linear optimization problem commonly solved through Newton-based methods. The significant property of the gradient-based approaches (e.g. steepest descent, conjugate gradient, and Landweber iterative scheme) is their fast convergence toward the final solution, but the local linearization of the inverse solution hinders reliable uncertainty appraisal (i.e. underestimated or overestimated uncertainty). In addition, these algorithms may get trapped in local minima if the initial model is far from the convergence region of a global minimum of the cost function so that slight variations in starting model can lead to a notably different subsurface model (Ghanati & Müller-Petke, 2021; Tafaghod Khabaz & Ghanati, 2023). A remedy to the exiting limitations is to employ derivative-free global direct search techniques, e.g. sampling of the posterior probability distribution using Markov Chain Monte Carlo (MCMC) algorithms in the Bayesian framework. The MCMC sampling is essentially a guided-random walk through the probable parts of the posterior model space (Tarantola, 1987; Sambridge & Mosengaard, 2002). With MCMC algorithms, the influence of the initial model

Cite this article: Tafaghod Khabaz, Z., & Ghanati, R. (2023). Markov Chain Monte Carlo Non-linear Geophysical Inversion with an Improved Proposal Distribution: Application to Geo-electrical Data. *Journal of the Earth and Space Physics*, 48(4), 107-124. DOI: http//doi.org/10.22059/jesphys.2022.339477.1007407

Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/jesphys.2022.339477.1007407

Print ISSN: 2538-371X
Online ISSN: 2538-3906

diminishes as the model space sampling progresses. The implementation of the MCMC algorithms requires the sampling of a large number of models, and consequently, the forward calculation of these models can be computationally expensive. Despite the advantages of the Bayesian inversion over the Newton-based optimization approaches, computational efficiency remains a critical factor for high-dimensional model spaces. However, with increasing computational power, the application of the Bayesian inversion methods toward large-scale problems is highly growing. In recent years, several variants of the MCMC algorithms have been proposed in the mathematical and geophysical literature. Beginning from the Metropolis-Hastings algorithm and Gibbs method (Metropolis & Ulam, 1949; Hastings, 1970; Geman & Geman, 1984; Gelfand & Smith, 1990; Gelfand et al., 1990), different investigations on enhancing the efficiency of the MCMC algorithms have been implemented for better performance of the classical approaches. For instance, Haario et al. (2001) proposed a self-tuning algorithm namely the adaptive metropolis algorithm in which proposal values are sampled from a multivariate normal distribution with covariance matrix generated using the accepted samples of the chain. However, it was shown by Cui et al. (2011) that the proposal distribution used by the adaptive Metropolis method can be suboptimal in cases where the posterior distribution is non-Gaussian. Later, they suggested the delayed rejection adaptive Metropolis method (Haario et al., 2006). Ter Braak (2006) introduced the differential evolution Monte Carlo algorithm. Vrugt et al. (2008) applied the idea of combining the differential evolution technique and the adaptive Metropolis to hydrological data. An accelerated variant of the MCMC technique is the parallel tempering algorithm (Swendsen & Wang, 1986) in which multiple chains are simulated but are allowed to swap information. This method can be much more efficient than the classic Metropolis-Hastings algorithm for problems involving multi-modal posterior distributions (Higdon et al., 2002; Dettmer & Dosso, 2012; Sambridge, 2013). Recent examples of parallel tempering as applied to geophysical inversion can be found in Dosso et al. (2012), Ray et al. (2013), Blatter et al. (2018), and Blatter et al. (2021). Despite significant improvements in MCMC sampling methods, choosing an appropriate proposal distribution, used to generate trial moves in the Markov chain, is of crucial importance for an efficient inversion process that creates a well-mixed Markov-Chain, that is, chain samples widely over the model parameters space avoiding both high rejection rates and small steps. Note that the mixing of a Markov chain is the number of steps the Markov chain must take before its probability distribution reaches the stationary distribution (Andrieu & Thoms, 2008). The common choice for proposal density is a Gaussian distribution centered on the current model and variance tuned by the user. The variance parameter determines the step length at each MCMC iteration. When the width of the proposal distribution is too wide, the acceptance ratio is small, and therefore the chain will not mix efficiently and converge only slowly to the target distribution. On the other hand, if the width of the proposal distribution is too narrow, nearly all candidate models are accepted, but the chain mixes again very slowly. In this paper, to improve the functionality of the Metropolis-Hastings sampler in terms of the faster convergence of the chain and good mixing properties, we develop an efficient proposal distribution based on perturbing the model parameters through an eigenvalue decomposition of the model covariance matrix using principal component analysis (PCA). The covariance matrix is retrieved from an early burn-in sampling, which itself commences using a linearized covariance estimate. The proposed algorithm is first applied to a hydrogeological problem to recover the hydrogeological parameters (i.e. transmissivity and specific yield). Then, our algorithm is used for nonlinear inversion of synthetic and real geo-electrical sounding data based on multiple depth layers of the fixed boundary. The article is organized as follows: in section 2 we briefly provide the theoretical framework of the Metropolis-Hastings algorithm and the presented proposal distribution. In section 3, we apply our algorithm to hydrogeological and geophysical synthetic and real data and compare the results to conventional Gaussian

proposal distribution. Finally, we conclude the paper with a brief discussion in section 4.

2. Methodology

The aim of solving an inverse problem in the Bayesian inference framework is to numerically sample the probability density called the posterior distribution for a given set of observations and prior information. By the rules of the conditional probabilities, the posterior can be inverted based on the Bayes' theorem in which the prior probability distribution and the likelihood function are combined and formulated as (Tarantola, 2005):

$$\Omega(\mathbf{m}|\boldsymbol{d}) = \frac{\Omega(\boldsymbol{d}|\mathbf{m})\Omega(\mathbf{m})}{\Omega(\boldsymbol{d})} \quad (1)$$

The term $\Omega(\boldsymbol{d}|\mathbf{m})$ can be interpreted as the likelihood function ($L(\boldsymbol{d}|\mathbf{m})$) that is the density function of the observed data $\boldsymbol{d} \in \mathfrak{R}^{m\times 1}$ given the model parameters $\mathbf{m} \in \mathfrak{R}^{n\times 1}$. The likelihood function depends on the statistics of the noise distribution. The unconditional distribution of the unknowns, $P(\mathbf{m})$, is called the prior distribution. This describes the knowledge about the unknowns that existed before or exists independent of, the current observations. The quantity $\Omega(\boldsymbol{d})$, known as the evidence, is a normalization constant to guarantee that the sum of the posterior probability distribution is unity (i.e. $\int \Omega(\mathbf{m}|\boldsymbol{d})\, d\mathbf{m} = 1$). Here, italics present scalar quantities, boldface lowercase letters vectors, and boldface capital letters matrices. The superscript T is used to indicate the transpose of an operator.

Assuming that the data error to be independent uncorrelated zero-mean Gaussian, we define the likelihood function by the relation (Tarantola, 2005):

$$L(\boldsymbol{d}|\mathbf{m}) = \frac{1}{2\pi^{m/2}(\prod_{i=1}^{m}\sigma_i^2)^{1/2}}.exp\left\{-0.5\big(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m})\big)^T \boldsymbol{C}_d^{-1}(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m}))\right\} \quad (2)$$

Equivalently

$$L(\boldsymbol{d}|\mathbf{m}) = \frac{1}{2\pi^{m/2}(\prod_{i=1}^{m}\sigma_i^2)^{1/2}}.exp\left(-\sum_{i=1}^{m}\frac{(\boldsymbol{d}_i - \boldsymbol{F}(\mathbf{m})_i)^2}{\sigma_i^2}\right) \quad (3)$$

Whereas computing the log-likelihood is preferable, taking the logarithm of Equation (3) results in:

$$\log L(\boldsymbol{d}|\mathbf{m}) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log(\prod_{i=1}^{m}\sigma_i^2) - \frac{1}{2}\big(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m})\big)^T \boldsymbol{C}_d^{-1}(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m})) \quad (4)$$

If $\sigma_i = constant$,

$$\log L(\boldsymbol{d}|\mathbf{m}) = -\frac{m}{2}\log(2\pi) - \frac{m}{2}\log(\sigma_i^2) - \frac{1}{2}\big(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m})\big)^T \boldsymbol{C}_d^{-1}(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m})) \quad (5)$$

where $\boldsymbol{C}_d$ is the data error covariance matrix, $\boldsymbol{F}$ is the non-linear forward operator, and σ is the data error.

We incorporate smoothness constraints into the model parameters by imposing independent normal distributions to the vertical model gradient. Hence, we define zero-mean normal prior distributions concerning the vertical resistivity gradient as follows:

$$p(\mathbf{m}) = \frac{1}{2\pi\beta^2}\exp[-\frac{1}{2\beta^2}(\mathbf{m}^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{m})] \quad (6)$$

Taking the logarithm of Equation (6) results in the following statement, which is called log-prior probability density function:

$$\log p(\mathbf{m}) = -\log(2\pi\beta^2) - \frac{1}{2\beta^2}(\mathbf{m}^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{m}) \quad (7)$$

where β is analogous to model regularization weights used in the deterministic inversion. The smaller the value of β, the higher the weight given to the regularization. To effectively solve the inverse problem, the posterior distribution $\Omega(\mathbf{m}|\boldsymbol{d})$ is sampled by the Metropolis-Hastings (MH) sampler (Metropolis et al., 1953; Hastings, 1970), which is an MCMC algorithm. The MH algorithm draws a sequence of random samples (solutions) from the posterior density that has the ergodic property, i.e. that allows expectations over the posterior distribution to be replaced by averages over the chain. Informally, we think of an ergodic chain as one that spends time in each region of parameter space proportional to the posterior probability of that region (Cui et al., 2011). The MH method proceeds in two steps. At the first step, a candidate model m^p is created using a proposal density $p(\mathbf{m}^p|\mathbf{m}^c)$ based on the current model $\mathbf{m}^c$ (i.e., the proposed method generates perturbed values, which are added to current parameters leading to new samples); at the second step, it is decided that the candidate (proposal) model is either accepted or rejected using the MH acceptance condition. The MH algorithm is summarized in Table 1 in the context of our new proposal distribution.

Table 1. Metropolis-Hastings based MCMC algorithm.

Input: $\mathbf{m}^0 \in \mathbb{R}^{m\times 1}$, forward function $F(\mathbf{m})$, M (number of iterations), proposal distribution $q(\mathbf{m}^p\|\mathbf{m}^c)$,$\rho^{lb}$(lower boundary), ρ^{ub}(upper boundary), $\gamma = 1$ and $\mu = 0$.
Output: a matrix $m \times M$ of the model parameters
for $k = 1$ to $M - 1$ do
Compute a new model proposal $\mathbf{m}^p$ from $q(\mathbf{m}^p\|\mathbf{m}^c)$ where $\mathbf{m}^p$ is generated in principal component space
Check whether the resistivity of each layer falls within the interval $[\rho^{lb}, \rho^{ub}]$, otherwise draw a new model proposal $\mathbf{m}^p$
Calculate $\alpha(\mathbf{m}^p\|\mathbf{m}^c) = \min(1, \frac{\mathrm{p}(\boldsymbol{d}\|\mathbf{m}^p)\mathrm{p}(\mathbf{m}^p)\ \mathrm{q}(\mathbf{m}^p\|\mathbf{m}^c)}{\mathrm{p}(\boldsymbol{d}\|\mathbf{m}^c)\mathrm{p}(\mathbf{m}^c)\mathrm{q}(\mathbf{m}^c\|\mathbf{m}^p)})$
Draw U$\sim\mathcal{N}(\mu,\gamma)$
If $\log U < \log \alpha\,(\mathbf{m}^p\|\mathbf{m}^c)$ then
Accept $\mathbf{m}^0 = \mathbf{m}^p$
else
$\mathbf{m}^0 = \mathbf{m}^{k-1}$
end If
end for

The acceptance probability, $\alpha(\mathbf{m}^p|\mathbf{m}^c)$, of the candidate model judges whether the new model is accepted to the Markov Chain or rejected. In other words, the acceptance probability is the key to guaranteeing that the Markov chain converges to the posterior distribution since it draws samples according to the posterior's density. In the case of symmetric proposal density, where the move from $\mathbf{m}^c$ to $\mathbf{m}^p$ is equally as likely as a move from $\mathbf{m}^c$ to $\mathbf{m}^p$, the acceptance probability can be written as:

$$\alpha(\mathbf{m}^p|\mathbf{m}^c) = \min(1, \frac{\mathrm{p}(\boldsymbol{d}|\mathbf{m}^p)\mathrm{p}(\mathbf{m}^p)}{\mathrm{p}(\boldsymbol{d}|\mathbf{m}^c)\mathrm{p}(\mathbf{m}^c)}) \qquad (8)$$

It is assumed that all the parameters within $\mathbf{m}^p$ fall within the prior bounds for the allowed variation of electrical resistivity. The proposal distribution from which we choose new points for the chain can be quite arbitrary, but the choice of a distribution that most closely resembles the true target density can considerably accelerate the convergence of the values created to the correct distribution. The closer the proposal distribution q is to the actual target $\Omega(\mathbf{m}|\boldsymbol{d})$, the better the chain mixes and the better a short sequence represents a random draw from the posterior. It is particularly true in multi-dimensional problems and when there is a correlation between the elements of the model parameters. The usual choice for proposal density is a Gaussian distribution centered on the current model and standard deviation adjusted by the user. Using this strategy for cases with many model parameters is usually laborious in terms of computing time and user input since many short trial runs have to be made. To overcome this difficulty, we present an efficient proposal scheme based on the principal component transformation and eigenvectors decomposition of the covariance matrix estimated from the deterministic inversion. To generate new (perturbed) model parameters using our proposal method, one needs to implement the following stages:

1) Calculate the model covariance matrix using the linearized inversion (deterministic inversion) results, that is, $\boldsymbol{C}_m = J^{\dagger}J$, where J is the jacobian matrix and $J^{\dagger}$ is the regularized generalized inverse of the matrix J.

2) Decompose the covariance matrix $\boldsymbol{C}_m$ in terms of the diagonal matrix $\boldsymbol{D}$ of eigenvalues and matrix $\boldsymbol{V}$ whose columns are the corresponding right eigenvectors

3) Transform the model parameters (the current model) into principal component space leading to the rotated parameters: $\widehat{\mathbf{m}} = \boldsymbol{V}^T\mathbf{m}$, where T is the transpose.

4) Draw parameter perturbations using a Cauchy proposal distribution with $\boldsymbol{D}$ degrees of freedom (scaling factor). In the Matlab environment,

$\mathbf{m}^{perturb} = \mathrm{trnd}(\mathrm{diag}(\boldsymbol{D}), m, 1)$, where m is the number of model parameters.

5) Calculate the candidate model (proposal model) through adding the rotated parameters ($\widehat{\mathbf{m}}$) to the parameter perturbations derived from step 4, that is, $\widehat{\mathbf{m}} = \widehat{\mathbf{m}} + \mathbf{m}^{perturb}$

6) Rotate the proposal model back to the original parameters $\mathbf{m} = \boldsymbol{V}\widehat{\mathbf{m}}$.

Step 1 is implemented only at the first iteration. Note that the starting model of the MH algorithm corresponding to the first iteration is obtained from the linearized inversion. In the case of the generation of parameter perturbations in stage 4, Gaussian distribution can be used instead of Cauchy distribution, but note that Gaussian perturbations may not be large enough to transit between potentially disjoint regions of high probability. The Cauchy distribution has heavier tails causing much wider sampling, and consequently better exploring the model space. Hence, we employ a Cauchy proposal distribution in rotated space, scaled according to the estimated rotated covariance matrix. As earlier mentioned, the covariance matrix obtained from the deterministic inversion is used during the burn-in phase, and as sampling progresses the covariance matrix is adaptively replaced with an alternative computed based on averaging over successive models along the Markov chain, as follows:

$$\boldsymbol{W}_{\mathbf{m}} = \sum_{i=1}^{n}(\mathbf{m}_i - \mathbf{m}_{i-1})(\mathbf{m}_i - \mathbf{m}_{i-1})^T \qquad (9)$$

From the theory of the MCMC algorithm, each sample is only dependent on the sample drawn directly before it and not earlier samples in the chain, however, using the adaptive covariance matrix constructed by the previous values violates the Markovian property. However, it can be proved that the proposed algorithm maintains the correct ergodicity properties.

We also impose lower and upper limits of the resistivity values based on the physical meaning of the resistivity distribution after drawing the proposed model. It means that the candidate models, which are out of the prior information are discarded and the sampling process is repeated. In the following, we first apply the proposed algorithm to two synthetic examples in terms of hydrogeological and geo-electrical problems and a real 1D resistivity data set.

3. Numerical examples

In this section, the functionality of our proposed scheme in comparison with the conventional Gaussian proposal distribution is tackled by synthetic and real examples in terms of the convergence rate of the chain and mixing properties.

3-1. Hydrogeological example (Slug test)

We first compare the presented algorithm with the traditional by inverting the hydrogeological parameters. A common practice in hydrogeology for determining the storage coefficient and transmissivity of a water-bearing layer is called the slug test. A known volume of water (Q) is injected into a well, and the variations of the water level (y) at an observation well a distance (d) away from the injection well are measured at different times. The water level recorded at the observation well increases rapidly and then decreases more slowly. The objective is to estimate the transmissivity (S) and the storage coefficient (T) based on the field observations. Mathematically, the function corresponding to the Slug test is defined as (Ferris & Knowles, 1963).

$$y = \frac{Q}{4\pi Tt}\exp(-Sd^2/4Tt) \qquad (10)$$

It is seen that the data (water level) are nonlinearly related to the unknown parameters leading to a non-linear optimization problem with a convex cost function. Using Equation (2) and the parameters defined in Table 2, the synthetic data are generated and corrupted by 5 percent uncorrelated Gaussian-distributed noise of zero mean. The noise level (standard deviation of the Gaussian distribution) is defined based on the amplitude of each datum.

Table 2. True parameters defined for the Slug test.

$Q\ (m^3)$	T	$S\ (m^2/hr)$	$d\ (m)$	$t\ (hr)$
50	0.58	0.002	60	$[5, 10, 20, 30, 40, 50]$

To estimate the model parameters of the Slug test, the MCMC algorithm based on the Gaussian proposal density and the PCA-based proposal distribution is implemented. The conventional MCMC algorithm is initialized from $[S_0 = 0.1, T_0 = 3]$ with a total of 1,000,000 iterations while the proposed sampling method starts with the model parameters derived from the linearized inversion with a total of 250,000 iterations with burn-in of 50 percent. It is well-known that the samples generated/accepted at the beginning of the chain during the burn-in phase may not precisely represent the PPD function. As a result, these samples are usually discarded in the computation of posterior distribution. It is also noticed that the length of the burn-in stage depends on the complexity of the problem being solved and on the shape and size of the proposal distribution used to implement the MCMC algorithm. Figures 1 and 2 illustrate the outcome of both algorithms. It is observed that both algorithms well sample the target distribution, however, the proposed method converges to the target distribution with less number of iterations compared to the conventional proposal density (100,000 iterations versus 500,000 iterations). This property can be significant when the forward calculation is computationally expensive, and it is ideal to explore the target distribution with less number of the forward modeling evaluations. The autocorrelation plots of the log-posteriors in terms of different lags of the samples generated by the Gaussian proposal distribution and the presented algorithm are shown in Figures 1 and 2, respectively. From the autocorrelation results, there is higher autocorrelation in the samples obtained from the conventional scheme compared to those of the proposed strategy. The autocorrelation function is a useful proxy to evaluate the sample variability and the mixing of the Markov chain. A high autocorrelation at lags simply indicates a rather poor mixing of the chain.

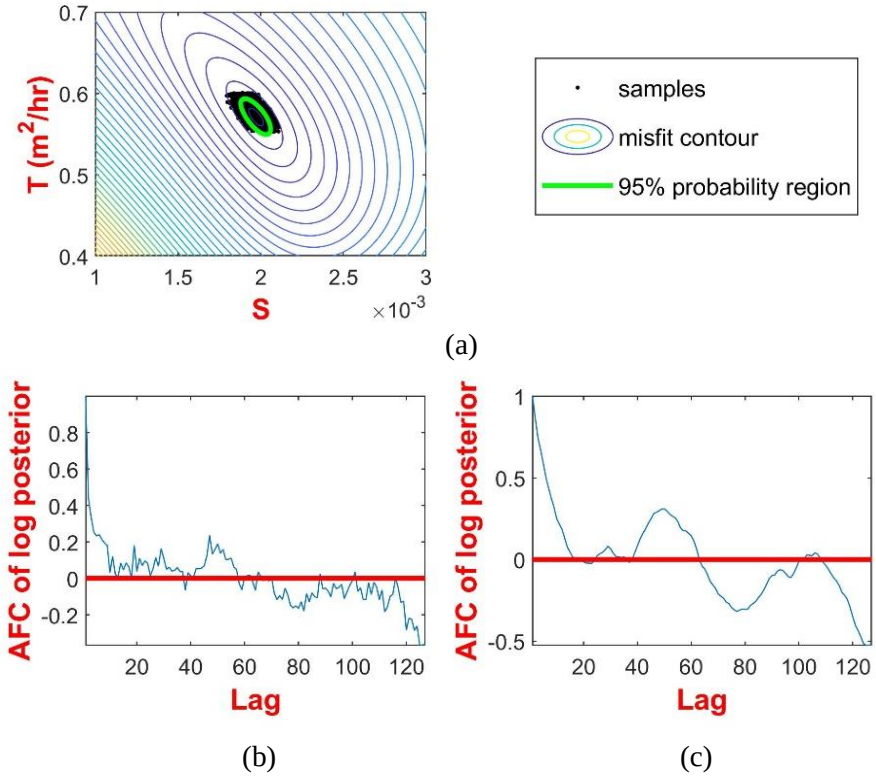


Figure 1. The resulting MCMC sampling using the conventional Gaussian proposal distribution: a) Two-dimensional scatter plot of the samples from the Slug test, b) the autocorrelation plot of different lags from sample S, and c) the autocorrelation plot of different lags from sample T generated by using the conventional Gaussian proposal distribution-based MCMC sampling method.

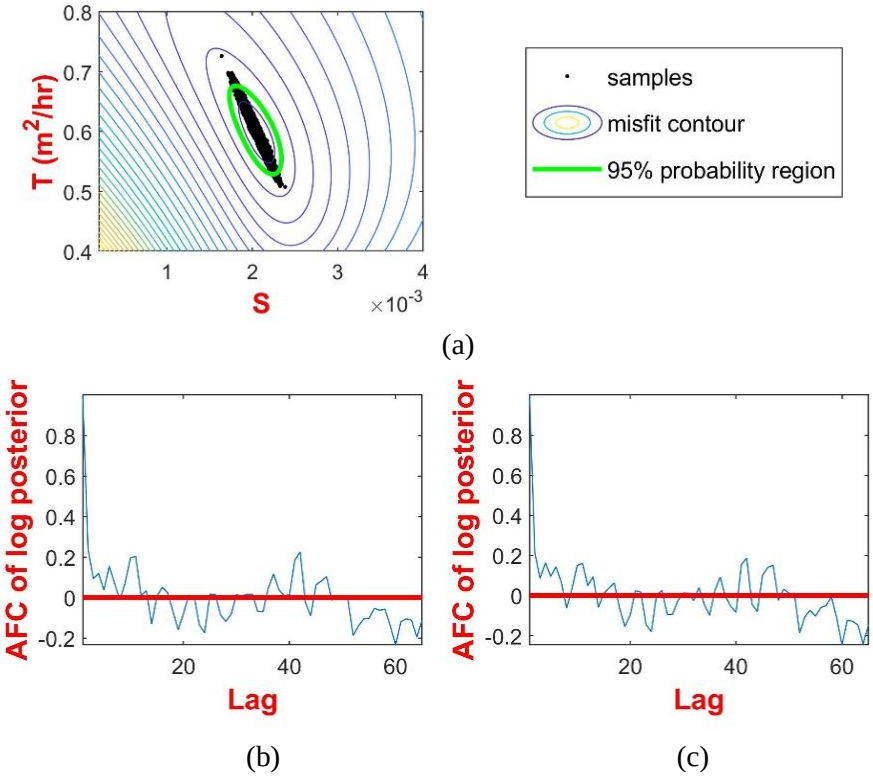


Figure 2. The resulting MCMC sampling using the presented proposal distribution: a) Two-dimensional scatter plot of the samples from the Slug test, b) the autocorrelation plot of different lags from sample *S*, and c) the autocorrelation plot of different lags from sample *T* generated by using the PCA-based proposal algorithm.

To better demonstrate the potential superiority of the PCA-based proposal for parameter perturbations, Figures 3 and 4 display the MCMC sampling history for two parameters, *S* and *T* generated by applying a Gaussian proposal density, and by applying the PCA proposal scheme, respectively. Comparing the sampling histories, the PCA proposal ameliorates mixing of the chain and creates an MCMC chain varying more rapidly (appears to have higher frequency content).

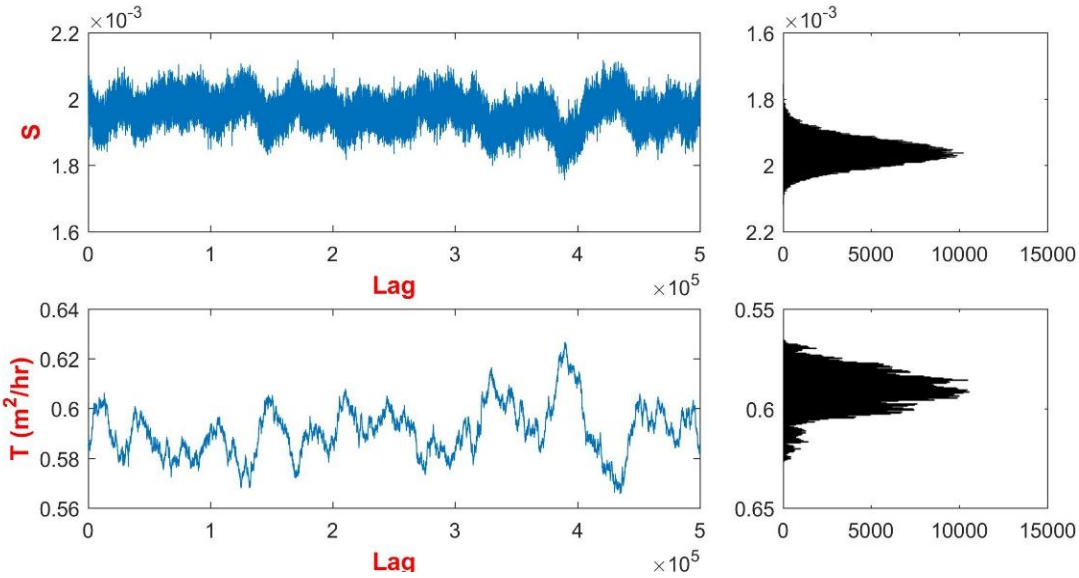


Figure 3. The trace plots of samples corresponding to the Slug test generated using the conventional Gaussian proposal distribution-based MCMC sampling method on the left and the resulting histogram on the right.

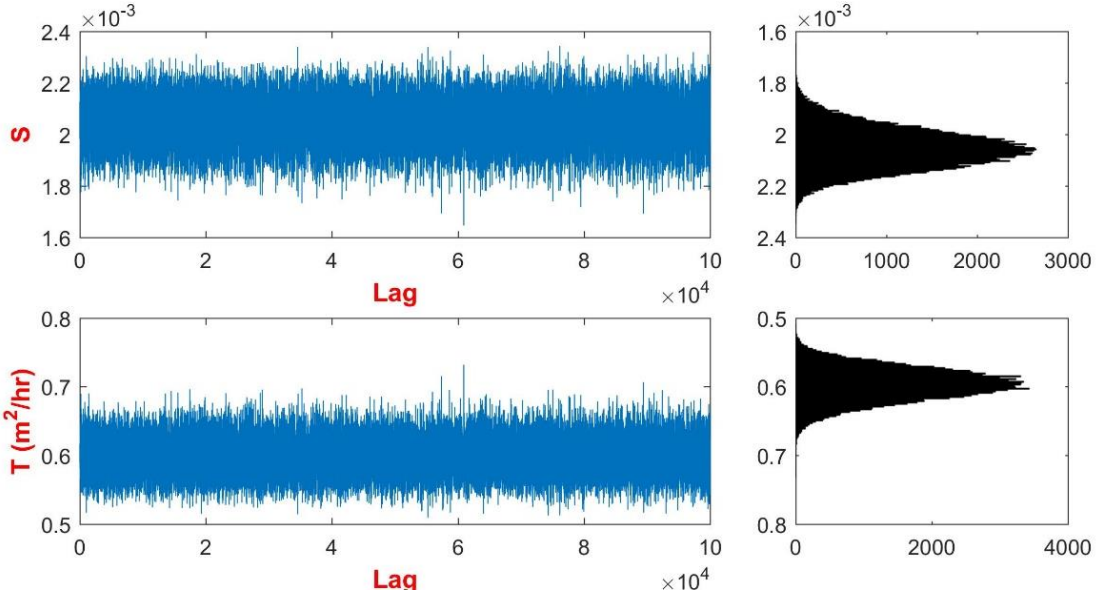


Figure 4. The trace plots of samples corresponding to the Slug test generated using the PCA proposal distribution on the left and the resulting histogram on the right.

This ameliorated sampling is demonstrated further by the marginal densities displayed to the right of the trace plots of samples. The marginal density for the PCA proposal converges to relatively smooth distributions (in particular, T) while those for the Gaussian proposal are rough and appear under-sampled.

3-2. Synthetic data examples

Here, we evaluate our novel proposal distribution versus the conventional method by applying the proposed algorithm to synthetic data computed for two models based on multiple depth layers of fixed boundary (i.e. invariant geometry). In invariant geometry inversion (also called smooth inversion) with predefined layer thicknesses of logarithmically increasing with depth, a distribution of the subsurface resistivity is recovered. As a general rule, the main advantage of using the smooth inversion compared to the block inversion in 1D geophysical modeling (e.g. electromagnetic and geo-electrical sounding) is that the smooth inversion requires little a priori knowledge about the subsurface layering characteristics bringing about smoothly varying features in the resultant model. However, in cases where a known number of layers are present, the block inversion strategy may be preferable allowing to construct piecewise constant resistivity profiles. The example uses a synthetic earth model including five-layered earth with a shallow and a deep unconfined aquifer that is separated by a low resistive layer acting as an aquiclude basement to the shallow aquifer. Table 3 summarizes the geoelectrical parameters associated with the synthetic case.

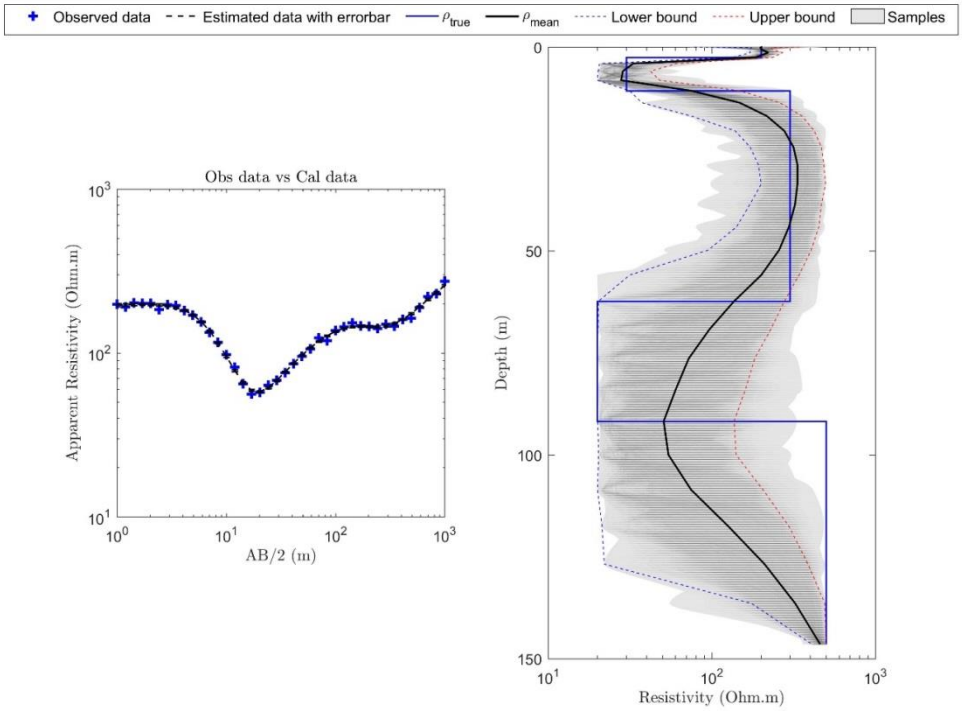
Table 3. True geo-electrical parameters of the synthetic five-layer earth example

geo-electrical parameters					
Resistivity ($\Omega.m$)	$\rho_1 = 200$	$\rho_2 = 30$	$\rho_3 = 300$	$\rho_4 = 20$	$\rho_5 = 500$
Thickness (m)	$h_1 = 2.5$	$h_2 = 8.7$	$h_3 = 51.3$	$h_4 = 30$	Half-space

Knowing the physical theory of direct current geo-electrical sounding relating the model parameters **m**, composed of M uniform isotropic layers with the resistivity values $[\rho_1, \rho_2, \dots, \rho_M]$ and layer thicknesses $[h_1, h_2, \dots, h_{M-1}]$, and the apparent resistivity values ρ_a, the forward response in terms of the Schlumberger configuration is calculated by an integral equation as follows:

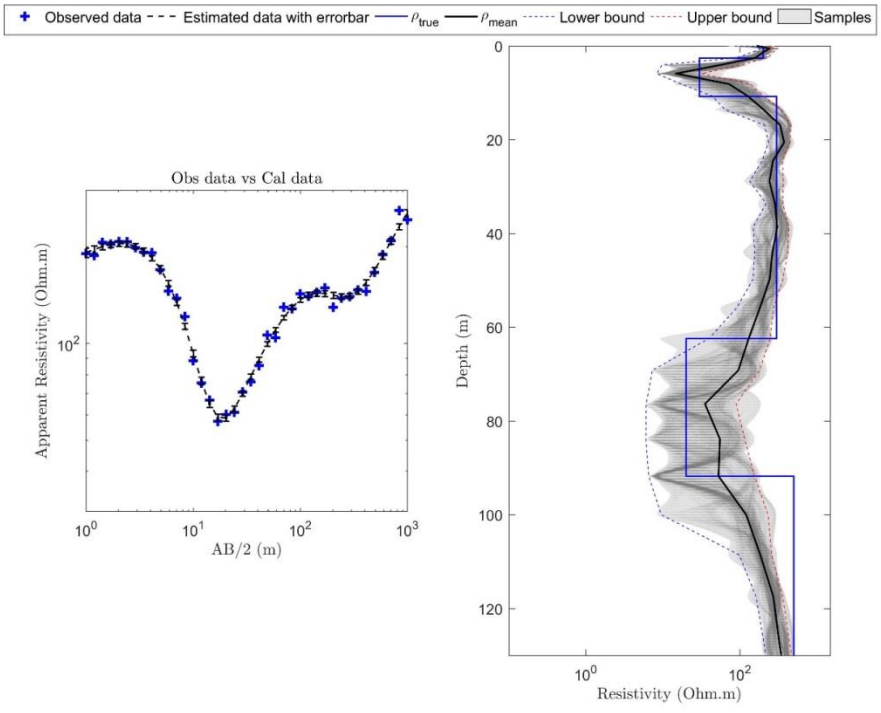
$$\rho_{sa}(L,m) = S^2\left\{\int_0^\infty W(\rho_i, h_i; \lambda)\, J_1(\lambda S)\lambda d\lambda\right\}; i = (1, 2, \dots, M) \quad (11)$$

where S denotes half the current electrode spacing, λ is the integration variable, $J_1(\lambda S)$ is the first-order Bessel function of the first kind, and $W(\rho_i, h_i; \lambda)$ stands for the resistivity transform function calculated using the recurrence relationships from the bottom to the surface. This integral equation can be solved using a linear digital filter based on the Johansson method. Appendix A provides a detail of the recurrence formulae for all kernel functions. Two approaches for solving the geo-electrical sounding inverse problem are commonly used in terms of layer boundaries. These methods are defined based on models with variant and invariant geometry. For variant geometry inversion, the models are divided into a few layers with variable boundaries where both the values of the resistivity and thickness in each layer are allowed to vary. The invariant geometry inversion is based on the assumption that the Earth is divided into many layers with fixed boundaries so that only the resistivity in each layer is allowed to vary. In this paper, we follow the latter strategy. The synthetic measurements are created using a Schlumberger array with 16 $AB/2$ spreads ranging from 1.25 to 1000 m on the simulated earth model with the geo-electrical parameters represented in Table 3. Then, the data are contaminated with uncorrelated Gaussian random noise having a standard deviation σ^{prior} of 5% of the amplitude of each datum, which was then considered as the prior data noise level during inversion. To explore the posterior model space, the MCMC sampler is implemented in two steps; at the first stage, a model including the subsurface resistivity distribution is proposed using the proposal density function, and at the second stage, the candidate model is either accepted or rejected depending on the likelihood of the model compared to the likelihood of the last accepted model. We run the MCMC algorithm with this synthetic dataset using the Gaussian proposal density and the PCA-based proposal distribution for 500,000 and 250,000 iterations, respectively. During the Bayesian inversion, lower and upper limits for the model parameters are defined such that the resistivity values are confined between $[0, 1000]\, \Omega.m$. The resulting Bayesian inversion corresponding to the Gaussian proposal distribution and the PCA proposal scheme are illustrated in Figures 5 and 6, respectively. Both MCMC algorithms capture the true model in the sense that it lies within the 95 percent confidence interval at nearly all depths. In addition, the MCMC sampling provides uncertainty estimates for resistivity at each depth-estimates that depend on the data, the forward modeling and the model parameterization. To further assess the results of this example, we perform MCMC diagnostics to evaluate the MCMC sample quality. We generate individual trace plots for four parameters ($\mathbf{m}_5, \mathbf{m}_{10}, \mathbf{m}_{15}$, and $\mathbf{m}_{20}$) of the subsurface layers obtained by applying a Gaussian proposal density (see Figure 7), and by applying the PCA proposal method (see Figure 8) to assess the chain convergence. Therein, we observe that the proposed algorithm is better mixing and has reached the stationary region of the target density with less number of iterations compared to the conventional method.



(a) (b)

Figure 5. Posterior model results derived from the Gaussian proposal density-based MCMC method for the five-layer synthetic case: a) observed data versus calculated data with the error bars, b) the posterior probability density of resistivity as a function of depth. The black line shows the mean posterior models and the blue line indicates true values of the model parameters. Lower and upper intervals computed based on 95 percent confidence interval are displayed by dashed blue and red lines, respectively.



(a) (b)

Figure 6. Posterior model results derived from the PCA proposal density-based MCMC method for the five-layer synthetic case: a) observed data versus calculated data with the error bars, b) the posterior probability density of resistivity as a function of depth. The black line shows the mean posterior models and the blue line indicates true values of the model parameters. Lower and upper intervals computed based on 95 percent confidence interval are displayed by dashed blue and red lines, respectively.

3-3. Real data example

Given the success of the synthetic example, we finally provide a field data set with known geology information aiming at further appraising the efficacy of the proposed procedure. Again, we compared the results of the proposed scheme with those of the conventional method. The field data has been acquired on the German North Sea Island Borkum, utilizing the Schlumberger configuration including 23 apparent resistivity records with current electrode spacing ranging logarithmically from 1.5 m to 150 m.

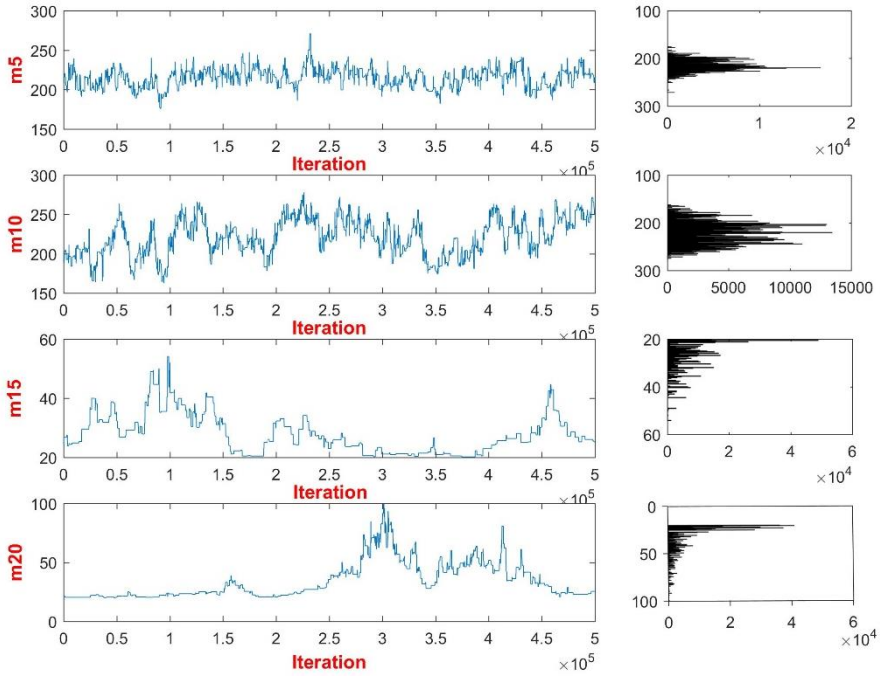


Figure 7. The trace plot of posterior samples corresponds to the synthetic example generated by the conventional Gaussian proposal distribution-based MCMC sampling method on the left and the resulting histogram on the right. Here, only four model parameters variation in terms of different iterations is shown.

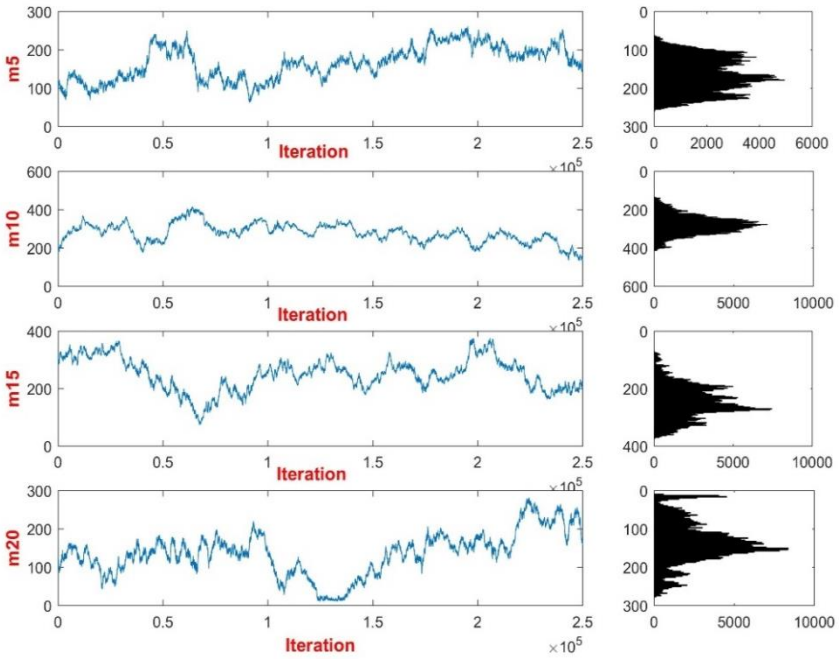


Figure 8. The trace plot of posterior samples corresponds to the synthetic example generated by the PCA proposal distribution-based MCMC sampling method on the left and the resulting histogram on the right. Here, only four model parameters variation in terms of different iterations is shown.

This site was chosen based on the accessibility of nearby lithology borehole information. Based on the drill hole, the geology section is characterized by a vadose zone with a thickness of 3 m, followed by the first aquifer composed of well-sorted fine sands until a depth of 23 m, followed by alternating sequences of fine sand and clay to a depth of 32 m as an aquitard basement to the shallow aquifer. The second fresh water-bearing layer at about 32 to 49 m can also be identified, which is composed of poorly sorted fine sands mixed with fine clay layers. There is a transition zone from fresh to saline water separating aquifer 2 and aquifer 3 at this depth (Günther & Müller-Petke, 2012). We follow the inversion strategy described for the synthetic experiment. We implement the MCMC sampler with the Gaussian proposal density and the PCA-based proposal distribution for 300,000 and 150,000 iterations, respectively. From the geological information, during the Bayesian inversion, lower and upper limits for the model parameters are defined such that the resistivity values are confined within the interval $1 \leq \rho(\Omega.m) \leq 5000$. Figure 9 shows the result provided by the conventional approach whereas Figure 10 illustrates the posterior models estimated by the PCA proposal density-based MCMC algorithm. When viewing Figures 9 and 10, we notice that both strategies give rise to relatively similar results; however, the geo-electric profiles of the Gaussian proposal density show higher variability than the PCA proposal procedure. In other words, the regions with larger uncertainty are characterized by broader distribution. Referring to the resulting geo-electric profiles and as would be expected since the variant geometry inversion requires a priori information about the layer thicknesses, the estimated uncertainty is consistently larger than the invariant geometry inversion.

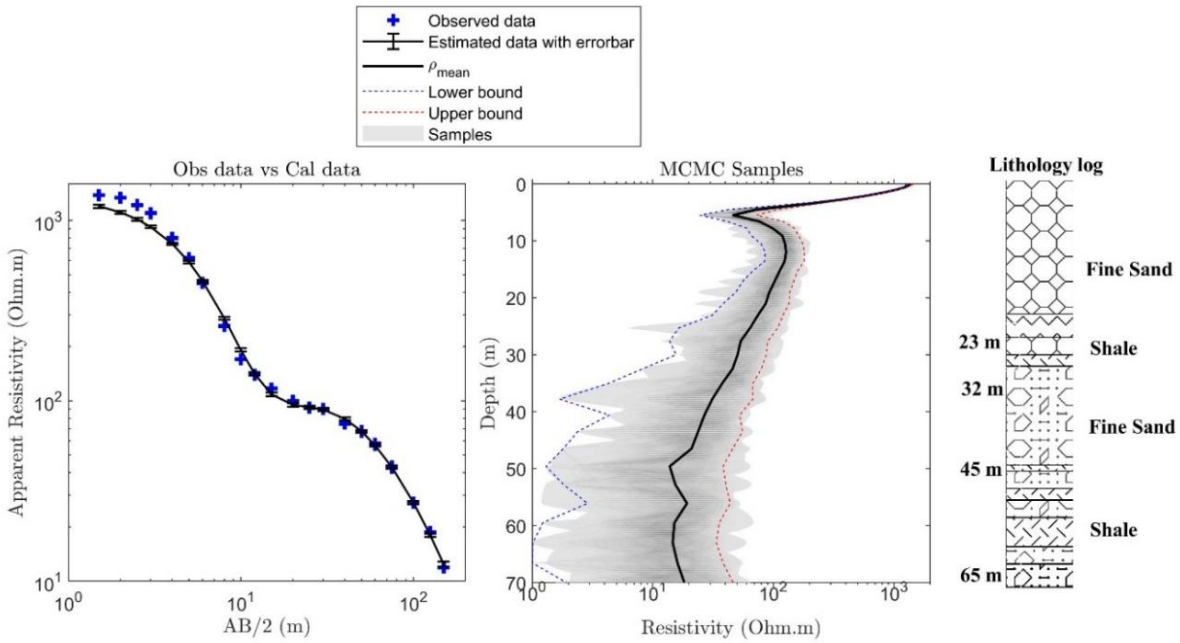


Figure 9. Posterior model results derived from the Gaussian proposal density-based MCMC method for the real case: a) observed data versus calculated data with the error bars, b) the posterior probability density of resistivity as a function of depth, c) information of a borehole lithology. The black line shows the mean posterior models and the blue line indicates true parameter values. Lower and upper intervals computed based on 95 percent confidence interval are displayed by dashed blue and red lines, respectively.

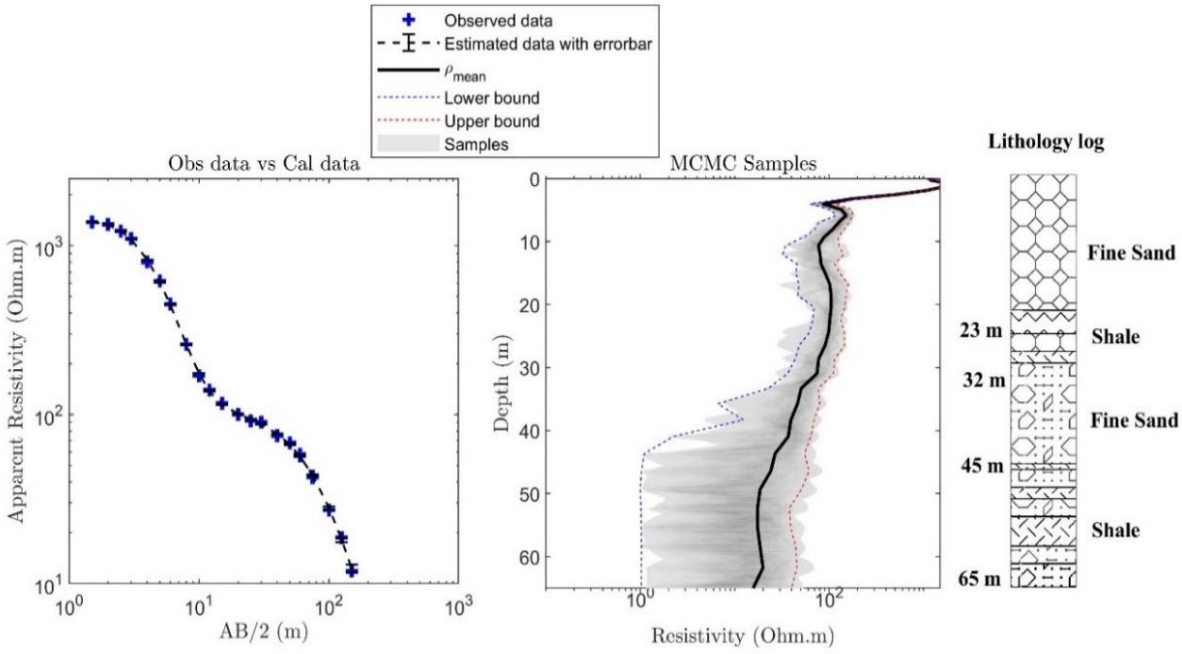


Figure 10. Posterior model results derived from the PCA proposal density-based MCMC method for the five-layer synthetic case: a) observed data versus calculated data with the error bars, b) the posterior probability density of resistivity as a function of depth, c) Information of a borehole lithology. The black line shows the mean posterior models and the blue line indicates true parameter values. Lower and upper intervals computed based on 95 percent confidence interval are displayed by dashed blue and red lines, respectively.

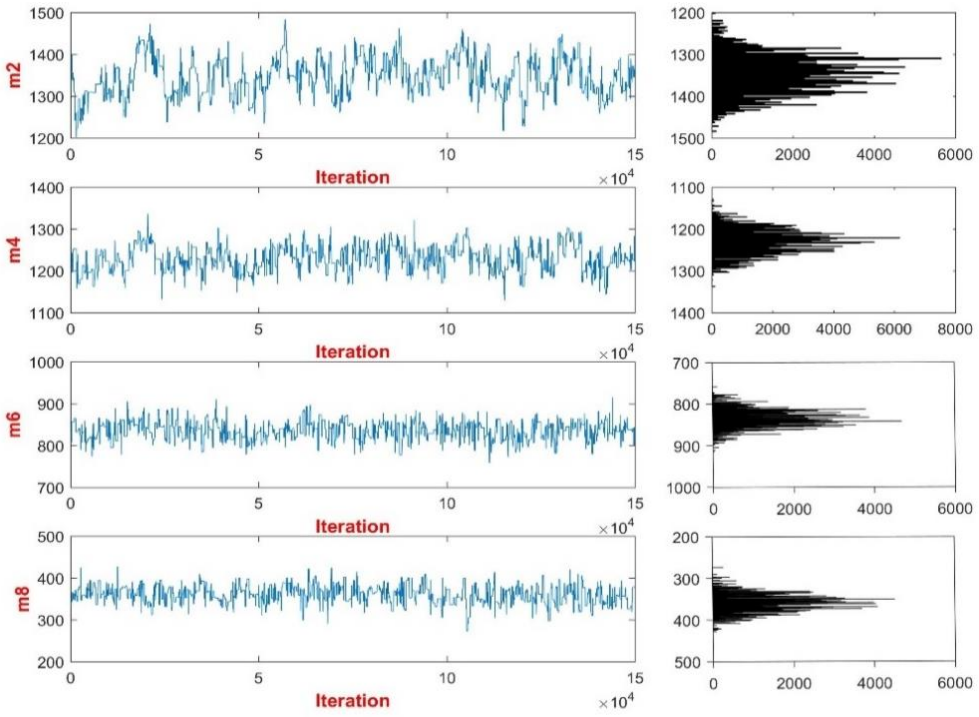


Figure 11. The trace plot of posterior samples corresponds to the real example generated by the conventional Gaussian proposal distribution-based MCMC sampling method on the left and the resulting histogram on the right. Here, only four model parameters variation in terms of different iterations is shown.

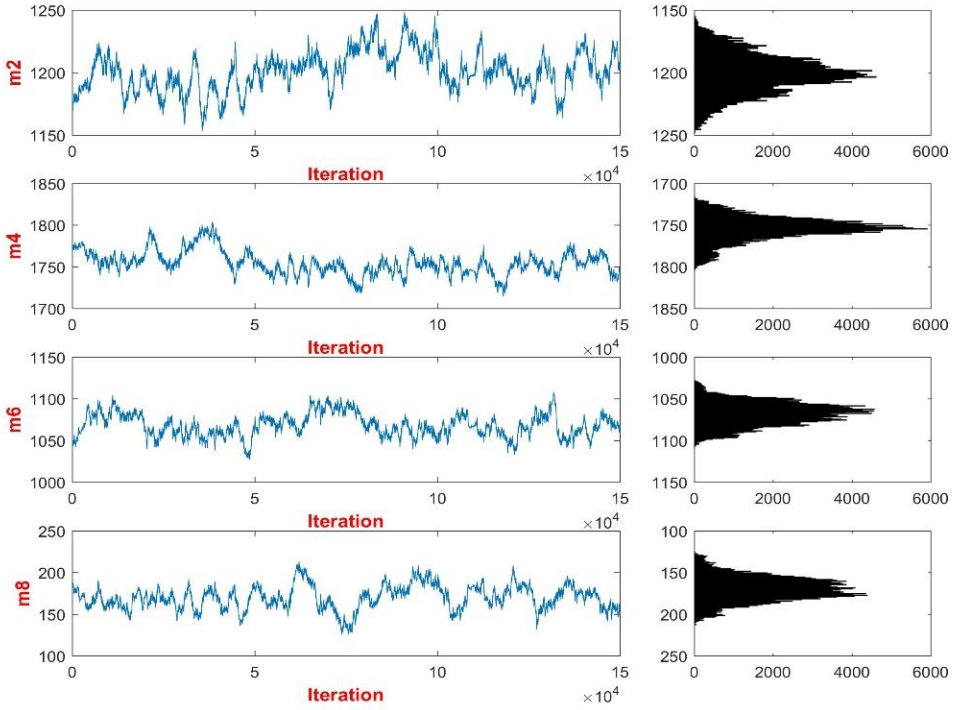


Figure 12. The trace plot of posterior samples corresponds to the real example generated by the presented proposal distribution-based MCMC sampling method on the left and the resulting histogram on the right. Here, only four model parameters variation in terms of different iterations is shown.

To further assess the results of this example, we perform MCMC diagnostics to evaluate the MCMC sample quality. We produce individual trace plots for four neighboring model parameters, ($\mathbf{m}_2, \mathbf{m}_4, \mathbf{m}_6$, and $\mathbf{m}_8$) of the subsurface layers obtained by applying a Gaussian proposal density (see Figure 11), and by applying the PCA proposal method (see Figure 12) to assess the chain convergence. Therein, the Gaussian proposal scheme exhibits a slower mixing rate and higher correlation for the corresponding parameters leading to not extensively exploring the sample space compared to the proposed algorithm. Furthermore, our algorithm has reached the stationary region of the target density with less number of iterations (here, the trace plot of posterior samples are shown after discarding the burnin phase samples) compared to the conventional method.

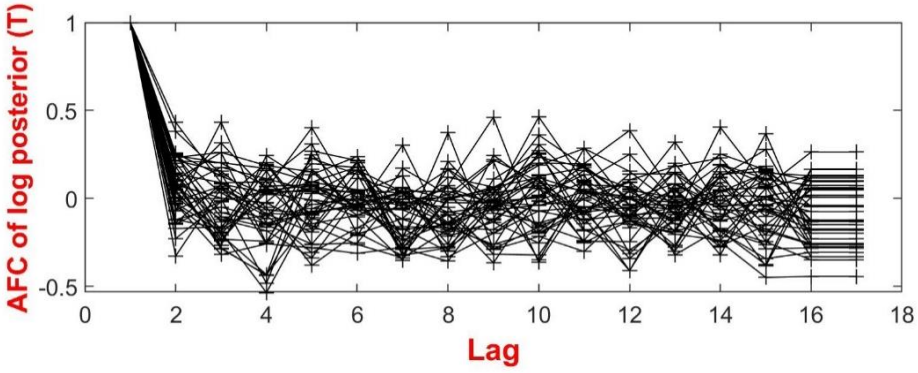


Figure 13. The autocorrelation plot of different lags for the posterior density derived from the Gaussian proposal distribution in the real case.

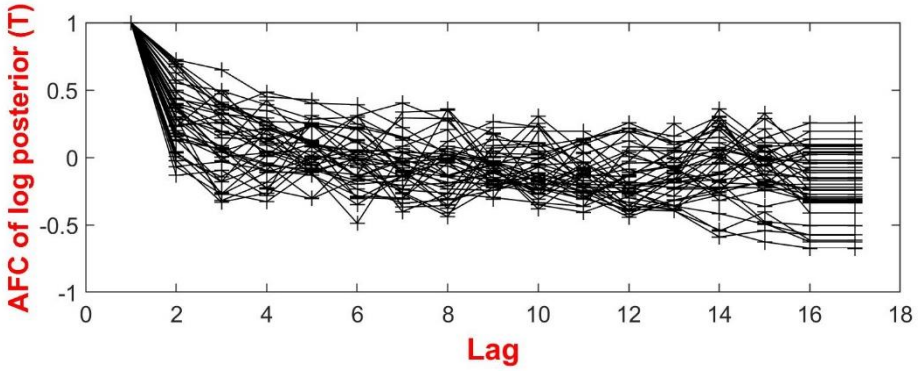


Figure 14. The autocorrelation plot of different lags for the posterior density derived from the PCA proposal distribution in the real case.

The autocorrelation plots of the log-posteriors in terms of different lags of the samples generated by the Gaussian proposal distribution and the presented algorithm are shown in Figures 13 and 14, respectively. From the autocorrelation results, there is higher autocorrelation in the samples obtained from the conventional scheme compared to those of the proposed strategy. Despite the advantage of the proposed algorithm over the conventional method, it should also be noticed that the implementation of the proposed approach depends upon the availability of the linearized inversion of the corresponding inverse problem. In addition, it is required to consider a strategy to monitor the convergence of the proposed MCMC algorithm. It can be a future work to detect convergence; i.e. how long should we run an MCMC chain.

4. Conclusion

The Markov Chain Monte Carlo inversion is aimed at accurately exploring the posterior model. A significant reason of the Bayesian inversion success depends on the shape and size of the proposal distribution used to implement the MCMC algorithm. Choosing an appropriate proposal density for the Metropolis-Hastings acceptance step is non-trivial. As the rejection rate increases, the computational cost for accepting a proposal sample increases. To overcome this problem, we developed an adaptive Metropolis-Hastings based on perturbing the model parameters through an eigenvalue decomposition of the model covariance matrix using principal component analysis (PCA). The Bayesian inversion commences with a linearized covariance estimated from a deterministic inversion. The covariance matrix obtained from the deterministic inversion is used during the burn-in phase, and as sampling progresses the covariance matrix is adaptively replaced with an alternative computed based on averaging over successive models along the Markov chain. This approach applies perturbations in principal component space where rotated parameters are uncorrelated, with the eigenvectors providing the rotation matrix and the eigenvalues providing appropriate perturbation length scales. The proposed scheme was examined for the inversion of simulated Slug test to recover the hydrogeological parameters, i.e. transmissivity and specific yield as well as for inversion of synthetic and real geo-electrical sounding data based on multiple depth layers of the fixed boundary. The numerical results indicated that the PCA perturbation proposal scheme significantly improved the efficiency of the MCMC sampler over the conventional Gaussian perturbation in terms of the faster convergence of the chain and good mixing properties. In future work, we would provide an application of the proposed procedure to large-scale non-linear geophysical problems such as electrical resistivity tomography.

References

Andrieu, C., & Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing*, 18, 343–373.

Blatter, D., Key, K., Ray, A., Foley, N., Tulaczyk, S., & Auken, E. (2018). Trans-dimensional Bayesian inversion of airborne transient EM data from Taylor Glacier, Antarctica. *Geophysical Journal*

International., 214(3), 1919–1936.

Blatter, D., Ray, A., & Key, K. (2021). Two dimensional Bayesian inversion of magneto-telluric data using trans dimensional Gaussian processes. *Geophysical Journal International*, 226(1) 548–563.

Cui, T., Fox, C., & O'Sullivan, M.J. (2011). Bayesian calibration of a large scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. *Water Resource Research*, 47, 26.

Dettmer, J., & Dosso, S.E. (2012). Trans dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains. Journal of the Acoustical Society of America, 132(4), 2239.

Dosso, S.E, Holland, C.W., & Sambridge, M. (2012). Parallel tempering in strongly nonlinear geoacoustic inversion. *Journal of the Acoustical Society of America*, 132(5), 3030–40.

Ferris, J.G., & Knowles, D.B. (1963). The slug-injection test for estimating the coefficient of transmissibility of an aquifer. Methods of Determining Permeability, Transmissibility and Drawdown, pages 299–304. U.S. Geological Survey.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., & Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of American statistics Association*, 85, 972–985.

Gelfand, A., & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of American statistics Association*, 85, 398–409.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721-741.

Ghanati, R., & Müller-Petke, M. (2021). A homotopy continuation inversion of geoelectrical sounding data. *Journal of Applied Geophysics*, 191.

Tafaghod Khabaz, Z., & Ghanati, R. (2023). Investigation of Smooth and Block inversion Characteristics of Electrical Resistivity Data. *Scientific Quarterly Journal of Geosciences*, doi: 10.22071/gsj.2022.319630.1962.

Günther, T., & Müller-Petke, M. (2012). Hydraulic properties at the North Sea island of Borkum derived from joint inversion of magnetic resonance and electrical resistivity soundings. *Hydrology and Earth System Sciences*, 16, 3279–3291.

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm, *Bernoulli*, 7, 223-242.

Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16, 339-354.

Hastings, H. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97-109.

Higdon, D., Lee, H., & Bi, Z. (2002). A Bayesian approach To Characterizing uncertainty in inverse problems using coarse and fine-scale information. *IEEE Transactions on Signal Processing*, 50(2), 389–399.

Koefoed, O. (1979). Geosounding Principles, Elsevier, Amsterdam.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., & Teller, A.H. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1092.

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of American Statistics Association*, 44, 335–341.

Parasnis, D.S. (1986). Principles of Applied Geophysics. Chapman and Hall, London.

Ray, A., Alumbaugh, D.L., Hoversten, G.M., & Key, K. (2013), Robust and accelerated Bayesian inversion of Marine controlled source electromagnetic data using parallel tempering. *Geophysics*, 78(6), E271–E280.

Sambridge, M., & Mosengaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Review Geophysics*, 40(3). 3-1-3-29.

Sambridge, M., Bodin, T., Gallagher, K., & Tkalcic, H. (2013). Trans-dimensional inference in the geosciences. *Philosophical Transactions of the Royal Society A*, 371, 20110547.

Swendsen, R.H., & Wang, J.S. (1987). Non

universal Critical dynamics in Monte Carlo simulations. *Physics Review Letter*, 58, 86–88.

Tarantola, A. (1987). Inverse Problem Theory: Methods for Data Fitting and Model Parameter estimation, Elsevier, Amsterdam.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM, 342.

Ter Braak, C.J.F. (2006). A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16, 239-249.

Vrugt, J.A., Ter Braak, C.F.F., Gupta, H.V., & Robinson, B.A. (2008). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling. *Stochastic Environmental Research and Risk Assessment*, 23(7), 1011–1026.

Appendix A.

Recurrence relationships for kernel functions to compute the surface electrical potential at any point in a layered medium.
Referring to Koefoed (1979) and Parasnis (1986), the electrical potentials in the first layer and the substratum read:

$$V_1 = \frac{I\rho_1}{2\pi r}\int_0^\infty \exp(-\lambda z)J_0(\lambda r)d\lambda + \int_0^\infty A_1(\lambda)[\exp(-\lambda z) + \exp(\lambda z)]J_0(\lambda r)d\lambda \quad \text{(A-1)}$$

$$V_n = \int_0^\infty A_n(\lambda)[\exp(-\lambda z)]J_0(\lambda r)d\lambda \quad \text{(A-2)}$$

And the electrical potential in any layer $i(i \neq 1, or\ n)$ is:

$$V_i(r,z) = \int_0^\infty [A_i(\lambda)\exp(-\lambda z) + B_i(\lambda)\exp(\lambda z)]J_0(\lambda r)d\lambda \quad \text{(A-3)}$$

where J_0 is the Bessel function of the first kind of order-zero and z shows the positive downward. $A_1, \dots, A_n$ and $B_2, \dots, B_{n-1}$ are unknown functions of the earth parameter and the real number λ. Conventionally, one is only interested in finding the potential at the surface of the earth, so only the coefficient A_1 needs to be found. The recurrence formulae for A_1 was given by Koefoed (1979) and Parasnis (1986). However, we need the potential at any point in a layered earth. Thus, we have to find all coefficients $A_1, \dots, A_n$ and $B_2, \dots, B_{n-1}$. They can be determined by solving the system of $2(n-1)$ linear equations obtained from the continuity conditions of the potential and the normal current density at the layer interfaces $H_i(\ i = 1, \dots, n-1)$. The solution is straightforward but tedious. The solutions are:

$$A_1 = \frac{I\rho_1}{2\pi}\exp(-2\lambda h_1)\frac{P_{1,2}}{1-P_{1,2}\exp(-2\lambda h_1)} \quad \text{(A-4)}$$

$$A_2 = \frac{I\rho_1}{2\pi}\frac{1+P_{1,2}}{1-P_{1,2}\exp(-2\lambda h_1)}\frac{1}{1+P_{2,3}\exp(-2\lambda h_2)} \quad \text{(A-5)}$$

$$B_2 = P_{2,3}\exp(-2\lambda H_2)A_2 \quad \text{(A-6)}$$

$$A_n = (1 + P_{n-1,n})A_{n-1} \quad \text{(A-7)}$$

$A_3, \dots, A_{n-1}$and $B_3, \dots, B_{n-1}$ are given by

$$A_i = \frac{1+P_{i-i,i}}{1+P_{i,i+1}\exp(-2\lambda h_i)}A_{i-1} \quad \text{(A-8)}$$

$$B_i = P_{i,i+1}\exp(-2\lambda H_i)A_i \ \ i = 3, \dots, n-1 \quad \text{(A-9)}$$

where

$$P_{i,i+1} = \frac{W^{(i+1)}-\rho_i}{\rho_i+W^{(i+1)}} \quad \text{(A-10)}$$

$$W^i = \rho_i\frac{W^{(i+1)}+\rho_i\tanh(\lambda h_i)}{\rho_i+W^{(i+1)}\tanh(\lambda h_i)} \qquad W^n = \rho_n \qquad for\ i = 1, \dots, n-1 \quad \text{(A-11)}$$

where an n-layer medium is discretized into the resistivities $\rho_1, \rho_2, \dots, \rho_n$ and the layer thicknesses $h_1, h_2, \dots, h_{n-1}$. H_i is the depth of the bottom of the ith layer and assumes the nth layer to extend to infinity, i.e. $h_n = \infty$ and $H_n = \infty$.