

Bayesian approach to magnetotelluric tensor decomposition

Václav Červ^{1,*}, Josef Pek¹, Michel Menvielle²

¹ Institute of Geophysics, Acad. Sci. Czech Rep., Prague, Czech Republic

² CNRS, Centre d'Études des Environnements Terrestre et Planétaires, Saint Maur des Fossés, France

Article history

Received July 7, 2009; accepted January 28, 2010.

Subject classification:

Magnetotelluric tensor, Decomposition, Bayesian approach, MCMC method, Gibbs sampler, Metropolis sampler.

ABSTRACT

Magnetotelluric directional analysis and impedance tensor decomposition are basic tools to validate a local/regional composite electrical model of the underlying structure. Bayesian stochastic methods approach the problem of the parameter estimation and their uncertainty characterization in a fully probabilistic fashion, through the use of posterior model probabilities. We use the standard Groom-Bailey 3-D local/2-D regional composite model in our Bayesian approach. We assume that the experimental impedance estimates are contaminated with the Gaussian noise and define the likelihood of a particular composite model with respect to the observed data. We use non-informative, flat priors over physically reasonable intervals for the standard Groom-Bailey decomposition parameters. We apply two numerical methods, the Markov chain Monte Carlo procedure based on the Gibbs sampler and a single-component adaptive Metropolis algorithm. From the posterior samples, we characterize the estimates and uncertainties of the individual decomposition parameters by using the respective marginal posterior probabilities. We conclude that the stochastic scheme performs reliably for a variety of models, including the multisite and multifrequency case with up to several hundreds of parameters. Though the Monte Carlo samplers are computationally very intensive, the adaptive Metropolis algorithm increase the speed of the simulations for large-scale problems.

1. Introduction

Magnetotelluric (MT) directional analysis and impedance tensor decomposition have since long become standard MT data analysis techniques that have largely extended possibilities of the MT interpretation of data with evidently a 3-D character [e. g., Zhang et al. 1987, Bahr 1988, Groom and Bailey 1989, Bahr 1991, Smith 1995, Jones and Groom 1993, Groom et al. 1993, Lilley 1998a, Lilley 1998b, McNiece and Jones 2001]. MT composite models reflect well the natural conditions in which the main distortions to the MT data often come from very complex near-surface inhomogeneities, while the deeper structure shows smoother conductivity trends and often a higher degree of symmetry. Since shallow inhomogeneities distort the MT impedances in solely a static way starting from a certain period, a possibility exists to separate the static and inductive parts of the impedance

tensor by utilizing their different frequency dynamics.

Various schemes of the MT decomposition have been suggested, and those by Bahr [1991] and Groom and Bailey [1989] have become standards in this respect. They are both primarily single site and single frequency approaches that may often produce largely scattered estimates of the decomposition parameters if noisy data and strong distortions are involved. Then, in order to obtain sufficiently stable decomposition results, an iterative procedure must be applied based on successively correcting the composite model parameters with respect to the data considered for a series of neighbouring periods. In general, superior estimates of distortion and regional MT parameters are obtained by fitting, in a least-squares sense, a parametric physical model of distortion to the observed impedance tensor, and the fit of the model tested statistically [McNeice and Jones 2001], as has been done in many studies since the beginnings of the decomposition approach [e.g., Bailey and Groom 1987, Groom and Bailey 1989, Groom and Bailey 1991, Groom et al. 1993]. Recently, McNeice and Jones [2001] have published a practical linearized multi-site and multi-frequency inverse procedure which stabilizes the MT decomposition by least-squares optimizing the composite model jointly for a series of periods and a whole array of sites.

Static distortions frequently cause the deep regional structure to become largely blurred in the surface MT data. As the regional parameters are of primary interest for the interpretation, a proper characterization of the uncertainties of their estimates is of essential importance. Bayesian inference is a stochastic approach frequently used in similar situations. In the Bayesian approach, prior information about the decomposition parameters is updated by combining it with the experimental impedance observations. The updated posterior probability distribution of the parameters conditioned on the experimental data then presents the most complete description of the parameter space for the

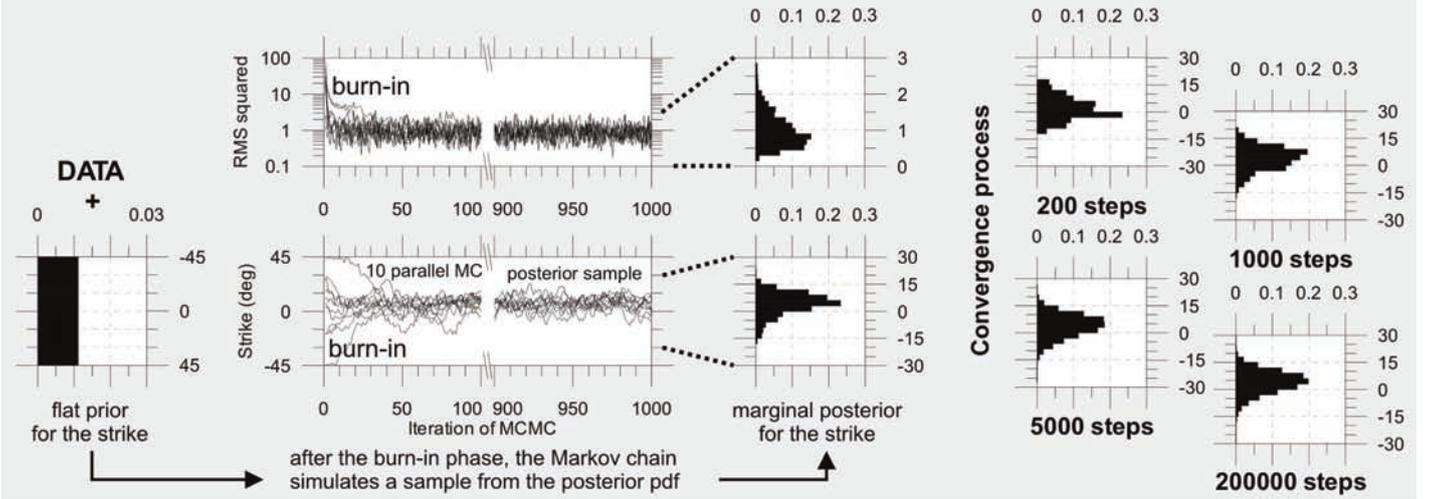


Figure 1. MCMC sampling procedure illustrated on the case of a stochastic estimation of the regional strike for one realization of the impedance tensor from Subsection 4.1. In the left, the data are combined with a non-informative, flat prior for the strike. Ten Markov chains, with seven decomposition parameters in (7), are run in parallel starting at different points in the parameter space. The normalized misfit and the strike are shown in the top and bottom evolution panels, respectively, for 10000 iterations of the Gibbs sampler. After about 100 iterations (burn-in period), the chains stabilize around a common strike value. After 10000 iterations, the histogram for the strikes may be considered a good approximation of the marginal probability density function of the strike. The histograms to the right show that the form of the marginal distribution of the strike is practically stable after the first 1000 iteration steps.

decomposition problem. Though the bayesian approach is sometimes criticized as a procedure that may inject subjective views into the inference via the choice of the prior probability, this is usually not an issue if non-informative flat priors are used for the parameters, when the inference results can be often rigorously shown to be close to those provided by classical statistics [e.g., Gelman et al. 2004].

The outstanding feature of the bayesian techniques is that they explicitly operate with probability distributions related to the composite model under study, and are thus capable of providing the most exhaustive quantitative information on the model parameter space [e. g., Gelman et al. 2004]. Clearly, this exhaustive probabilistic mapping of the model parameter domain needs an extensive exploration of the parameter space, which is often a computationally extremely intensive task.

In this contribution, we present simple Monte Carlo (MC) procedures to analyze the distorted MT data and conclude on both the point estimates of the decomposition parameters and their uncertainties by simulating marginal posterior probability density functions for the parameters. The structure of the paper is as follows: in Section 2, we briefly recall the basics of the MT distortions, MT composite models and decomposition procedures. In Section 3, we present a bayesian formulation of the MT decomposition problem for the classical Groom-Bailey factorization of a 3-D local/2-D regional composite model and summarize the main ideas on the numerical sampling procedures used, specifically the Markov chain Monte Carlo (MCMC) method with Gibbs sampler [Geman and Geman 1984] and an adaptive single-component Metropolis algorithm adopted from [Haario et al. 2004]. In the subsequent Section 4, we apply our stochastic decomposition procedure to the

synthetic as well as practical MT data sets presented by McNiece and Jones [2001] in their recent multi-site, multi-frequency decomposition study, and discuss the performance and efficiency of the stochastic approach for those data sets. Finally, we outline some perspectives of the stochastic decomposition in the conclusion, Section 5.

2. MT tensor decomposition

2.1. 3-D local/2-D regional composite model

Static distortions of the MT impedance tensor due to shallow 3-D electrical inhomogeneities can be formally described by

$$\mathbf{Z}^{obs}(r, T) = \mathbf{A}^{dist}(r) \mathbf{Z}^{reg}(r, T), \quad (1)$$

where $\mathbf{A}^{dist}(r)$ is a frequency independent distortion tensor, and $\mathbf{Z}^{obs}(r, T)$ and $\mathbf{Z}^{reg}(r, T)$ are, respectively, the observed and regional impedance tensors at the location r and period T . For a 2-D regional structure, the regional impedance in the direction of the regional strike is given by an antidiagonal tensor,

$$\mathbf{Z}^{reg2D}(r, T) = \begin{bmatrix} 0 & \mathbf{Z}_E(r, T) \\ -\mathbf{Z}_H(r, T) & 0 \end{bmatrix}. \quad (2)$$

Two factorizations of the distortion tensor in terms of more elementary distortion factors are widely used. Bahr's approach expresses the distortion matrix in terms of telluric deviations and anisotropic gains [Bahr 1991],

$$\mathbf{A}^{dist}(r) = \begin{bmatrix} 1 & \tan \beta_H \\ \tan \beta_E & 1 \end{bmatrix} \begin{bmatrix} a_E & 0 \\ 0 & a_H \end{bmatrix}, \quad (3)$$

while that by Groom and Bailey [1989] factorizes it as a product

of elementary distortion types, twist t , shear e , anisotropy s , and gain g ,

$$A^{dist}(r) = \begin{bmatrix} 1 & -t \\ t & 1 \end{bmatrix} \begin{bmatrix} 1 & e \\ e & 1 \end{bmatrix} \begin{bmatrix} 1+s & 0 \\ 0 & 1-s \end{bmatrix} g. \quad (4)$$

By comparing (3) and (4), a simple relation between the directional distortion parameters of those two factor sets can be written immediately,

$$\begin{aligned} \beta_E &= \varepsilon + \tau, & \beta_H &= \varepsilon - \tau, \\ \varepsilon &= \frac{1}{2}(\beta_E + \beta_H), & \tau &= \frac{1}{2}(\beta_E - \beta_H) \end{aligned} \quad (5)$$

with $\varepsilon = \arctan e$, $\tau = \arctan t$.

2.2. Fitting the composite model

MT decomposition is an ambiguous task. From a single-site single-frequency impedance tensor, we can uniquely recover only the regional strike θ_{reg2D} , the directional distortions twist t and shear e , and scaled (by unknown real factors) regional impedances, $Z_E^{scaled}(r, T)$, $Z_H^{scaled}(r, T)$.

MT decomposition leads to the solution of a system of eight real, non-linear algebraic equations that result from the condition of fitting the experimental impedances to those produced by the composite model,

$$\begin{aligned} Z^{obs.mod}(r, T) &= R(\theta_{obs} - \theta_{reg2D}) \begin{bmatrix} 1 - te & e - t \\ e + t & 1 + et \end{bmatrix} \times \\ &\times \begin{bmatrix} 0 & Z_E^{scaled}(r, T) \\ -Z_H^{scaled}(r, T) & 0 \end{bmatrix} R(\theta_{reg2D} - \theta_{obs}), \end{aligned} \quad (6)$$

where $R(\theta)$ is a 2-D rotation matrix through θ .

As the near-surface distortion often mask the deep structure to a considerable degree, estimates of regional parameters, and especially those of the regional strike θ_{reg2D} , are often unstable and largely scattered if they are evaluated separately for individual frequencies within some frequency range. In the procedure by McNeice and Jones [2001] the observed impedance data are fitted by a composite model for a whole range of periods and for multiple sites simultaneously. Specifically, the multi-site multi-frequency decomposition minimizes the target

$$\Phi(d_{ij}) = \sum_i \sum_j \sum_{\alpha, \beta \in \{x, y\}} \sum_{Pa \in \{Re, Im\}} \sum_{(sites) (periods)} \left[\frac{Pa Z_{\alpha\beta}^{obs.exp}(r_i, T_j) - Pa Z_{\alpha\beta}^{obs.mod}(d, r_i, T_j)}{\delta Z_{\alpha\beta}^{obs.exp}(r_i, T_j)} \right]^2, \quad (7)$$

where d are the decomposition parameters, i. e., the regional strike common to all sites and periods, the twist and shear parameters common to all periods at a specific site, and the regional impedance pairs specific for each period and each site. In what follows, we use the normalized value of $\Phi(d)/N_D$ to characterize the misfit between the observed and model data, where N_D is the total number of the data items. If experimental (complex) impedances are available for N_T periods at each of N_S sites, then $N_D = 8 N_T N_S$. The total number of decomposition parameters to be recovered by a decomposition procedure is $N_p = 4 N_T N_S + N_S + 1$, where the

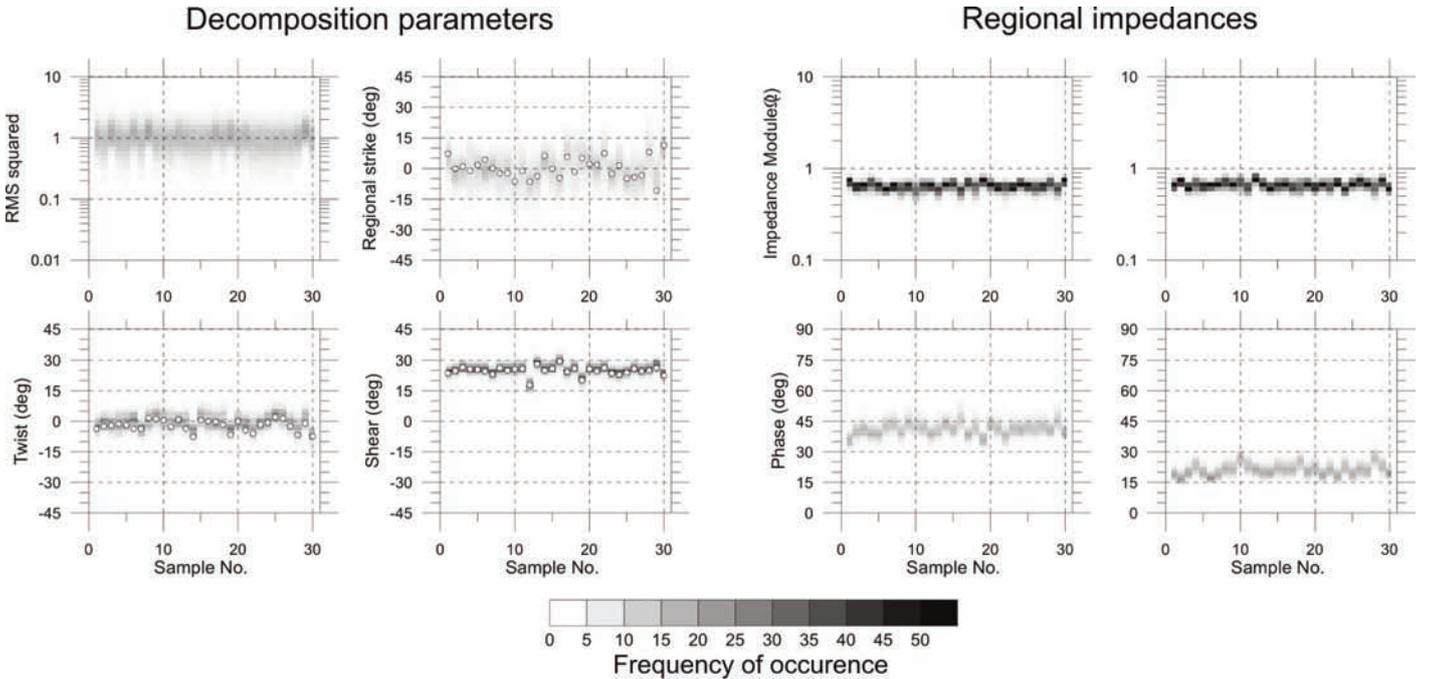


Figure 2. MT decomposition parameters from the MCMC sampling applied individually to each of 30 noisy realizations of the impedance tensor derived from (17). Left panels show the normalized misfit (RMS squared) for each decomposition run, and histograms, transformed into gray-shade maps, for the regional strike, twist and shear parameters. For comparison, point estimates of the latter three parameters from Bahr's decomposition [Bahr 1988] are shown by white circles. Right panels show the approximate marginal probability densities for the recovered regional impedances, in terms of their modules and phases. Exact values of the parameters from the generating impedance tensor (14) are: $\nu_{reg2D} = 0^\circ$, $\arctan t = -2.2^\circ$, $\arctan e = 24.95^\circ$, $\varphi_E = 40.630$, $\varphi_H = 20.59^\circ$.

three summands are for the number of the (real) scaled regional impedances, number of twists and shears, and one common value of the regional strike, respectively.

The problem of minimizing equation (7) with respect to the parameters d is a standard non-linear optimization problem. McNeice and Jones [2001] use an efficient iterative procedure based on a sequential quadratic programming algorithm to minimize the difference between the observed and model impedances and to obtain point estimates of the decomposition parameters in the least-squares sense. To quantitatively characterize the parameter uncertainties recovered from the non-linear minimization of (7), McNiece and Jones [2001] use a slightly modified bootstrap procedure of Groom and Bailey [1991] to derive the confidence limits for the individual decomposition parameters.

3. Bayesian approach to the MT decomposition

3.1. Bayesian formulation of the MT decomposition problem

In a bayesian approach, both the parameter estimation and the assessment of the parameter uncertainties are treated as problems of determining the posterior probability of the composite model conditioned on the observed data, i. e., according to the Bayes rule,

$$\text{Prob}(d | Z^{obs.exp}, M) \propto \frac{\text{Prob}(Z^{obs.exp} | d, M) \times \text{Prob}(d | M)}{\text{Prob}(Z^{obs.exp}, M)}, \quad (8)$$

[see, e. g., Gelman et al. 2004]. The posterior probability density function $\text{Prob}(d | Z^{obs.exp}, M)$ is considered a solution to the inverse problem (7) and is further used to evaluate point

estimates for the parameters and to derive their confidence intervals.

In the general formula (8), the prior probability, $\text{Prob}(d | M)$, describes the available knowledge about the decomposition parameters prior to the data being observed. The symbol M stands for the assumptions made on the decomposition model *a priori*. In our particular decomposition problem, M represents a class of 3-D local/2-D regional composite models (1) used throughout this paper.

As we generally do not assume any particular knowledge about the decomposition parameters a priori, we use flat (constant) priors on the individual parameters within reasonable physical bounds, specifically

$$\begin{aligned} \theta_0 \leq \theta_{reg} \leq \theta_0 + 90^\circ, \quad -2 \leq t(r_i) \leq 2, \quad -1 \leq e(r_i) \leq 1, \\ 0.5 \sqrt{10\rho_{min}/T_j} \leq \text{Pa } Z_{E,H}(r_i, T_j) \leq 0.5 \sqrt{10\rho_{max}/T_j}, \end{aligned} \quad (9)$$

where θ_0 is used to adjust limits of the regional strike range, and ρ_{min} , ρ_{max} define wide enough limits to accomodate sufficiently large impedance shifts. The lower and upper bounds for the twist and shear correspond to the limit angles of $\pm \arctan 63.4^\circ$ for the twist and $\pm \arctan 45^\circ$ for the shear. Of course, one of the strengths of the bayesian analysis is that more informative priors on the parameters can be introduced into (8) if additional structural information is available *a priori*.

The other fundamental term in (8), the likelihood, $\text{Prob}(Z^{obs.exp} | d, M)$, represents the probability of obtaining the observed impedances given a particular set of values for the decomposition parameters, and can be written in the form

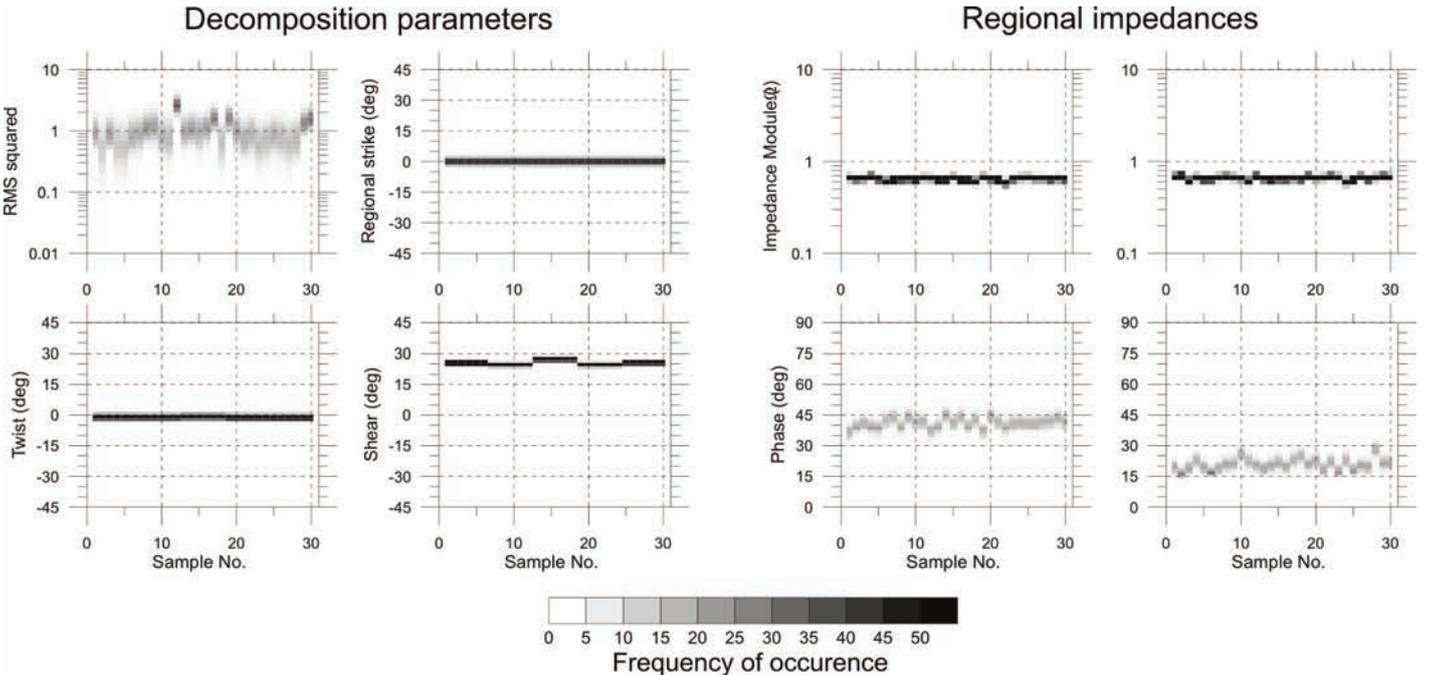


Figure 3. MT decomposition parameters form the stochastic sampling applied to five groups of six realizations the impedance tensor derived from (17). For each group, the twist and shear parameters are considered constant. The regional strike is assumed to be the same for all data. For details on the figure structure, see caption to Figure 2.

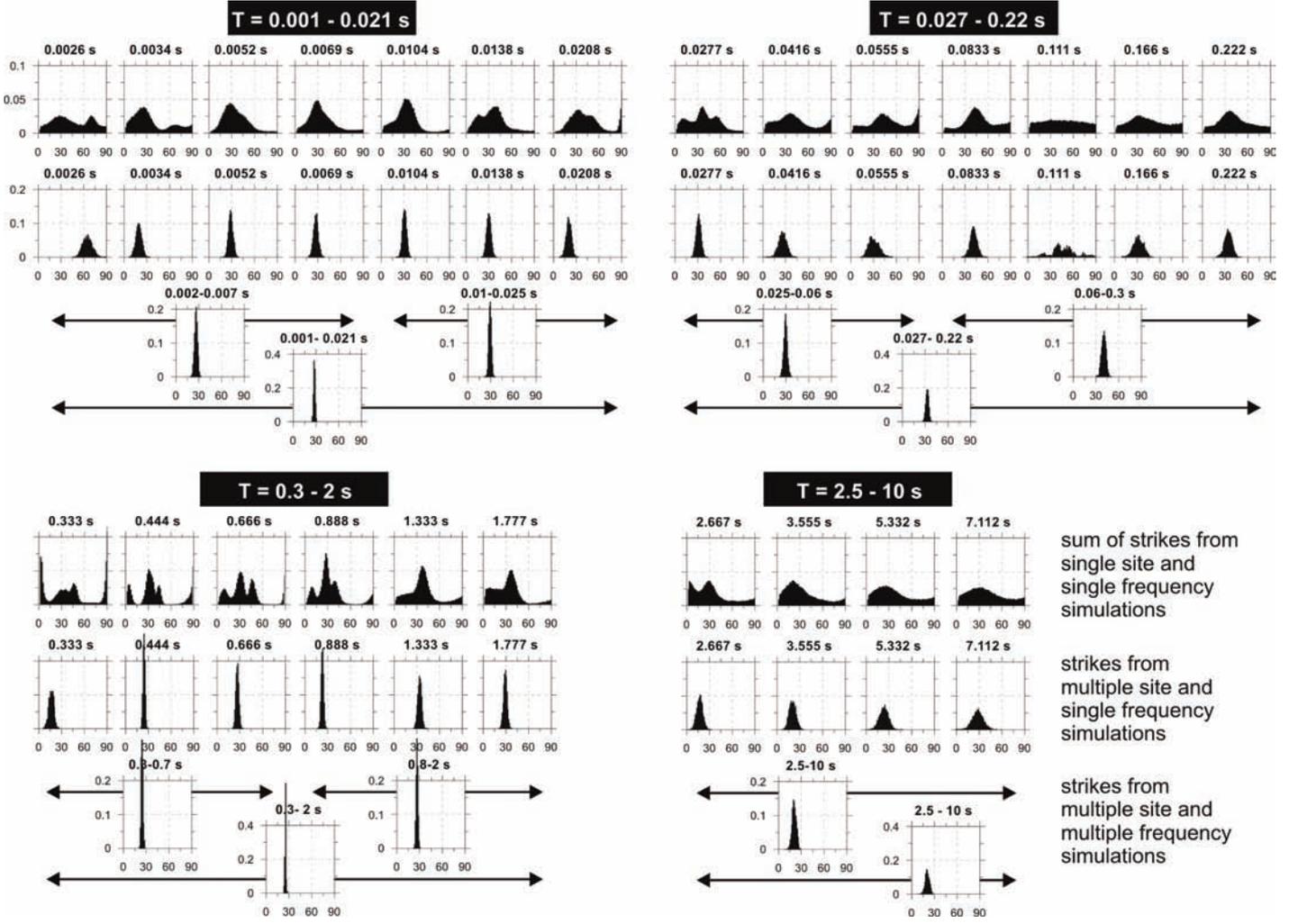


Figure 4. Marginal probability densities for the regional strike from the PNG101 through PNG108 data set [Jones and Schultz 1997]. Top row: Sum of strikes estimated for the sites and periods separately. Middle row: Strikes obtained from a multi-site, single frequency decomposition applied to all eight sites jointly. Bottom rows: Strikes from a multi-site, multi-frequency decomposition over all eight sites and over period bands of, respectively, half a decade and one decade width.

$$\text{Prob} (Z^{obs.exp} | d, M) \propto \exp \left[-\frac{1}{2} \Phi(d) \right], \quad (10)$$

if Gaussian noise distribution in the observed data is assumed. Here, $\Phi(d)$ is the misfit defined by (7). The likelihood function allows us to rate models according to their fit to the particular experimental data observed. By folding the likelihood with the prior information on the parameters, we arrive at the parameters' posterior probability distribution. The denominator in (8), $\text{Prob} (Z^{obs.exp}, M)$, plays just a role of a constant scaling factor which guarantees that the posterior probability distribution integrates to one over the admissible parameter space domain.

Clearly, the posterior probability (8) with the specific likelihood (10) combined with the uniform priors on the parameters (9) shifts the bayesian formulation very close to the standard Gaussian stochastic model. Classical maximum likelihood estimation (MLE) maximizes the likelihood function over the parameter space,

$$d_{MLE} = \arg \max_d \text{Prob} (Z^{obs.exp} | d, M). \quad (11)$$

The maximum a posteriori parameter estimate (MAP) from (8) is

$$\begin{aligned} d_{MAP} &= \arg \max_d \text{Prob} (Z^{obs.exp} | d, M) = \\ &= \arg \max_d \left[\text{Prob} (Z^{obs.exp} | d, M) \text{Prob} (d | M) \right] = d_{MLE} \end{aligned} \quad (12)$$

for a constant prior, $\text{Prob} (d | M) = \text{const} > 0$. To assess the parameter uncertainties, either a full stochastic sampling from the likelihood must be carried out, or, as most common in practice, an approximation of the likelihood by a multivariate normal distribution in the vicinity of the MAP solution can be used [Laplace approximation, e. g. MacKay 2003],

$$\begin{aligned} \text{Prob} (d | Z^{obs.exp}, M) &\propto \\ &\propto \exp \left\{ \ln \left[\text{Prob} (Z^{obs.exp} | d, M) \text{Prob} (d | M) \right] \right\} \approx \\ &\approx \exp \left\{ \ln \left[\text{Prob} (Z^{obs.exp} | d_{MAP}, M) \text{Prob} (d_{MAP} | M) \right] - \right. \\ &\quad \left. - \frac{1}{2} (d - d_{MAP})^T A (d - d_{MAP}) \right\} \sim \text{Normal} (d | d_{MAP}, A^{-1}), \end{aligned} \quad (13)$$

where $A_{ij} = -\delta^2 \ln \left[\text{Prob} (Z^{obs.exp} | d, M) \text{Prob} (d | M) \right] / \delta d_i \delta d_j$ evaluated at the point d_{MAP} . A full stochastic sampling may

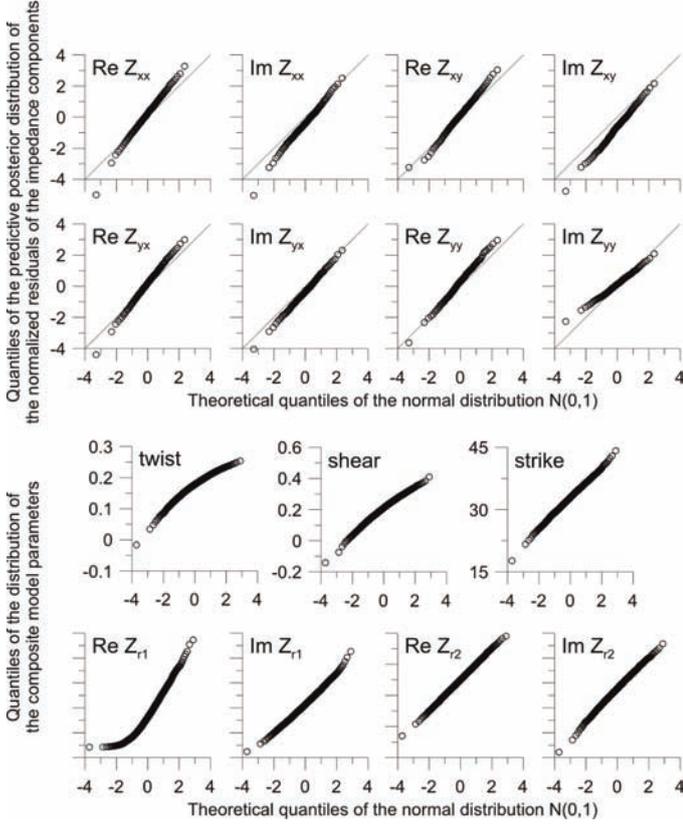


Figure 5. Fit of the data generated by the composite model and the experimental impedances for the site PNG101 from the MT-DIW2 Papua-New Guinea data set [Jones and Schultz 1997]. The decomposition was run for one decade of periods, 1 to 10 s (6 periods). Rows 1 and 2: Distribution of normalized residuals between the experimental impedance components and the corresponding data replicas given the simulated decomposition parameters compared with the standardized normal distribution by means of the q - q plots for one arbitrarily chosen period 3.56 s. Rows 3 and 4: q - q plots for selected parameters of the decomposition model from the Markov chain with respect to the respective best-fit Gaussian curves. The anomalous q - q plot for the real part of the first regional impedance is due to our setting the lower limit for that parameter too high by mistake.

then be useful in capturing deviations from this approximate normal distribution.

3.2. MT decomposition via stochastic sampling

Analytic solutions to bayesian inference problems are rare and are mostly limited to linear statistical models in low-dimensional settings and to simple standard probability distribution functions. Most of the practical applications of the bayesian methods, especially in higher dimensions and with non-linear models involved, are based either on qualified approximations of the target probability distributions by simpler standard probability densities, like Gaussians or their mixtures, or on generating samples from the posterior probability function numerically, e. g., by Monte Carlo simulation procedures.

In our study, we have used a version of the Monte Carlo method with Markov chains (MCMC) to simulate samples from the posterior probability of MT composite models conditioned on the observed impedances. Without going

into details of the MCMC technique [for details see, e. g., Gelfand et al. 2004 or, within a geoelectrical context, Grandis et al. 1999], the procedure consists (i) in constructing an ergodic Markov chain with the limit probability distribution equal to our target posterior probability (8), and (ii) in obtaining a partial realization from the corresponding Markov chain. After a certain burn-in period, during which the chain transits to its stationary state, the samples from the Markov chain realization can be considered approximate draws from our target posterior distribution (8).

MCMC sampling algorithms represent general rules for the construction and generation of the Markov chains with the above desired properties. Here, we have at first tested the standard Gibbs sampling procedure [e. g., Geman and Geman 1984, Grandis et al. 1999, Gelman et al. 2004], which proceeds as follows: starting from the latest state of the Markov chain, say k -th, with parameters $\vec{d}(k)$, the Gibbs sampler loops through all the components of the vector \vec{d} , and, for each individual component, d_i , updates its value by drawing from the univariate conditional probability density

$$d_i^{(k+1)} = \text{Draw} \left\{ \text{Prob} \left(d_i \mid \underbrace{d_1^{(k+1)}, \dots, d_{i-1}^{(k+1)}}_{\text{updated values}}, \underbrace{d_{i+1}^{(k)}, \dots, d_{N_p}^{(k)}}_{\text{original values}}, M \right) \right\} \quad (14)$$

After all the N_p components of the parameter vector have been updated in this way, one iteration step of the Gibbs sampler, and the transition to its new state $\vec{d}^{(k+1)}$, is completed.

The convergence of the MCMC procedure to the target probability is theoretically guaranteed for Markov chains of infinite length only. In practice, various indicators are used to assess the convergence, the simplest being the stability of the marginal probabilities of the model parameters over long enough sections of the chain. To assist the convergence, several parallel chains can be started from different points in the parameter space. Figure 1 illustrates the basic steps of the MCMC sampling procedure for the regional strike assessment from a synthetic impedance tensor realization treated later in Section 4.1.

After a sample from the posterior probability density function is obtained, basic bayesian integrals (mean values, covariance matrices, etc.) can be easily evaluated from the posterior sample and used to assess both the decomposition parameters (mean values) and their uncertainties and inter-dependencies (variance-covariance matrix, correlations).

Since the conditional probabilities in (14) are not given in a closed form, they are standardly approximated on a grid of points within the parameter domain. As the likelihood function (10) has to be evaluated at each grid point, that approximation may require extreme computing times, especially if the direct problem solution is demanding, or if vast domains of the parameter space with very low likelihood are sampled. In some of our practical experiments,

Site	True twist	True shear	MNJ twist	MNJ shear	ACM twist	ACM shear
SYN001	-20	20	-20.1	20.1	-20.0 ± 0.1	19.9 ± 0.1
SYN002	40	-10	40.2	-10.1	39.9 ± 0.1	-10.2 ± 0.1
SYN003	-15	25	-15.1	25.2	-15.2 ± 0.2	25.0 ± 0.1
SYN004	20	40	19.7	39.9	19.9 ± 0.2	40.0 ± 0.2
SYN005	-40	-25	-40.0	-25.0	-40.0 ± 0.1	-25.0 ± 0.1
SYN006	30	-20	30.1	-20.2	29.9 ± 0.1	-20.0 ± 0.1
SYN007	-50	-35	-50.1	-34.9	-50.1 ± 0.2	35.2 ± 0.1
SYN008	-10	25	-10.1	-25.1	-10.3 ± 0.2	25.0 ± 0.1
SYN009	-5	35	-5.3	35.1	-5.1 ± 0.2	35.1 ± 0.1
SYN0010	45	15	45.1	14.8	44.9 ± 0.1	15.0 ± 0.1

Table 1. Strike and twist parameters, in degrees, for the synthetic data produced by the 2-D conductive block in Section 4.2. True parameters were used to distort 2-D impedances of the box model. MNJ are results of the reverse decomposition presented by McNeice and Jones [2001]. ACM parameters are results of the adaptive componentwise Metropolis sampling used in this paper. For the common regional strike, we have 30° for the true strike, 30.3° from the MNJ analysis, and $30.1 \pm 0.2^\circ$ from the stochastic procedure.

we have used grid steps as small as 0.5° for the regional strike, 0.01 for the twist and shear parameters, and 0.005 for the logarithms of the components of the regional impedances. Considering the parameter limits specified in (9), with $\rho_{\min} = 10^{-2} \Omega m$ and $\rho_{\max} = 10^5 \Omega m$ used, the number of misfit (7) evaluations totals to more than 25 millions in such cases for one iteration step of the Gibbs sampler. Coarsening the grid over areas with small likelihood, by using more sophisticated, data driven approximation schemes, such as the neighbourhood interpolation suggested by Sambridge [1997], or by narrowing the parameter bounds helps in reducing the computation burden.

As an alternative to the Gibbs sampler, we have also tested a slightly simplified version of the componentwise adaptive Metropolis algorithm suggested recently by Haario et al. [2003] in the context of upper atmosphere studies. The algorithm proceeds in similar cycles as the Gibbs sampler above except that the updates to the individual component d_i are generated by an adaptive Metropolis rule. For this, first a proposal draw is made from a normal distribution centered at the current value $d_i^{(k)}$ with a data adaptive variance, specifically

$$\text{proposal}_i^{(k)} = \text{Draw} \left\{ \text{Normal} \left[d_i^{(k)}, s(\text{var}_i^{(k)} + \varepsilon) \right] \right\} \quad (15)$$

where $\text{var}_i^{(k)}$ is the variance of d_i estimated from the previous steps of the sampler, the factor s is a multiplicative constant tuned experimentally to optimize the rejection/acceptance ratio of the algorithm (here, $s = 2.4$ has been used according to the suggestion by Haario et al. 2003), and ε is a small regularizing factor. Then, the Metropolis decision step is made, i. e., the candidate point is accepted, $d_i^{(k+1)} = \text{proposal}_i^{(k)}$ with the probability

$$\pi(\text{accept}) = \min \left\{ 1, \frac{\text{Prob}(\text{proposal}_i^{(k)} | d_1^{(k+1)}, \dots, d_{i-1}^{(k+1)}, d_{i+1}^{(k)}, \dots, d_{N_p}^{(k)})}{\text{Prob}(d_i^{(k)} | d_1^{(k+1)}, \dots, d_{i-1}^{(k+1)}, d_{i+1}^{(k)}, \dots, d_{N_p}^{(k)})} \right\}. \quad (16)$$

If the proposal is rejected, the old value of the parameter is retained, i.e., $d_i^{(k+1)} = d_i^{(k)}$ with the probability $1 - \pi(\text{accept})$. As in this algorithm a longer history of the chain is employed for adapting the variances in (15), the chain is evidently not Markovian any more. Nonetheless, Haario et al. [2003] have proved its convergence to the target posterior. As compared to the original Gibbs sampler, the adaptive Metropolis procedure requires only one solution to the direct problem per component and per iteration. The adaptive variance in (15) should take care of a quasi-optimality of the acceptance/rejection ratio for updating the model parameters in the chain evolution, regulating thus the convergence of the chain.

4. Numerical experiments

4.1. Synthetic example I:

multiple realizations of a single impedance tensor

To have a possibility to compare results of our numerical experiments with a reliable reference, we have tested the stochastic MT decomposition procedure on several data sets presented lately by McNiece and Jones [2001] in their multi-site, multi-frequency decomposition study. Their first example uses a single synthetic impedance matrix

$$Z = \underbrace{\begin{pmatrix} 1.26 & 0.44 \\ 0.53 & 0.86 \end{pmatrix}}_{\text{distortion}} \underbrace{\begin{pmatrix} 0 & 4.72 + 4.05i \\ -8.25 - 3.10i & 0 \end{pmatrix}}_{\text{2-D impedance}} \times 10^{-4} \Omega \quad (17)$$

They further generate a series of 31 realizations of this distorted impedance tensor by contaminating its components by Gaussian noise with the standard deviation of 4.5 % of the largest impedance element. Though artificial, the example is suitable for estimating the impact of the noise on the decomposition results under otherwise equivalent conditions as regards the regional structure as well as the local distorter.

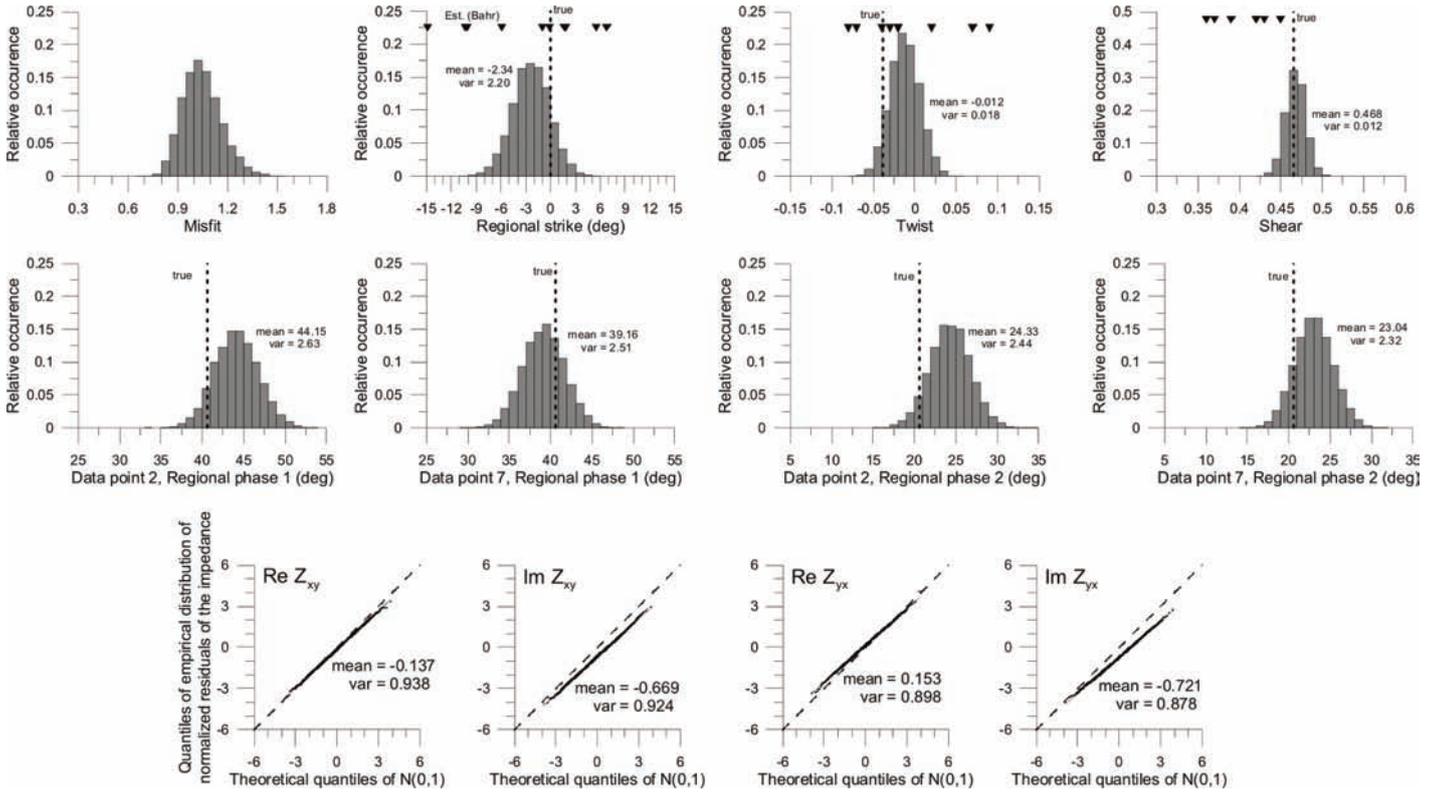


Figure 6. Histograms for selected decomposition parameters generated after 100k MCMC steps for 10 experimental impedances contaminated with Gaussian noise. Top row: Histograms of the misfit, regional strike, twist and shear along with point estimates of the decomposition parameters by Bahr's [1988] formulas (black triangles) and the true parameter values (dashed vertical line). Middle row: Histograms of the regional phases 1 and 2 (E and H) at two arbitrary selected data points, 02 and 07 from the set of 10 experimental impedances. Bottom row: q - q plots of normalized residuals of the principal impedance elements versus the standardized normal distribution for one arbitrary data point, here point 02 from the 10 experimental impedances.

In Figures 2 and 3, we show results of the stochastic decomposition runs for the set of thirty data realizations, which were combined in various ways to simulate different decomposition modi. Specifically, Figure 2 displays results of the decomposition applied to each of the thirty tensors individually. We applied the Gibbs sampler with 10k steps. The first 2k samples were used to equilibrate the chain (burn-in period). Histograms of the individual parameters from the remaining 8k samples were used as approximates to their marginal probability densities. The relative frequencies of occurrence of particular values were mapped onto a gray-shade scale, and are shown in Figure 2. For comparison, point estimates of the regional strike, twist and shear obtained by Bahr's decomposition analysis [Bahr 1988] are also indicated.

While the previous example illustrated the stochastic decomposition approach in a setting typical for a single site and single frequency decomposition, Figure 3 illustrates results of a simulation for a multiple site and multiple frequency case from the same data set. Now, the noisy impedance tensors are arranged into five groups, with six impedance realizations in each of them. Each group simulates a site, with a common value of the twist and shear parameters. Each realization within a particular group simulates one frequency. The regional strike is assumed to have the same value common to all the data.

4.2. Synthetic example II:

multi-site multi-frequency decomposition over a 2-D block model

The second example adopted from McNeice and Jones [2001] is their synthetic study of distorted impedances generated by a simple 2-D model. A 50 Ωm 2-D body, 5 km deep with a 4 km depth extent and a width of 25 km, was embedded in a 1000 Ωm half-space. Observations were made at ten sites equispaced at 8 km intervals across the surface. At each site, impedances were modelled numerically for 31 periods within the range 0.01 to 1000 s and subsequently distorted by a predefined, site-specific distortion matrix. The distorted tensors were then rotated away from the regional strike direction (by -30°), and contaminated with Gaussian noise with the standard deviation corresponding to 2 % of the maximum impedance element at each period.

Though apparently simple and straightforward, this example bears some specific features that might be rather unfavourable for the stochastic decomposition approach. First, stochastic global optimization and sampling procedures are known to fail frequently for problems with a large number of variables. In the model above, the total number of decomposition parameters to be resolved is 1261, which may be considered very large for a stochastic approach. Especially, this number of variables practically prevents us from using the simple Gibbs sampler within the MCMC, as the number of solutions to the direct problem

would be prohibitively large if no coarsening strategy for the approximation of the parameters' conditionals could be suggested.

Second, any componentwise sampling suffers from the presence of highly correlated parameters within the set of variables. In the present synthetic example, the 2-D manifestation of the regional structure is relatively weak, and is, moreover, obscured by excessively strong artificial distortions. In such a case, correlations between the parameters of the composite model occur, with degraded performance of the sampling procedure as a consequence. We applied the componentwise adaptive Metropolis algorithm by Haario et al. [2003] to perform the MCMC sampling for the above model. We have met serious difficulties in sampling for the sites individually, especially because of strong correlations between the decomposition parameters. For the whole set consisting of all 10 sites and all 31 frequencies, the procedure behaved much more regularly than for the sites considered individually, and produced satisfactory results after about 100k iterations of the adaptive Metropolis algorithm. For comparison, we present our results for the strike, twist and shear estimates together with those published by McNiece and Jones [2001] in Table 1.

4.3. Practical example:

MT-DIW2 Papua-New Guinea MT data set

Here we present a few illustrative results of the analysis of the Papua New Guinea data [PNG, Jones and Schultz 1997], which were a subject of detailed investigations within the MT-DIW2 project, and have been studied extensively by McNiece and Jones [2001] from the point of view of the directional and decomposition analysis. From the latter analysis, this data set has been shown to indicate a consistent regional structure for a series of eight sites, called PNG101 through PNG108.

To compare the performance of our algorithm with the optimization procedure by McNiece and Jones [2001] for a set of field data, we have used the PNG data set within our stochastic decomposition analysis. Here, we will only show a fraction of the results concerning the regional strike estimates. The results were obtained by using the Gibbs sampler within the MCMC procedure, typically with 10k iterations and 2k steps of a burn-in phase. The results are summarized in Figure 4, and can be compared directly with the estimates given in McNiece and Jones [2001], Figures 11 through 13.

The strike estimates were obtained by applying the stochastic decomposition to various combinations of the PNG data items. First, the single strike, single frequency decomposition was carried out. The partial strike histograms at individual frequencies were then merged into a single histogram to show the aggregate directional information for the region at individual frequencies (top row of histograms

in Figure 4). Obviously, for whole frequency ranges, this directional information is rather poor and excessively diffuse.

By assuming the same regional strike for all the eight sites considered at individual frequencies, a multiple site, single frequency directional analysis clearly improves the resolution with respect to the regional strike (middle row of histograms in Figure 4). Further sharpening of the deep directional image is achieved by aggregating the data over frequency ranges, as demonstrated by histograms in the bottom line in Figure 4.

4.4. Assessing MCMC model fit and stability

In addition to verifying the convergence of the Markov chain by comparing multiple chains started from different points in the parameter space, as described in Section 3.2, the final model has to be also assessed according to its fit to the original data. As the Monte Carlo procedure provides a sample from the posterior probability (8), the bayesian goodness-of-fit is estimated with respect to this probability sample rather than with respect to specific point estimates of the model parameters. A common way is to compare the experimental data with their replicas produced by the model via a predictive posterior probability.

Probability of hypothetical data replications $Z^{obs.rep}$ conditioned on the really observed experimental data $Z^{obs.exp}$ is given by [Gelman and Meng, chap. 11 in Gilks et al. 1999]

$$\begin{aligned} \text{Prob} (Z^{obs.rep} | Z^{obs.exp}, M) &= \\ &= \int_D \text{Prob} (Z^{obs.rep} | d, M) \times \text{Prob} (d | Z^{obs.exp}, M) dD \approx \\ &\approx \sum_k \text{Prob} (Z^{obs.rep} | d^{(k)}, M) \times \text{Prob} (d^{(k)} | Z^{obs.exp}, M), \end{aligned} \quad (18)$$

where the integration goes over all the parameter space D , and can be approximated by a summation over the states k of the posterior density sample providing the Markov chain has sampled the influential part of the parameter space sufficiently well. The first factor in the above integrand/sum is a probability of $Z^{obs.rep}$ if the model parameters are $d^{(k)}$ and the observation conditions (noise sources) are exactly the same as in the true experiment. Then, this probability can be approximated by sampling for $Z^{obs.rep}$ from the likelihood function (10) with the fixed model parameters $d^{(k)}$, i. e., by drawing samples from a Gaussian distribution centered at $Z(d^{(k)})$ and with the covariance matrix of the experimental data. The second factor in the integrand/sum (18) is the probability of the state $d^{(k)}$ from the posterior distribution (8), and can be taken directly from the Markov chain.

We present an example of the model fit in Figure 5. The decomposition was run for one decade of periods, 1 to 10 s (6 periods), for a single site PNG101 from the MT-DIW2 Papua-New Guinea data set [Jones and Schultz 1997]. Distribution of normalized residuals between the experimental impedances and the data replicas given the

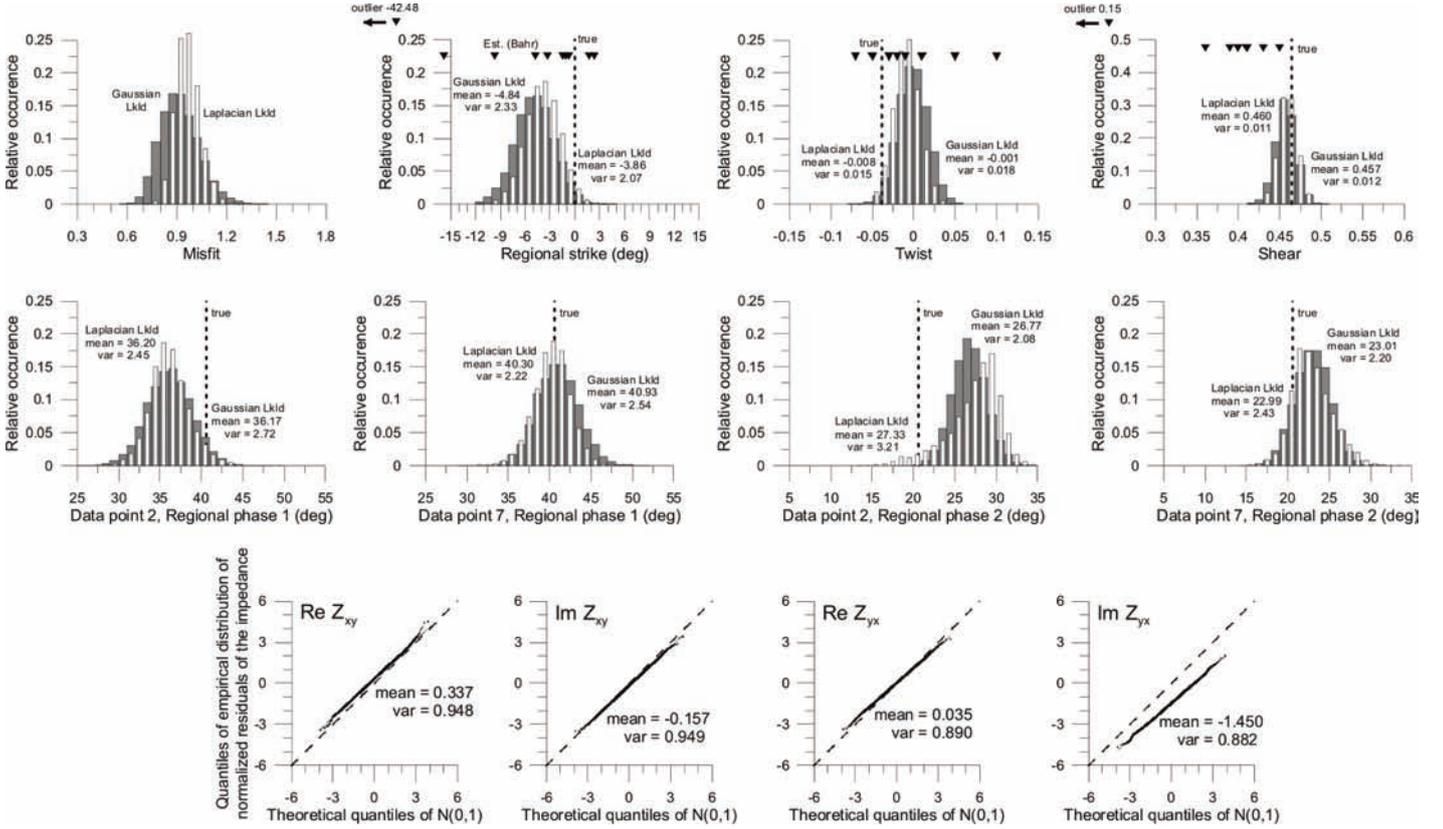


Figure 7. Histograms for selected decomposition parameters generated after $100k$ MCMC steps for 10 experimental impedances contaminated with Laplacian noise with the same variance as in the Gaussian case in Figure 6. Top row: Histograms of the misfit, regional strike, twist and shear generated by the MCMC procedure with the Gaussian (gray histograms) and Laplacian (white histograms) likelihood used in the MCMC procedure. Point estimates of the decomposition parameters from Bahr's [1988] formulas are shown by black triangles and the true parameter values indicated by dashed vertical lines. Middle row: Histograms of the regional phases 1 and 2 (E and H) at two arbitrary selected data points, 02 and 07 from the set of 10 experimental impedances. Point 02 shows the large outlier indicated in the top strike and shear histograms. Bottom row: q - q plots of normalized residuals of the principal impedance elements versus the standardized normal distribution for the data point 02, with an large outlier in $Im Z_{yx}$.

simulated decomposition parameters is compared with the standardized normal distribution, Normal ($x|0.1$), individually for each impedance data item in the top panel of Figure 5 by means of the q - q plots (for one arbitrarily chosen period, specifically 3.56 s here). The plots are constructed from 8000 quasi-independent states of the Markov chain and show a good degree of fit with respect to the reference normal distribution.

In the bottom panel of Figure 5, we also show, for the sake of completeness, q - q plots for selected parameters of the decomposition model from the Markov chain with respect to the respective best-fit Gaussian curves. As no distributional assumptions were made about the parameters, except for the uniform priors, the parameters are generally not normally distributed. It is especially evident for the twist and shear parameters which both show left-skewed distributions. The anomalous q - q plot for the real part of the first regional impedance is a result of our setting the lower limit for that parameter too high. This deficiency was corrected in later runs.

Stability of the MCMC inference with respect to the input data is another important factor of the model assessment, but its general analysis, involving the stability with respect to both the prior and to the likelihood, is far

beyond the scope of the present paper. The likelihood (10) misspecification is one of factors that may affect the bayesian inference made from practical MT data, especially if raw time signals or estimates of their spectral densities are not available. Chave and Jones [1997] and McNiece and Jones [2001] discuss serious inaccuracies in error bounds for the impedance estimates when parametric estimators were used in the data analysis codes. Incorrect variances of the experimental data in the misfit (8) and likelihood (10) result in false uncertainty assessment of the decomposition parameters, and often hamper the convergence of the MCMC procedure considerably.

Here, we present an example of incorrectly specified noise model for the experimental data and its effect on the MCMC inference outputs. The studied data are the simple single-point noisy synthetic impedances adopted from McNiece and Jones [2001] and analyzed as synthetic example I in Section 4.1. Similarly as earlier, we first generated a sample of 10 impedance tensors by contaminating the exact impedance (17) with Gaussian noise with the standard deviation of 4.5% of the largest impedance element. This model serves as a reference for subsequent tests. Histograms of selected decomposition parameters after $100k$ MCMC steps with the adaptive Metropolis sampler, which

approximate the marginal posterior probability densities of those parameters, are shown in Figure 6, along with the q - q plots of the normalized residuals for one set of principal impedance elements, illustrating the fit of the model data to the (simulated) experiment. Considering the relatively small size of the experimental data sample, the histograms show satisfactory resolution of the decomposition parameters, with the difference between the true values and the bayesian means not exceeding two times the sample variance of the respective histograms even in extreme cases.

In the second experiment, we generated the noise for the simulated data from a Laplacian distribution with the same variance as used in the Gaussian case. This effectively results in simulated experimental data showing more frequently larger outliers than would correspond to the normal noise model. For a sample of 10 noisy realizations of the impedance tensor, we again ran the *MCMC* sampling, but with the (misspecified) Gaussian likelihood (10). The gray histograms in Figure 7 show distributions of the decomposition parameters after 100k *MCMC* steps for this case. The resolution of the aggregate parameters, common to all the experimental data in the sample, is comparable to the case of the true Gaussian noise above. For regional phases, the resolution is poorer for data items showing large outliers, as is the case of the data point 02 (Figure 7). For that particular data point, we also display q - q plots of the normalized residuals with respect to the normal distribution in Figure 7, which show the largest misfit in the $Im Z_{yx}$ component, which is the most serious outlier in the simulated data set. For comparison, white histograms in Figure 7 show results of the *MCMC* sampling with the correct Laplacian likelihood applied in (8) and (10), but the differences between the cases of the (misspecified) Gaussian likelihood and the (true) Laplacian likelihood are not conclusive in this experiment.

A serious factor in applying *MCMC* sampling procedures is the computer intensity given by the necessity of sampling relatively large domains of the parameter space in often a multi-dimensional setting. In most cases, the evaluation of the direct model is the most time consuming part of the *MCMC* iterations, though for a rather simple *MT* composite model (6) other steps of the code may be of comparable complexity (random number generation, online evaluation of the variance-covariance matrix, etc.). In our experiments, typical CPU times were about 100 s for a set of experimental data with 10 impedance tensors (43 variable model parameters) and 100k *MCMC* iteration steps on a mini-notebook with the Intel Atom N280 (1.66 GHz) CPU, 1 GB RAM, and using the Compaq Visual Fortran, version 6.6c, compiler upon the Win XP operation system. With the number of experimental data items increased to 100 (i. e., with 403 variable parameters), the computation time increased to about 71 minutes on the same machine.

5. Conclusion

The *MT* decomposition is a problem that targets not only the parameters of the underlying composite model, but is interested in their uncertainties as well. Relatively weak manifestation of the deep symmetric regional structure and its masking by static distortions, sometimes extreme, of near-surface origin may result in excessively blurred images of the deep conductors. By aggregating the data over frequency bands and groups of sites presents a way of effectively focusing on poorly resolvable features of the regional structure, as shown by McNeice and Jones [2001] in their multi-site, multi-frequency decomposition study.

Technically, the decomposition is an optimization procedure aiming at the minimization of the objective function (7). In the above study by McNeice and Jones [2001], a direct minimization procedure is used to solve the decomposition problem. If converged, it provides point estimates of the decomposition parameters for the optimal composite model. The error estimates for these parameters can be then obtained by either a linearized projection of the data covariance matrix into the parameter space [e.g., Menke 1989], or, if non-linearities are essential, by a stochastic search in a neighbourhood of the optimal model, or by studying changes in the composite model due to stochastic variations of the data, e.g., by using a bootstrap procedure as in Groom and Bailey [1991] and McNeice and Jones [2001].

We present an alternative multi-site, multi-frequency decomposition procedure that is based on the bayesian formulation of the decomposition problem and its solution via a stochastic sampling by a *MCMC* technique. The procedure generates a chain that approximates samples from the posterior probability distribution of the *MT* composite model conditioned on the observed data. Though computationally demanding, the advantage of the procedure is that, by providing the results in the form of a posteriori probability distribution, the uncertainty information on the decomposition parameters is an inseparable component of the solution.

This paper is just a feasibility study that compares the performance of the stochastic decomposition with the results presented by McNeice and Jones [2001]. The comparisons carried out above show that the bayesian averaging from the *MCMC* samples provides estimates to the decomposition parameters statistically equivalent to those obtained from the direct optimization procedure. Confidence intervals for the individual parameters are immediately available at the output of the *MCMC* sampling. The *MCMC* procedure could be used with success even in the unfavourable case of a 2-D box model, with a large number of variables and strong correlations between the parameters. In this case, however, the computational demands cannot compete with those of the direct minimization. Nonetheless, if a bootstrap error testing should be supplemented for that

model, the amount of computations required would undoubtedly rocket as well.

The presented bayesian analysis allows us to relatively easily extend the scope of problems that have to be addressed in relation to the *MT* decomposition. E. g., the generalization to the model that considers also static magnetic distortions [Chave and Smith 1994, McNeice and Jones 2001] is straightforward. Moreover, the bayesian model selection is a suitable tool to answer the question of whether incorporating the magnetic distortions into the composite model is really required by the data. Another problem not addressed here is that of poorly estimated data variances in (7), which is not a rare case in practice, and affects even the *PNG* data set used here [for details, see McNeice and Jones 2001]. Bayesian approach makes it possible to include the data variances into the parameter set as nuisance variables [e. g., Gelman et al. 2004], and thus allows us to cope with deficiencies originating from the data processing.

Acknowledgements. Financial assistance of the Grant Agency of the Czech Republic, contracts No. 205/04/0746 and No. 205/07/0292, and of the Grant Agency of the Acad. Sci. Czech Rep., contract No. IAA200120701, is highly acknowledged.

References

- Bahr, K. (1988). Interpretation of the magnetotelluric impedance tensor: regional induction and local telluric distortion, *J. Geophys.*, 62, 119-127.
- Bahr, K. (1991). Geologic noise in magnetotelluric data: a classification of distortion type, *Phys. Earth. Planet. Int.*, 66, 24-38.
- Bailey, R. C. and R. W. Groom (1987). Decomposition of the magnetotelluric impedance tensor which is useful in the presence of channeling, 57th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 154-156.
- Chave, A. D. and A. G. Jones (1997). Electric and magnetic field distortion decomposition of BC87 data, *J. Geomagn. Geoelectr.*, 49, 767-789.
- Chave, A. D. and J. T. Smith (1994). On electric and magnetic galvanic distortion tensor decompositions, *J. Geophys. Res.*, 99, 4669-4682.
- Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin (2004). *Bayesian data analysis*. 2nd ed., Boca Raton.
- Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6, 721-741.
- Gilks, W. R., S. Richardson and D. J. Spiegelhalter (1999). *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*, London.
- Grandis, H., M. Menvielle and M. Roussignol (1999). Bayesian inversion with Markov chains I. The magnetotelluric one-dimensional case, *Geophys. J. Int.*, 138, 757-768.
- Groom, R. W. and R. C. Bailey (1989). Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion, *J. Geophys. Res.*, 94, 1913-1925.
- Groom, R. W. and R. C. Bailey (1991). Analytical investigations of the effects of near surface three-dimensional galvanic scatterers on MT tensor decomposition, *Geophysics*, 56, 496-518.
- Groom, R. W., R. D. Kurtz, A. G. Jones and D. E. Boerner (1993). A quantitative methodology for determining the dimensionality of conductivity structure and the extraction of regional impedance responses from magnetotelluric data, *Geophys. J. Int.*, 115, 1095-1118.
- Haario, H., E. Saksman and J. Tamminen (2003). Component-wise adaptation for MCMC. Reports of the Department of Mathematics, University of Helsinki, Preprint 342, pp. 1-20.
- Jones, A. G. and R. W. Groom (1993). Strike angle determination from the magnetotelluric tensor in the presence of noise and local distortion: rotate at your peril!, *Geophys. J. Int.*, 113, 524-534.
- Jones, A. G. and A. Schultz (1997). Introduction to the MT-DIW2 special issue, *J. Geomag. Geoelectr.*, 49, 727-738.
- Lilley, F. E. M. (1998a). Magnetotelluric tensor decomposition: Part I, Theory for a basic procedure, *Geophysics*, 63, 1885-1897.
- Lilley, F. E. M. (1998b). Magnetotelluric tensor decomposition: Part II, Examples of a basic procedure, *Geophysics*, 63, 1898-1907.
- MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*, Cambridge, UK, New York.
- McNeice, G. W. and A. G. Jones (2001). Multisite, multifrequency tensor decomposition of magnetotelluric data, *Geophysics*, 66, 158-173.
- Menke, W. (1989). *Geophysical data analysis: Discrete inverse theory*, International Geophysics Series, vol. 45, Academic Press, San Diego, California.
- Sambridge, M. (1999). Geophysical inversion with a neighborhood algorithm: I. Searching a parameter space, *Geophys. J. Int.*, 138, 479-494.
- Smith, J. T. (1995). Understanding telluric distortion matrices, *Geophys. J. Int.*, 122, 219-226.
- Zhang, P., R. G. Roberts and L. B. Pedersen (1987). Magnetotelluric strike rules, *Geophysics*, 52, 267-278.

*Corresponding author. Dr. Václav Červ, Institute of Geophysics, Acad. Sci. Czech Rep., v.v.i., Boční II/1401, 14131 Prague 4, Czech Republic; e-mail: vcv@ig.cas.cz