

Stochastic interpretation of magnetotelluric data, comparison of methods

Václav Červ⁽¹⁾, Michel Menvielle⁽²⁾⁽³⁾ and Josef Pék⁽¹⁾

⁽¹⁾ Geophysical Institute, Academy of Sciences of the Czech Republic, Prague 4, Czech Republic

⁽²⁾ Centre d'Études des Environnements Terrestre et Planétaire, Saint Maur des Fosses Cedex, France

⁽³⁾ Département des Sciences de la Terre, Université Paris Sud XI, Orsay Cedex, France

Abstract

Global optimization and stochastic approaches to the interpretation of measured data have recently gained particular attraction as tools for directed search for and/or verification of characteristic structural details and quantitative parameters of the deep structure, which is a task often arising when interpreting geoelectrical induction data in seismoactive and volcanic areas. We present a comparison of three common global optimization and stochastic approaches to the solution of a magnetotelluric inverse problem for thick layer structures, specifically the controlled random search algorithm, the stochastic sampling by the Monte Carlo method with Markov chains and its newly suggested approximate, but largely accelerated, version, the neighbourhood algorithm. We test the algorithms on a notoriously difficult synthetic 5-layer structure with two conductors situated at different depths, as well as on the experimental COPRODI data set standardly used to benchmark 1D magnetotelluric inversion codes. The controlled random search algorithm is a fast and reliable global minimization procedure if a relatively small number of parameters is involved and a search for a single target minimum is the main objective of the inversion. By repeated runs with different starting test model pools, a sufficiently exhaustive mapping of the parameter space can be accomplished. The Markov chain Monte Carlo gives the most complete information for the parameter estimation and their uncertainty assessment by providing samples from the posterior probability distribution of the model parameters conditioned on the experimental data. Though computationally intensive, this method shows good performance provided the model parameters are sufficiently decorrelated. For layered models with mixed resistivities and layer thicknesses, where strong correlations occur and even different model classes may conform to the target function, the method often converges poorly and even very long chains do not guarantee fair distributions of the model parameters according to their probability densities. The neighbourhood resampling procedure attempts to accelerate the Monte Carlo simulation by approximating the computationally expensive true target function by a simpler, piecewise constant interpolant on a Voronoi mesh constructed over a set of pre-generated models. The method performs relatively fast but seems to suggest systematically larger uncertainties for the model parameters. The results of the stochastic simulations are compared with the standard linearized solutions both for thick layer models and for smooth Occam solutions.

Key words magnetotelluric method – inverse problem – controlled random search – Markov chain Monte Carlo – neighbourhood algorithm

1. Introduction

Interpretation of geoelectrical induction data in seismoactive and volcanic areas often requires a directed search for characteristic structural details and quantitative parameters of the deep structure that could be directly compared to results of other, generally non-electrical investigations. Typical questions may be, *e.g.*, correlation

Mailing address: Dr. Václav Červ, Geophysical Institute, Academy of Sciences of the Czech Republic, Boční II/1401, 14131 Prague 4, Czech Republic; e-mail: vcv@ig.cas.cz

of clusters of seismic sources with boundaries of blocks with a significant conductivity contrast, detection, spatial mapping and conductivity estimates of magmatic plumes, delineation of groundwater layers and of deep aquifers, etc. Due to the inherent ambiguity of the inverse problem solution with noisy and incomplete data, those structural features may be missed single model linearized inversions, or smeared over large intervals of depths if smoothness of the inverse model is emphasized by applying a roughness penalty for regularization.

Stochastic approaches to the interpretation of measured data have recently gained attention in geophysical applications (*e.g.*, Mosegaard and Tarantola, 1995). They are particularly attractive as tools for estimating the model parameters from samples from their probability distributions, for quantifying uncertainties in the estimated model parameters, and for structural hypotheses testing. The advantage of the stochastic simulations is that they aim at performing the search throughout the parameter space and pick up the models according to their probabilities measured by the misfit and priorities for specific model features. The main difficulty of the stochastic inversions is that they are carried out via often extremely computer intensive simulations.

Mapping the parameter space has long been an ambition of global minimization methods in geophysical inversions (*e.g.*, Senn and Stoffa, 1995). Though these methods primarily aim at searching for a global minimum of a function, they always operate on a large population of models and rely on random factors within their optimization strategy to reduce the risk of being trapped at local minima of the target function. By this mechanism, the optimization procedure generates a topography map of the target function along a series of randomly perturbed paths in the parameter space, with usually dense coverage in the vicinity of the minima. The model population generated by a global optimization procedure does not generally follow the true probability density function of the models.

In this paper we aim at demonstrating how ensembles of models produced by selected global optimization procedures compare with results of single model linearized inversions as well as with results of a probabilistic sampling. As repre-

sentatives of the global minimization procedures, we have selected the Controlled Random Search (CRS) method (Price, 1977) and the Neighborhood Algorithm (NA) by Sambridge (1999a), both being relatively fast global minimizers with only low demands on a model-conditioned fine-tuning of the optimization parameters. A fully stochastic sampling is based on a simple implementation of the Markov Chain Monte Carlo (MCMC) procedure adopted from (Grandis *et al.*, 1999) and modified for thick layer 1D MT inversion with both variable resistivities and layer thicknesses. The results are also compared with the standard Occam smooth inversion (Constable *et al.*, 1987) and layered linearized inversion by Weaver and Agarwal (1993).

As to the model structures, we restrict ourselves to 1D layered MT models here that are better suited if structural details like sharp electrical boundaries, thin sheets or sandwich structures with a tectonic significance are targets of the search. Moreover, the easiness of 1D MT direct solutions allows us to carry out all the computer intensive tests more completely than would be possible under the serious limitations imposed by using 2D or even 3D direct modelling codes. Specifically, we test the selected optimization procedures on a simple COPROD1 benchmark model and then analyze a synthetic model with two conductive layers separated by a non-conductor in detail. The latter example presents a generally non-elementary problem, especially as regards the detection and resolution of the deeper conductor, and is typical of tectonic settings in many active areas.

2. Global optimization and stochastic sampling methods

2.1. Controlled Random Search

Controlled Random Search (CRS) is a simple yet often very efficient optimization algorithm first suggested by Price (1977). It starts with a sufficiently large pool of randomly generated models, say \mathbf{p}_i , $i=1, \dots, N_p$ with the target values $\varphi(\mathbf{p}_i)$, where each vector \mathbf{p}_i consists of N_S parameters that describe the physical model. The CRS algorithm then proceeds by employ-

ing a set of heuristic rules to generate a new test model from the current pool, say $\mathbf{p}_T = \mathbf{R}(\mathbf{p}_1, \dots, \mathbf{p}_{N_p}, \text{rand})$ where ‘rand’ symbolizes a random factor that diversifies the model population and reduces the risk of the algorithm being trapped in the local minima of the target function. If the new test model \mathbf{p}_T is better, in terms of the target misfit, than the currently worst model in the pool it is used to replace the latter. If, moreover, $\varphi(\mathbf{p}_T)$ is less than the currently best model misfit in the pool then many heuristics suggest carrying out an additional detailed local search in the vicinity of this successful point. By repeatedly applying the above steps to the model population the model pool develops and moves towards regions with better target values until a termination criterion is met, defined via a target threshold, by a minimum of a parameter change between successive iterations, or by setting a limit on the number of iterations.

Particular versions of the CRS algorithm largely differ by the heuristics used for generating the new test model. The original CRS algorithm by Price (1977) uses a simplex of $N_S + 1$ models randomly chosen from the current pool. The new test model is generated by mirroring the $(N_S + 1)$ -st model from the simplex through the centroid of the first N_S models,

$$\mathbf{p}_T = \frac{2}{N_S} \sum_{j=1}^{N_S} \mathbf{p}_j - \mathbf{p}_{N_S+1}.$$

This CRS version will be further referred to as CRS1.

A number of other heuristics are discussed and compared, *e.g.*, in Tvrdík *et al.* (2002). One of the approaches showing good performance in 1D MT inversions is a heuristics proposed by Montaz *et al.* (1997), referred to as CRS6 in what follows. It generates the new test model by component-by-component fitting parabolas through the currently best model and two further models selected randomly from the current pool. The parameters of the new trial model \mathbf{p}_T are then given by the minima of the individual parabolas.

For a given number of model parameters to be estimated the performance of the CRS algorithms in 1D MT inversions depends primarily on the size of the pool from which new test models are generated. The results indicate that

CRS6 is largely superior to CRS1 as regards the convergence speed. However, the CRS1 simplex algorithm seems to better explore the parameter space and shows lower tendency to end up in secondary local minima. In general, CRS with repeated runs presents a relatively fast and reliable global optimization algorithm, which provides us rapidly with information about the complete spectrum of the minima of the target function under study.

2.2. Neighborhood algorithm

Recently, a new global optimization procedure was suggested by Sambridge (1999a) for a seismic inversion, based on an iterative evolution of the initial model pool by randomly sampling the target function in the immediate vicinity of the pool models. The vicinity of a model is defined in a most natural way, as a subregion in the parameter space that comprises all the nearest points with regard to this particular model. Formally, the nearest neighborhood of a pool model \mathbf{p}_i , $i \in \{1, \dots, N_p\}$ consists of all models \mathbf{p}_j , which meet the condition $\|\mathbf{p} - \mathbf{p}_i\| \leq \|\mathbf{p} - \mathbf{p}_j\|$; $j \in \{1, \dots, N_p\}$, $j \neq i$. In this way, the whole parameter space is divided into a system of convex Voronoi cells, each of them representing the nearest neighborhood to a particular pool model.

The optimization by the Neighborhood Algorithm (NA) proceeds by carrying out N_p/N_r random steps within N_r selected Voronoi cells with the best target values, generating thus N_p new models in the regions of the minimum misfit in each algorithm step. The random walks within the Voronoi cells are carried out by using the standard Gibbs sampler (Gelman *et al.*, 1995; also see next section) restricted to the selected cell. Sambridge (1999a) demonstrated both the optimization and numerical efficiency of the NA algorithm which adapts fast according to the shape of the underlying target functions and relatively rapidly provides a map of domains around the target minima. A particular choice of the factor N_r controls the preference of the algorithm for a local search or for global exploration.

Numerical tests for a 1D inverse MT problem indicate that the NA shows quite a similar performance as the CRS algorithms above.

2.3. Monte Carlo with Markov Chains

Stochastic sampling methods are used to generate model samples distributed proportionally to the likelihood of the models. The sample probability distribution can then be used to numerically approximate Bayesian integrals that provide single parameter estimates or information about their credibility. The Monte Carlo methods with Markov Chains (MCMC) are algorithms to generate sample ensembles for the Monte Carlo integration.

In the Bayesian approach, the inverse problem solution is represented by a posterior probability distribution of the model given the data and prior information on the model (*e.g.*, Grandis *et al.*, 1999)

$$\text{Prob}(\mathbf{p}/\mathbf{d}) = \frac{\text{Prob}(\mathbf{p}/\mathbf{d})\text{Prob}(\mathbf{p})}{\text{Prob}(\mathbf{d})}$$

$$\text{Prob}(\mathbf{p}/\mathbf{d}) \propto \exp\left\{-\frac{1}{2}\sum_{l=1}^{N_D}\left[\frac{d_l^{\text{mod}}(\mathbf{p}) - d_l^{\text{obs}}}{\delta d_l^{\text{obs}}}\right]^2\right\}$$

where $\text{Prob}(\mathbf{p})$ is the prior probability distribution, and the likelihood function $\text{Prob}(\mathbf{d}/\mathbf{p})$ in the form given above assumes that the data items follow a normal distribution law $N(d_l^{\text{obs}}, \delta d_l^{\text{obs}})$, $l = 1, \dots, N_D$. If no specific prior information on the model parameters is available, we can assume, for a layered Earth's model, that logarithms of both the resistivity and thickness assume constant values within sufficiently broad limits for each layer. We use this non-informative prior in all the MCMC simulations presented in Section 4.

Without going further into detail here (see, *e.g.*, Gelman *et al.*, 1995; Grandis *et al.*, 1999, for reference), the idea behind the MCMC is to create a random walk, or Markov process, that has $\text{Prob}(\mathbf{d}/\mathbf{p})$ as its stationary distribution and then to run the process long enough so that the resulting sample closely approximates a sample from $\text{Prob}(\mathbf{d}/\mathbf{p})$.

The Gibbs sampler is one of the particular methods used to construct such a Markov process. The Markov chain relies on updating the model parameters in the process of successively scanning the parameter domain under study. In each scan, we update one single model param-

eter, say p_k , by drawing a new value from the one-dimensional conditional probability distribution $\text{Prob}(p_k/\mathbf{d}, p_1, \dots, p_{k-1}, p_{k+1}, \dots, p_{N_S})$, *i.e.* with all parameters except the k -th one fixed at their current values. One MCMC step is completed after all N_S model coordinates have been updated. In this way, an ergodic Markov chain is designed with an invariant probability law identical with $\text{Prop}(\mathbf{p}/\mathbf{d})$. The posterior marginal probability distributions for the parameters as well as various Bayesian integrals are then estimated from the successive simulated values of this Markov chain.

2.4. Resampling with the neighborhood algorithm (NAR)

MCMC is a computer intensive method since a series of direct problem solutions is required for each parameter scan within the Gibbs sampler. If an ensemble of direct solutions already exists from computations carried out previously, an effective resampling procedure can be applied to this model ensemble as suggested by Sambridge (1999b). The models generated, *e.g.*, by previous global optimization runs are used to approximate the misfit function by the neighborhood algorithm, *i.e.* by a piecewise constant function in a Voronoi mesh with cells centered at the models available. Then, the MCMC sampling is carried out for that surrogate target function in exactly the same way as the full MCMC simulation above, but without any additional direct problem solution required. Clearly, the success of this approach highly depends on how good a support the underlying model ensemble gives to the true misfit function, which is a delicate question in multidimensional parameter spaces, and might be also a problem-dependent issue.

3. Test data sets

We selected two specific data sets to analyze the performance of the stochastic optimization procedures in situations often encountered in various interpretations of electromagnetic induction data. The first Data Set (DSI) is synthet-

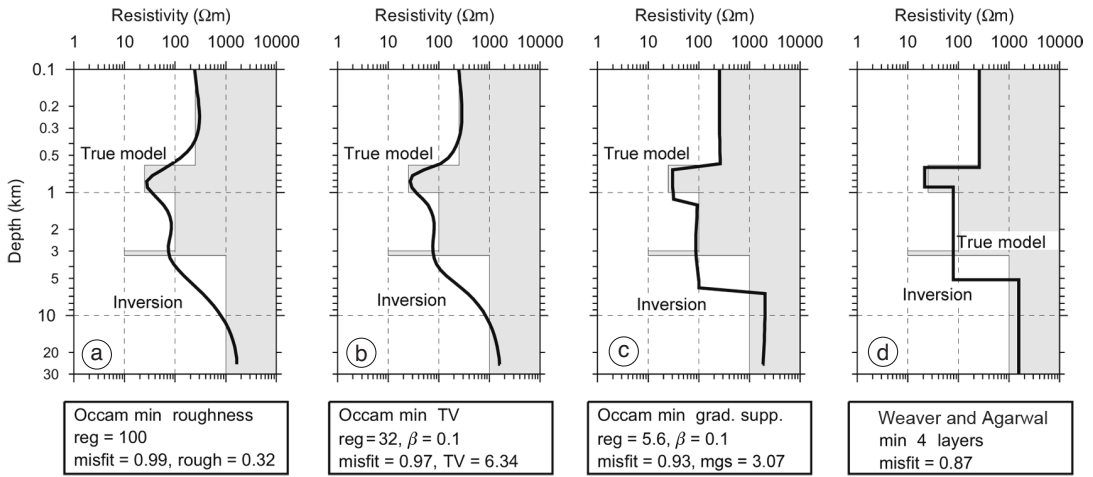


Fig. 1a-d. Results of four different linearized inversion routines applied to the synthetic data set DSI. The data generating model is marked by a gray-white contrast in all sub-panels, and its layer parameters are: layer 1 – $h_1=0.6$ km, $\rho_1=250$ Ωm ; layer 2 – $h_2=0.4$ km, $\rho_2=25$ Ωm ; layer 3 – $h_3=2$ km, $\rho_3=10$ Ωm ; layer 4 – $h_4=0.25$ km, $\rho_4=10$ Ωm ; basement 5 – $\rho_5=1000$ Ωm . MT data were generated from this model for 41 log-regularly spaced periods within the range from 10^{-3} to 10 s, and the synthetic impedances were further contaminated with Gaussian noise with the standard deviation of 5% of the maximum impedance element for each period. Inverse results in the sub-panels correspond to: a) Occam inversion with roughness penalty (Constable *et al.*, 1987); b) Occam inversion with total variation penalty (Portniaguine and Zhdanov, 1999); c) Occam inversion with gradient support penalty (Portniaguine and Zhdanov, 1999); d) inversion for a minimum number of layers according to Weaver and Agarwal (1993). In the boxes below the panels the respective regularization weights (reg), misfit (RMS^2) values, and other routine-specific parameters are given.

ic and is derived from a model with a shallow conductor over a deeper conductive target. This general setting occurs in a variety of practical situations, especially if a layer of conductive sediments screens deeper layers containing fluids, but also in large-scale settings like those of an electric asthenosphere screened by a conductive lower crust. It is generally known that the deep conductor is difficult to resolve unless its conductance is considerably higher than that of the shallower screening layer.

The particular data generating model used here was adopted from (Grandis *et al.*, 1999). Model for DSI is shown in fig. 1a-d, marked by a gray-white contrast in all sub-panels, and its parameters are given in the figure caption. Conductances of the layers, $S=h/\rho$, are 2.4, 16, 20 and 25, in Siemens, from the top to the bottom, and indicate that there is only little chance to resolve the deep thin conductor situated at the

depth of 2 km. It is clear from fig. 1a-d that for 5% of Gaussian noise added to the synthetic MT data no common linearized inversion gives any indication that a deeper conductor exists beneath the shallow conductive layer. They all rather extend the intermediate third layer to greater depths and smear the deep thin conductor over a depth range of more than 2 km into the high-resistivity basement.

The second Data Set (DSII) is the experimental COPROD1 set introduced and analyzed in detail by Jones and Hutton (1979a,b), later used for demonstrating the performance of the Occam 1D inverse approach by Constable *et al.* (1987), and commonly employed as a benchmark data set for performance tests of 1D inverse techniques in magnetotellurics. The COPROD1 data represent a set of 15 pairs of log apparent resistivity and MT phase data items within the period range of 20 to 2000 s and

correspond to a relatively simple structure with one pronounced conductive layer at a lower crustal/upper mantle depth. Nonetheless, the data set is by far not elementary, especially owing to the data error structure that does not seem to follow any simple noise model.

4. Search in the parameter space: model simulations for the test data sets

4.1. Synthetic data set

In the subsequent numerical tests we concentrate on the comparison of the performance of the selected global and stochastic optimization procedures for parameter estimation and uncertainty quantification in simple but practically relevant magnetotelluric settings. Availability of relatively simple and reliable linearized inverse procedures for 1D MT problems makes it possible to assess the extra information we can infer from sampling the whole parameter space as compared to searching directly for one specific inverse solution.

A first series of numerical experiments aimed at comparing the selected methods applied to the synthetic data set DSI. The underlying model for DSI is a 5 layer structure with a deep thin conductor that is clearly hard to detect in the surface MT data. We experimented successively with models consisting of 3 to 6 layers, including the basement, to test the outputs of the individual optimization procedures. We intentionally used very broad *a priori* limits for all layer parameters, specifically $10^{-3} \leq h_i \leq 10$ km, $1 \leq \rho_i \leq 10^4$ Ω m, $i=1, \dots$, number of layers to test the convergence as well as the degree of risk of the individual methods to end up in false minima.

With 3-layer test models the best attainable RMS² misfit was 1.88 and none of the methods could fit the data satisfactorily. Both the CRS6 (pool of 100 models, 5000 iterations, 200 runs, accepted all models with RMS² ≤ 4) and MCMC (50000 steps, burn-in 10000, thinning 100) could recover the resistivity of the first layer and that of the basement with high accuracy, but they were not able to satisfactorily simulate the fine resistivity structure in the intermediate part of the model section by one single layer. The second layer

does, however, provide the true integral conductance of the three original conductive layers with good accuracy, within less than 5% of the true value. Both methods compensate for the insufficient resistivity contrast between the first and second layer by systematically decreasing the depth to the conductor by about 100 m as compared to the true model. Figure 2 indicates the mean values of the individual parameters are displayed, and ranges comprising 50 and 90% of all the parameter values provided by the simulation techniques. Though not so comprehensive as complete parameter histograms, these plots qualitatively provide a good idea about both the correctness of the recovered parameters, their uncertainty and skewness of the underlying histograms.

After having increased the number of layers to four, the fit to the data improved substantially. Parameters of the CRS6 and MCMC procedures were the same as those specified for the case of 3 layers above; from the CRS outputs only models with RMS² ≤ 1 were accepted now for constructing the histograms.

In the 4-layer case, both the CRS6 and MCMC recover the top layer resistivity almost exactly, while the thickness of the first layer is systematically greater by about 50 m than the true value and is scattered over about 40 m for the MCMC estimate, if the interval between the 25 and 75% quartiles of the parameter samples is used to measure the uncertainty. As the CRS6 was always terminated after a fixed number of 5000 iterations, repeated values of h_1 during the final stage of the iteration process lead to over-optimistic estimates of the parameter uncertainties, specifically 20 m for h_1 , as compared to the probabilistic MCMC sampling. Contrary, the parameters of the second layer, *i.e.* the shallow conductor, are recovered only with large uncertainty, especially as regards its thickness. Nevertheless, both methods produce relatively sharp estimates of the conductance of this layer, within 1.5 S, which are by about 4 S lower than the true value of $S_2=16$ S. The deficit of the conductance in the second layer is compensated by increased conductance of the layer beneath, which tries to account for the summary effect of the third layer in the true model as well as for that of the deep thin conductor situated on the top of the basement.

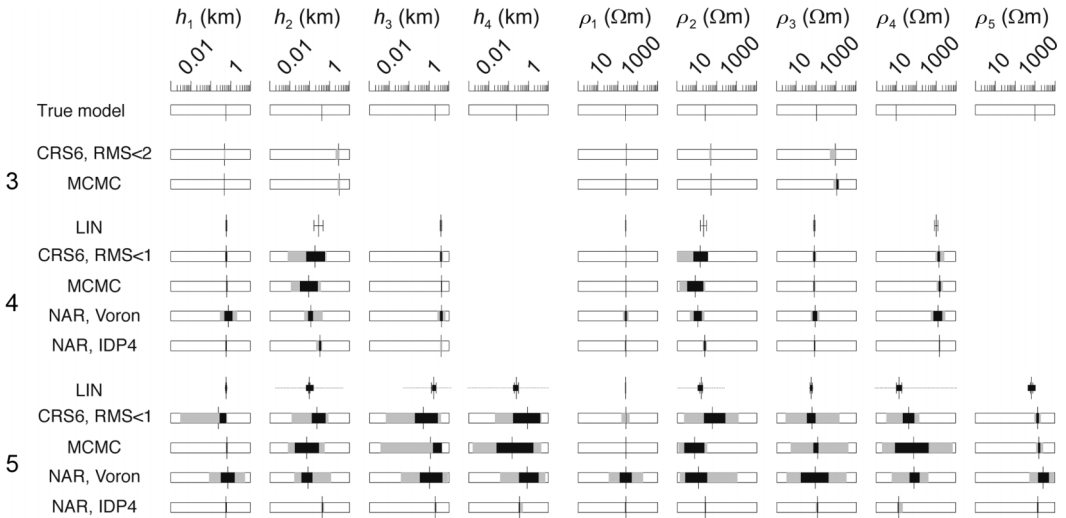


Fig. 2. Mean values (vertical bars) and ranges comprising 50 and 90% of all parameter values (black and gray zones, respectively) provided by various global optimization and stochastic sampling runs applied to the DSI data set with 5% of Gaussian noise. The vertical bars in the top row indicate the true values of the parameters of the synthetic model. For specific routine parameters see the text. For the 4 and 5-layer models, the row labeled LIN shows parameter estimates (long bars) and variances (short bars) obtained for classical linearized inverse solutions. The classical parameter variances for the 5-layer model from a truncated SVD analysis (only the seven largest from altogether nine singular values were used to compute the parameter variance-covariance matrix) are shown by full lines, and those obtained with addition of the nul space of the solution are indicated by dotted lines. Black rectangles show the restrictions to the parameter acceptability range for the non-linear model (models with RMS less than one here).

As the MCMC sampling is an extremely time-consuming procedure, even for a simple 4-layer MT model, we also tested the relatively fast neighborhood resampling algorithm (NAR) by Sambridge (1999b) on the present data set DSI. We applied this procedure to a complete set of models generated by the CRS6 algorithm in 20 successive runs with different initial pools, 100 models each, and with 5000 iteration steps in each run, which makes altogether about 100 000 to 150 000 individual models distributed all over the parameter space, but with most of them concentrated in close neighborhoods of the main minima of the target function. We performed 20 000 iterations of the NAR to produce a resampled set of models for further stochastic inference. We can see from fig. 2 that the NAR resampling gives results similar to those of the MCMC, but with clearly more diffuse statistical ranges for those model parameters that have

been recovered sharply by the MCMC sampling above. Comparison of true target misfits with those obtained by the NA interpolation in the Voronoi mesh during the NAR runs shows that the NA interpolation may extend the domain of the target minimum widely. As a consequence, models with large misfits may be accepted more frequently into the resampled chain than it would be appropriate to their true likelihood.

For the 4-layer MT model tested here, the difference between the MCMC and NAR outputs can be illustrated by stacked model samples shown in fig. 3 (top row of panels). Here, the gray scale is used to visualize the relative number of models from the sample ensemble which pass through the individual cells in the $\log \rho - \log z$ plane. The plots show limits for the resistivity indicated by the sample models as a function of depth. The fuzziness of the NAR results could be partly reduced by using a more

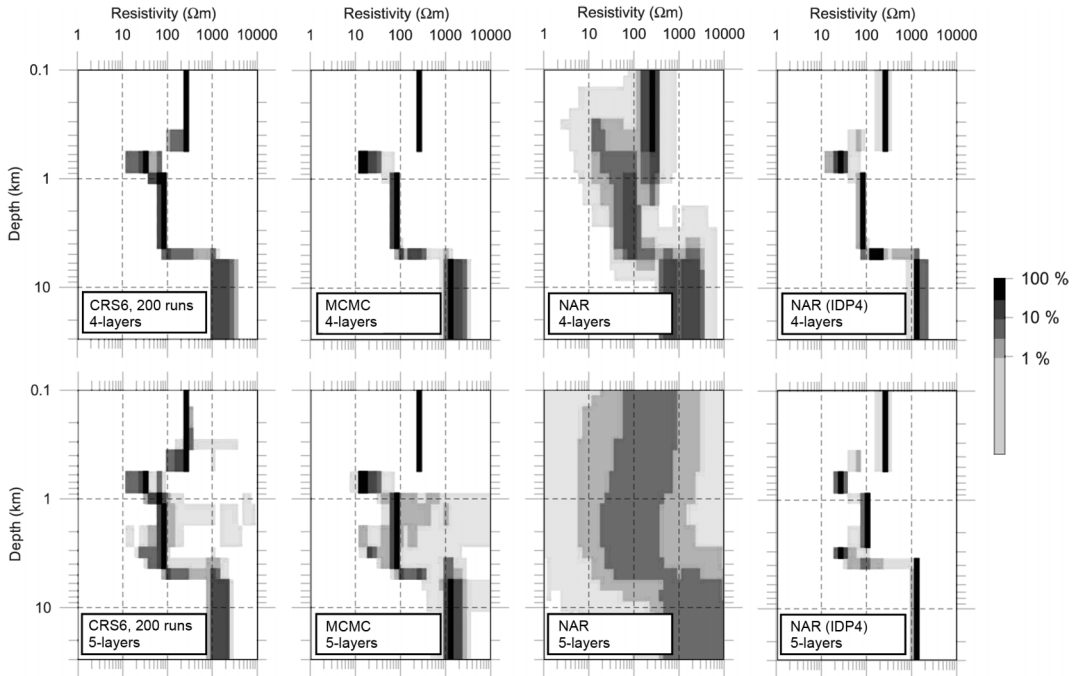


Fig. 3. Gray-shade plots of stacked models produced by CRS, MCMC and NAR runs applied to DSI data set with 5% of Gaussian noise. Top row of panels show the results for a 4-layer model approximation, bottom panels are for inversion with 5-layer models. The degree of gray color indicates the relative number of models that pass through 0.1×0.1 sized cells in the $\log\rho\text{-}\log z$ plane.

accurate, but also greatly more time-consuming, interpolation scheme employing a standard inverse distance power procedure (with the power of 4 here). Unfortunately, other experiments indicate that this improvement is not systematic and the performance of the NAR with various interpolants depends highly on the dimension of the problem and the particular distribution of the interpolated models throughout the model space.

For comparison with a classical single-model inverse solution, we also present in fig. 2 results of a *posteriori* linearized uncertainty analysis based on the Singular Value Decomposition (SVD) of the sensitivity matrix for the 4-layered model obtained earlier by the inversion algorithm by Weaver and Agarwal (1993) (see fig. 1a-d for the model structure). The structure of the singular vectors corresponding to the four

largest singular values of the sensitivity matrix for this model weighted by data variances indicates that the most significant parameters in the model are h_1 , $S_3 = h_3/\rho_3$, ρ_1 , and $S_2 = h_2/\rho_2$, while the two smallest singular values suggest the resistivity of the basement, ρ_4 , and the resistance of the second layer, $T_2 = h_2\rho_2$, to be the least significant parameters. This ranking is also clearly manifested in fig. 2 by the individual parameters' error bars, computed as square roots of the diagonal elements of the parameter covariance matrix (Menke, 1989). These error bounds approximately demarcate the range of models that are equivalent, as to their fit to the experimental data, with the original model analyzed. Though the classical variances are qualitatively in accord with the parameter uncertainties suggested by the stochastic simulations, their magnitudes are known to be underestimated (Menke, 1989).

Obtaining more realistic error bounds within the classical analysis would require more sophisticated approaches to the estimates of the parameter limits to be applied, as those suggested, *e.g.*, by Pous *et al.* (1985) for linear or by Meju and Hutton (1992) for non-linear problems, or simply performing a random search for acceptable models in a certain vicinity of the original inverse solution.

For test models with 5 or more layers, the inverse problem becomes ill-posed and most of the inverse methods suffer in some way from parameter indefiniteness or inter-dependence. Often, the physical sense of the parameters gets lost in multi-layered models since characteristic conductive/resistive features of the real structure cannot be attributed to the model parameters in an unambiguous way. In particular, the histograms of the model parameters are difficult to interpret in terms of marginal distributions since they can, in fact, mix together true parameters of more than one physical domain.

The bottom panels in figs. 2 and 3 show results of the global and stochastic sampling for 5-layer test models. Except for the resistivities of the first and last layer, which are well constrained by the data, all the other parameter histograms are diffuse and, on a more detailed scale, even multimodal. The concentrated histograms resulting from the NAR algorithm with the inverse distance power (power 4) interpolant rather suggest that the method stayed trapped in a vicinity of one specific minimum dictated by the particular distribution of the CRS set of models.

The stacked model plots in fig. 3 provide a clearer visual representation of the model samples with respect to the resistivity-depth section. They show very similar conductivity patterns for the shallow part of the structure down to the bottom of the first conductor for both the 4 and 5-layer models. The third layer is recovered sharply but already shows a typical high-resistivity 'shadow' in the 5-layer case, which is an artifact due to a low sensitivity of the MT field to resistive domains beneath the first, shallow conductor. As regards the manifestation of the deep thin conductor at the depth of 3 km, less than 10% of the models in both the CRS and MCMC patterns indicate a layer of in-

creased conductivity between 2 and 3.5 km. The standard NAR algorithm gives a largely diffuse image of the resistivity structure which indicates that the underlying set of CRS models (from 20 runs of CRS6, with the pool of 100 models and 5000 iteration steps) does not cover the parameter space properly. This situation did not change for better even when the support models were generated by the NA minimization algorithm instead of CRS.

Due to a massive inter-dependence of the parameters in multilayer cases, standard MCMC Gibbs sampler mixes poorly and requires very long chains, of the order of hundreds of thousands of samples, to walk throughout the parameter space with sufficient density. Nevertheless the resistivity pattern in fig. 3 remains quite stable after a few tens of thousands of the MCMC steps, though its detailed probabilistic interpretation may be questionable for chains so short.

The stacked model resistivity image shows stable structural features also for increasing number of layers in the test models, though especially the high-resistivity artifacts become more pronounced in runs with many layers.

The classical SVD analysis of the 5-layer linearized inverse solution allows similar conclusions to be made as regards the parameter uncertainty. Contrary to the 4-layer model studied earlier, the SVD spectrum for the 5-layer model indicates that at least two singular vectors form the nul space of the inverse solution, specifically those corresponding to the resistances of the second and fourth layers, $T_2 = h_2 \rho_2$ and $T_4 = h_4 \rho_4$. Then the linearized variances for the parameters involved in these combinations are not bounded within the range of the physical acceptability of the models. Figure 2 shows parameters of one particular classical 5-layer inverse solution along with their linearized variances for the case of a truncated SVD spectrum, with only seven largest singular values from nine left for the computation of the parameter covariance matrix, as well as for the case of the full SVD spectrum. In the latter case, the projection of the nul space solutions onto the space of the physical parameters gives very large, or almost infinite, error bounds for the poorly resolvable parameters. The predicted linearized parameter variances have to be further verified,

and properly reduced, by an extra procedure so that they do not exceed the region of acceptable models for the underlying non-linear problem.

One additional difficulty of the uncertainty analysis based on a linearized inverse single-model solution is that the SVD spectra may differ not negligibly for different inverse models due to the non-linearity of the underlying direct problem, especially if the experimental data are sufficiently noisy, suggesting thus sufficiently different nul spaces for different inverse models. In the stochastic sampling procedures, this equivalency across various classes of models may substantially deteriorate the convergence properties, but for long enough chains all the model classes should be visited and properly sampled in the course of the stochastic walk.

4.2. Experimental COPROD1 data set

We carried out a similar series of numerical simulations also for the experimental data set DSII (COPROD1) with the standard data errors adopted from Constable *et al.* (1987). This data set is specific especially due to its error structure. We show in fig. 4 a projection of all the RMS^2 misfit values obtained from twenty CRS6 runs (3 layer test models, pool size 100, 3000 iterations) as a function of the resistivity of the first layer. The figure clearly shows a bimodal character of the misfit, with a deep minimum at ρ_1 close to $170 \Omega\text{m}$, but with $\text{RMS}^2 > 1$ and with another broad, but very flat zone, $\rho_1 > 300 \Omega\text{m}$, with misfits $\text{RMS}^2 < 1$. The former target minimum attracts most of the solutions of both the CRS and MCMC runs, and especially for MCMC with 3-layer models does not allow the chain to leave its ‘gravity’ in reasonable time even in case that the chain has been started from the minimum of the target function, provided by the CRS. In this case, the CRS map of the target minima provides valuable information for constraining the parameters for the MCMC simulation.

Figure 5 summarizes the results of CRS, MCMC and NAR simulations with 4-layer models for the DSII data set in the form of stacked $\log\rho - \log z$ gray-shade plots. Similarly as in the synthetic case, very broad *a priori* parameter

ranges were chosen to simulate minimum prior knowledge of the parameter limits. Specifically, we have set the layer thicknesses to be between 3 and 300 km and the resistivities between 1 and $10^5 \Omega\text{m}$, except for the first layer with $\rho_1 > 180 \Omega\text{m}$ to avoid the ‘target trap’ described above.

The resistivity patterns from the CRS6 (200 runs, pool size 100, 3000 iterations) and MCMC (50000 steps, burn-in 10000, thinning 100) runs show very similar structures, though the CRS pattern is sharper due to excessive accumulation of the solutions close to the target minima. The models seem to prefer very high resistivity values of the top layer, but constraining this parameter does not change the deeper resistivity pattern noticeably. Similarly as in the synthetic case above, the NAR resampling algorithm smears the resistivity patterns over a broader range of values. In the case studied here, no dramatic improvement was achieved if the inverse distance power interpolant was used within the

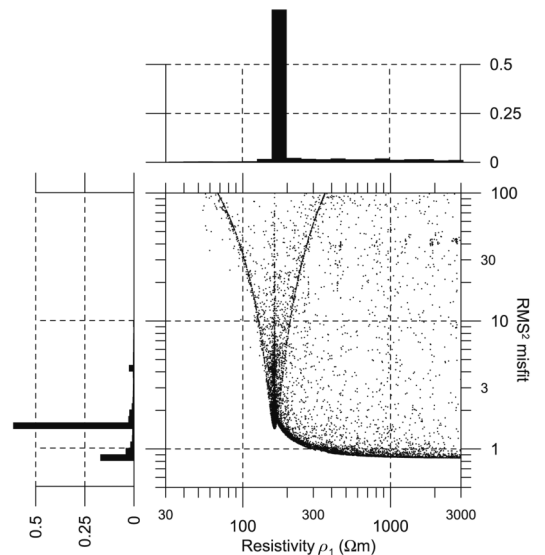


Fig. 4. Projection of all RMS^2 misfits versus the resistivity of the top layer obtained in the course of 20 CRS6 runs applied to the COPROD1 (DSII) data set with 3-layer test models. Along the axes, the corresponding histograms of the RMS^2 and ρ_1 values are shown.

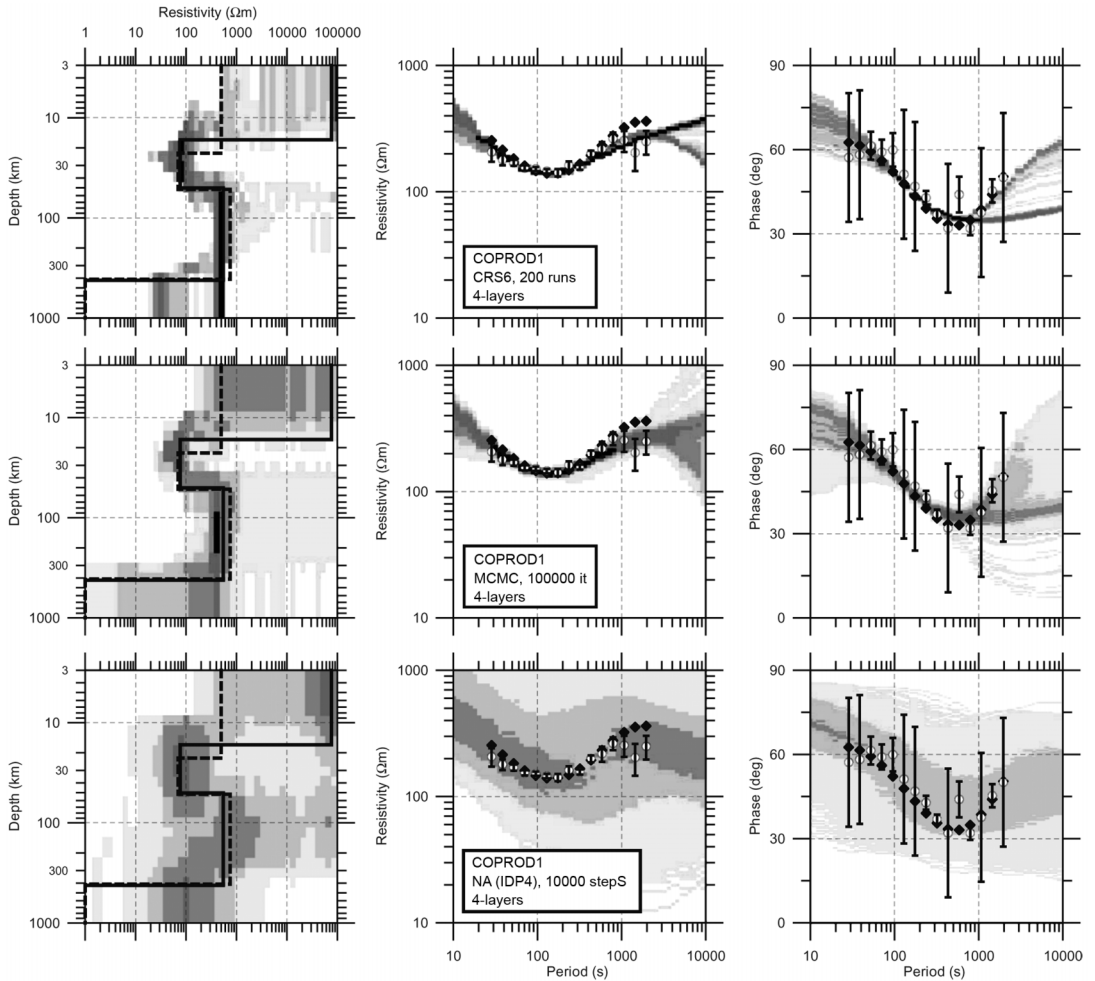


Fig. 5. Left column: gray-shade plots of stacked models obtained from CRS, MCMC and NAR runs applied to the COPROD1 (DSII) data set with 4-layer test models. The dashed line indicates a preferred 4-layer solution by Jones and Hutton (1979b) who minimized the misfit function without considering the confidence weights. The full line shows a 4-layer solution from the inverse procedure by Weaver and Agarwal (1993) applied to the data supplemented with error bars according to Constable *et al.* (1987). Central and right columns: data fit for the apparent resistivities and phases, respectively, for models generated by the CRS, MCMC and NAR runs. The experimental COPROD1 resistivities and phases are shown by white circles with error bars according to Constable *et al.* (1987), the black dots are solutions for the model by Jones and Hutton (1979b).

NAR algorithm instead of the Voronoi tessellation. Figure 5 also presents the fit of the 4-layer model data from the above simulation runs to the experimental apparent resistivities and phases.

5. Discussion and conclusions

We applied three common methods of global and stochastic optimization, specifically the CRS, MCMC and NA/NAR, to simple 1D mag-

netotelluric models to test their performance in problems that target at searching for and at verifying distinct structural features in geoelectrical sections, such as boundaries with a large conductivity contrast or conductors screened by shallow conductive structures. Similar structural elements are often typical of tectonically active areas, and their detection, as well as estimation of their geometrical and electrical parameters, along with a proper specification of the uncertainty of these estimates, are indispensable prerequisites for carrying out any analysis of possible correlations between electrical structural factors and non-electrical indicators of the tectonic activity.

CRS and NA are relatively simple, in terms of the tuning complexity, but effective global optimization algorithms that are suitable both for providing a quick inspection of the target function topography and for searching for the inverse problem solution. Performing repeated runs of these algorithms with different randomly generated initial model populations gives not only a good idea of the character and multimodality of the underlying target function, but also draws an image of the parameter bounds (uncertainties), though not in quite fair proportions with regard to the likelihood of the models. The parameter distributions provided by the CRS or NA are affected by the tuning conditions of the minimization procedure and may be, *e.g.*, excessively peaked if the procedure is terminated after a fixed (and large enough) number of iteration steps, or diffuse and shifted if the optimization is stopped after a certain portion of models in the pool, or all of them, reach a predefined misfit threshold.

Stochastic sampling techniques aim at providing unbiased model samples distributed according to the likelihood of the models. For layered MT models with very broad *a priori* limits for the model parameters, the MCMC Gibbs sampler procedure shows a good convergence provided the underlying test models are not largely underdetermined. In the opposite case, the Markov chain may mix poorly unless closer and more realistic parameter bounds are specified, especially for the layer thicknesses. With broad parameter ranges, MCMC shows frequent irregular transitions between fine modes of the target func-

tion and does not provide stable parameter marginal distributions unless a very large number of MCMC steps is carried out. It can, however, provide relatively quickly a stable approximate image of the resistivity *versus* depth distribution via stacking the models from the chain.

Resampling model ensembles produced by relatively fast global optimization procedures, such as CRS or NA, is another way to estimating the model parameters and their true uncertainty. The NAR resampling based on approximating the misfit function by constants in an ensemble-related Voronoi mesh gives results similar to those of MCMC if sufficiently narrow bounds for the MT model parameters can be specified *a priori*. For broad parameter limits and highly underdetermined models, NAR tends to produce more diffuse posterior distributions for the parameters as compared to MCMC. One of the likely reasons for this behavior may be that models produced by a fast converging CRS procedure do not provide sufficient support for the NA interpolation throughout the parameter space. As a consequence, the high probability region may expand via the NA interpolation far beyond its true limits. In some cases, using a more accurate interpolation technique (*e.g.*, inverse distance power interpolant here) could improve the results, but it cannot be generalized.

As compared to standard single-model linearized inversions, the global and stochastic inverse methods provide additional information, especially as regards the possible ranges of model parameters and uncertainty of the parameter estimates conditioned on the observed data, though for the price of higher, and sometimes enormous computation times. They are of particular interest in situations like, *e.g.*, directed search for specific structural features, inversion with various kind of prior information, or in jointly solving a problem of parameter estimation and model selection (Malinverno, 2002).

Acknowledgements

Financial assistance of the Grant Agency of the Czech Republic under grants No. 205/04/0740 and No. 205/04/0746, and of the Ministry of Education, Youth and Sports of the Czech Republic,

grant No. ME 677 is gratefully acknowledged. We thank V. Spichak and an anonymous reviewer for helpful comments.

REFERENCES

- CONSTABLE, S.C., R.L. PARKER and C.G. CONSTABLE (1987): Occam's inversion: a practical algorithm for generating smooth models from EM sounding data, *Geophysics*, **52**, 289-300.
- GELMAN, A., J.B. CARLIN, H.S. STERN and D.B. RUBIN (1995): *Bayesian Data Analysis* (Chapman & Hall, New York), pp. 552.
- 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.
- JONES, A.G. and R. HUTTON (1979a): A multi-station magnetotelluric study in Southern Scotland, I. Fieldwork, data analysis and results, *Geophys. J. R. Astron. Soc.*, **56**, 329-349.
- JONES, A.G. and R. HUTTON (1979b): A multi-station magnetotelluric study in Southern Scotland, II. Monte-Carlo inversion of the data and its geophysical and tectonic implication, *Geophys. J. R. Astron. Soc.*, **56**, 351-358.
- MALINVERNO, A. (2002): Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**, 675-688.
- MEJU, M.A. and V.R.S. HUTTON (1992): Iterative most-squares inversion: application to magnetotelluric data, *Geophys. J. Int.*, **108**, 758-766.
- MENKE, W. (1989): *Geophysical Data Analysis: Discrete Inverse Theory* (Academic Press, London), 2nd edition, pp. 289.
- MONTAZ, A., A. TORN and S. VIITANEN (1997): A numerical comparison of some modified controlled random search algorithms, *TUCS Technical Report No. 98*.
- MOSEGAARD, K. and A. TARANTOLA (1995): Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, **100**, 12431-12447.
- PORTNIAGUINE, O. and M.S. ZHDANOV (1999): Focusing geophysical inversion images, *Geophysics*, **64**, 874-887.
- POUS, J., X. LANA and A.M. CORREIG (1985): Generation of Earth stratified models compatible with both ellipticity and phase velocity observations of Rayleigh waves, *Pure Appl. Geophys.*, **123**, 870-881.
- PRICE, W.L. (1977): A controlled random search procedure for global optimization, *Computer J.*, **20**, 367-370.
- SAMBRIDGE, M. (1999a): Geophysical inversion with a neighbourhood algorithm, I. Searching a parameter space, *Geophys. J. Int.*, **138**, 479-494.
- SAMBRIDGE, M. (1999b): Geophysical inversion with neighbourhood algorithm, II. Appraising the ensemble, *Geophys. J. Int.*, **138**, 727-746.
- SENN, M.K. and P.L. STOFFA (1995): Global optimization methods in geophysical inversion, *Advances in Exploration Geophysics* (Elsevier, Amsterdam), **4**, pp. 294.
- TVRDÍK, J., L. MIŠÍK and I. KŘIVÝ (2002): Competing heuristics in evolutionary algorithms, in *Intelligent Technologies: from Theory to Applications*, edited by P. SINCÁK, V. KVASNICKA, J. VASCÁK and J. POSPÍCHAL (IOS Press, Amsterdam), 159-165.
- WEAVER, J.T. and A.K. AGARWAL (1993): Automatic 1D inversion of magnetotelluric data by the method of modelling, *Geophys. J. Int.*, **112**, 115-123.