Stochastic methods for model assessment of airborne frequency-domain electromagnetic data

Burke J. Minsley
U.S. Geological Survey
Denver, Colorado, USA
bminsley@usgs.gov

James Irving
University of Lausanne
Lausanne, Switzerland
james.irving@unil.ch

Jared D. Abraham
U.S. Geological Survey
Denver, Colorado, USA
jdabraha@usgs.gov

Bruce D. Smith
U.S. Geological Survey
Denver, Colorado, USA
bdsmith@usgs.gov

SUMMARY

Bayesian Markov chain Monte Carlo (MCMC) algorithms are introduced for the analysis of one- and two-dimensional airborne frequency-domain electromagnetic datasets. Substantial information about parameter uncertainty, non-uniqueness, correlation, and depth of investigation are revealed from the MCMC analysis that cannot be obtained using traditional least-squares methods.

In the one-dimensional analysis, a trans-dimensional algorithm allows the number of layers to be unknown, implicitly favouring models with fewer layers. Assessment of data errors and systematic instrumentation errors can also be incorporated. An example from western Nebraska shows that the MCMC analysis reveals important details about the subsurface that are not identified using a single 'best-fit' model.

A geostatistical facies-based parameterization is introduced in order to reduce the number of underlying parameters for the two-dimensional MCMC analysis. This parameterization naturally incorporates lateral constraints in the proposed models, which is important for efficiently sampling the model space. A trans-dimensional component can be optionally incorporated in the two-dimensional algorithm by allowing the number of facies to vary, but with models that contain fewer facies implicitly favoured.

Key words: Bayesian MCMC, airborne electromagnetics, geostatistics, uncertainty, hydrogeophysics.

INTRODUCTION

Several airborne frequency-domain electromagnetic (FDEM) surveys have recently been acquired by the U.S. Geological survey to inform groundwater models by delineating the geometry of hydrogeologic structures in the subsurface (Ball et al., 2011; Smith et al., 2010). While traditional 'best-fit' models derived from these data provide useful details about subsurface structures, they do not capture the full range of plausible solutions that are consistent with the measured data, often leading to incomplete or inaccurate interpretations. A robust assessment of model uncertainty is a critical part of any parameter estimation problem (Tarantola and Valette, 1982). Understanding parameter uncertainty, non-uniqueness, and correlation is just as important as estimating parameter values themselves when interpreting the constraints on plausible solutions provided by a measured dataset.

Recent applications of Bayesian Markov chain Monte Carlo (MCMC) methods to geophysical parameter estimation problems have incorporated a trans-dimensional aspect, wherein the number of unknown parameters is one of the unknowns (Malinverno, 2002; Sambridge et al., 2006; Hopcroft et al., 2007; Bodin and Sambridge, 2009). The trans-dimensional approach provides substantial flexibility in model parameterization, and allows for natural parsimony by favouring solutions that fit the data with the fewest number of parameters. Chosing the model parameterization with MCMC methods is an especially challenging issue when assessing heterogeneous distributions of geophysical properties in two- or three-dimensions, in that a balance must be obtained between forward model complexity and computational expense. In an attempt to deal with this issue, Irving and Singha (2010) proposed a strategy where complex structures were assessed using a geostatistical facies-based parameterization. In their case, the underlying continuously varying subsurface model was approximated by a two-facies system of block pixels with constant geophysical properties within each facies, in order to significantly reduce dimensionality of the problem.

Here, we adapt several existing MCMC approaches to address model assessment and uncertainty analysis of airborne frequency domain electromagnetic (FDEM) data. For one-dimensional (1D) soundings, a trans-dimensional MCMC algorithm is used to solve for the distribution of resistivity values at depth (Minsley, 2011). Additionally, we incorporate flexibility in the algorithm to assess any of several optional parameters, including system elevation, the level of random noise in the data, and systematic instrument errors, all of which can alter the uncertainty in resistivity values with depth. For two-dimensional (2D) models, we adapt the geostatistical basis approach of Irving and Singha (2010). Heterogeneous resistivity models are generated using a sequential simulation algorithm (Deutsch and Journel, 1997) that perturbs a smaller set of primary geostatistical parameters.

METHODS AND RESULTS

One-dimensional algorithm

A trans-dimensional Bayesian MCMC algorithm for the assessment of individual FDEM soundings is presented in Minsley (2011), which builds on the work by Malinverno (2002). In addition to generating realizations from the posterior distribution of resistivity with depth, this algorithm is modified to also assess uncertainty in the system elevation, the level of random noise in the data, and systematic instrument errors. Because of the computational expense involved with a comprehensive MCMC analysis, this algorithm is currently being utilized to assess uncertainty in several characteristic regions of a larger airborne survey.

Figure 1 shows a general flowchart for the trans-dimensional MCMC algorithm for 1D soundings. A layered earth model is initialized with the best-fitting half-space model, with an interface in the middle of the model (layers on both sides of the interface have the same resistivity). A new resistivity model is then proposed in two steps: (1) layer geometry is perturbed by adding a layer, deleting a layer, perturbing one layer interface depth, or leaving the interfaces unchanged, and (2) resistivity values within each new layer are perturbed, with variance proportional to the linearized posterior covariance of the current model (Minsley, 2011). Next, any of the optional parameters describing the system or data errors can be updated.

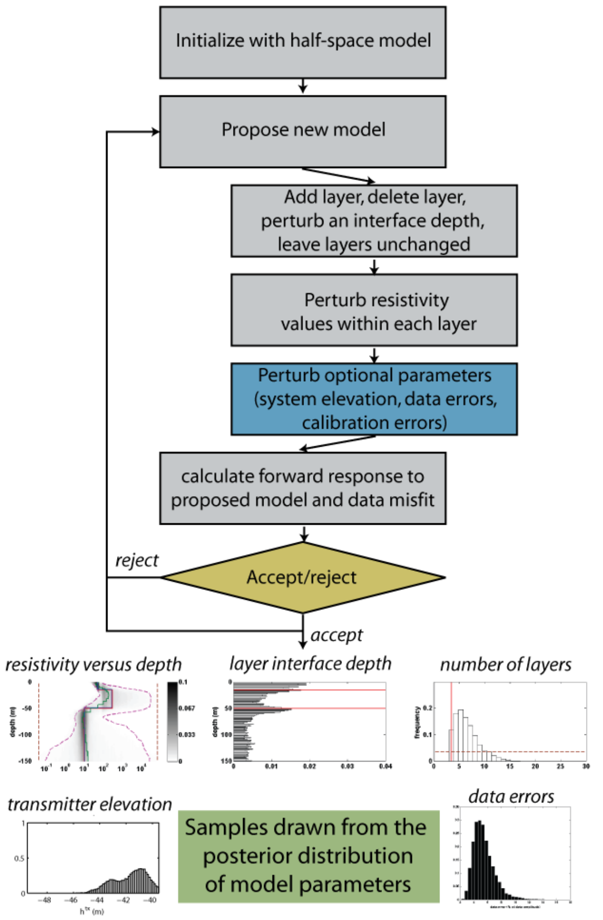

Figure 1. Flowchart for trans-dimensional Bayesian MCMC algorithm applied to 1D soundings (top), with examples of resulting parameter values drawn from the posterior distribution (bottom). Optional steps are shaded in blue.

Once the new model is proposed, its forward response is calculated and is compared with the measured data assuming either a Gaussian or Laplace distribution of errors. The prior probability of each model is calculated by assuming a Gaussian distribution that favours models with small vertical gradients in resistivity. Models are accepted or rejected according to the Metropolis-Hastings criterion (Metropolis et al., 1953; Hastings, 1970), and the algorithm is repeated either for a fixed number of MCMC iterations or until some convergence conditions are met.

A wealth of information about model uncertainty, correlation, and non-uniqueness can be obtained by assessing the ensemble of accepted MCMC models. Examples shown at the bottom of Figure 1 include (clockwise from top left): 2D histogram of resistivity versus depth, distribution of likely interface depths, distribution of number of layers in the model, distribution of data errors (as a percentage of amplitude at each frequency), and the distribution of plausible transmitter elevations.

An example of this algorithm applied to a field dataset acquired in western Nebraska, USA (Smith et al., 2010), is illustrated in Figure 2. Airborne FDEM data were acquired over a borehole where stratigraphic (Figure 2B) and downhole resistivity (Figure 2A, yellow curves) data were logged. Elevated resistivity in the borehole log coincides with gravels, which comprise the principal aquifer in this area, while decreased resistivity is observed in the silts and clays.

A traditional least-squares ‘best fit’ model derived from the airborne data at this location (Figure 2A, green curve) is superimposed on the shaded histogram of MCMC models, along with the most-probable MCMC model (Figure 2A, blue curve), and the region that contains 95 per cent of the MCMC models (Figure 2A, magenta curves).

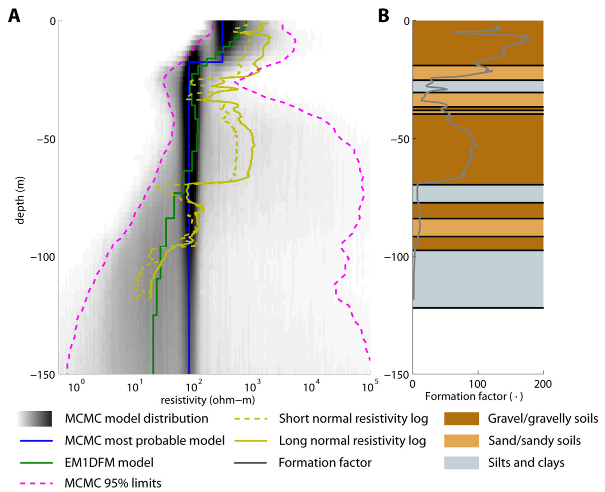

Figure 2. (A) Distribution of MCMC models compared with short-and long-normal resistivity log data (yellow) from a borehole collocated with the airborne survey, and the traditional least-squares inversion result (green). (B) Stratigraphic information from the borehole.

While the traditional least-squares result in Figure 2A captures the general trend of the resistivity logs, it does not appear to be sensitive to the resistive layer between ~40 – 75 m depth. This discrepancy leads to two possible conclusions: (1) the airborne data are insensitive to this layer, or (2) the data are inaccurate and inconsistent with the downhole logs. Analysis of the MCMC results, however, provides a more complete picture. Although not the most probable solution, there are many plausible solutions consistent with the data and prior information that do include this resistive layer, as evidenced by the darker shading (more models) at elevated resistivity values. In addition, the distribution of MCMC models provides a more robust assessment of uncertainty on the depth to the transition to lower resistivity values ~ 100 m depth, as well as bounds on the likely values of resistivity.

Two-dimensional algorithm

Airborne data are densely sampled along multiple long (typically 20 km or more) survey lines, which facilitates a 2D analysis of the data. In order to make the MCMC approach tractable for 2D problems, a geostatistical parameterization is used to reduce the number of underlying parameters. Here, these underlying parameters include the resistivity mean and variance for each facies, as well as the vertical and horizontal correlation lengths, although any of these parameters can remain fixed to simplify the problem. In addition, the trans-dimensional approach discussed for the 1D case can be implemented by allowing the number of facies to vary for each proposed model. By proposing new models using a geostatistical basis, lateral continuity along the profile is naturally enforced.

In the 2D case, the Markov chain is initialized using information derived from traditional least-squares inversion of the data (Figure 3), which reduces the time needed to begin sampling relatively high-probability regions of the model space. An initial facies model is defined by thresholding the least-squares result, selecting a mean resistivity value for each facies, and defining vertical and horizontal correlation lengths.

New models are proposed in several steps, though this process can be simplified by keeping some optional parameters fixed in order to reduce the number of unknowns (Figure 3). In the first (optional) step, the number of facies and/or facies properties are perturbed. Allowing the number of facies to vary incorporates the trans-dimensional concepts applied to the 1D problem, where models with fewer facies are implicitly favoured. The properties of each facies- specifically, the resistivity mean and variance, and correlation lengths, can also be optionally perturbed.

Next, a new distributed facies model is proposed by running a sequential indicator simulation (Deutsch and Journel, 1997) using the previously defined number of facies and correlation lengths. In addition, the indicator simulation is constrained by a fixed percentage of values from the previous model, which ensures that the perturbed model remains close to the previous model. The number and locations of values that remain fixed can be chosen to achieve a desired MCMC acceptance rate, and can also be selected according to locations in the model with acceptable data fit.

In the final proposal step, fixed resistivity values can be assigned to each facies, or (optionally) a sequential Gaussian simulation (Deutsch and Journel, 1997) can be used to incorporate heterogeneity within each facies according to the previously defined resistivity mean and variance. Again, the sequential simulation is constrained by the same fixed values from the previous model in order to ensure modest changes in the proposed model.

Once resistivity values are assigned, the forward response is calculated and compared with the measured data. The forward response can be calculated in 1D at each location in the model, or using 2D or 3D algorithms that account for lateral sensitivity of the airborne system. In either case, this step is easily parallelized, with each forward calculation assigned to a different processor. Because the system footprint is small (up to several hundred meters) compared with the profile length (tens of kilometres), even 2D or 3D forward algorithms can be run in parallel using different segments of the profile (Cox and Zhdanov, 2008).

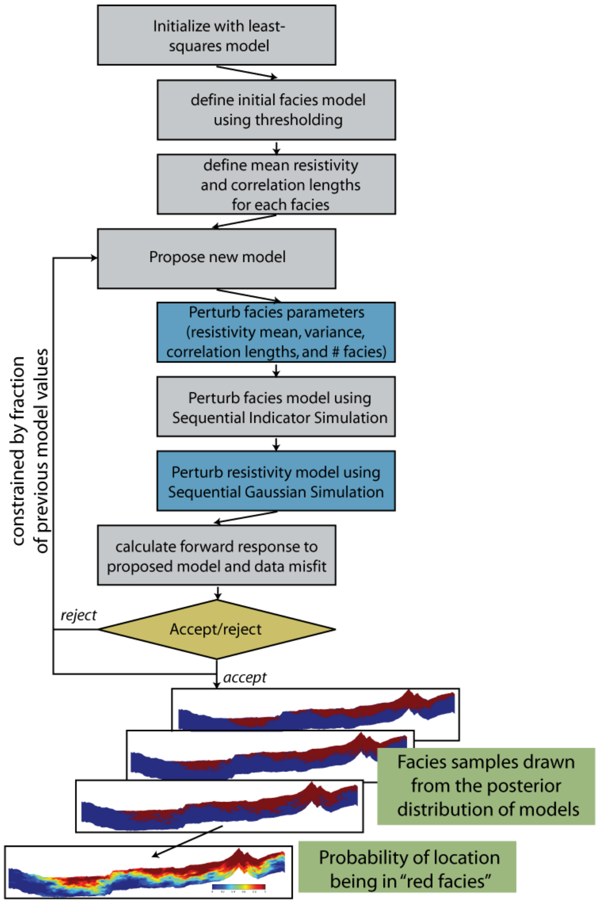

Figure 3. Flowchart for 2D Bayesian MCMC algorithm (top), with examples of resulting facies models drawn from the posterior distribution (bottom) that can be used to produce secondary products such as the probability of any location falling within a particular facies. Optional steps are shaded in blue.

Models are accepted or rejected according to the Metropolis-Hastings criterion (Metropolis et al., 1953; Hastings, 1970), and the algorithm is repeated either for a fixed number of MCMC iterations or until some convergence conditions are met. Given the limited lateral sensitivity of the data, portions of the model with acceptable data fit can be accepted and regions with poor fit can be rejected. A fraction of the values from the accepted parts of the model are then used to constrain the next MCMC sample.

The ensemble of accepted MCMC models drawn from the posterior Bayesian distribution now include 2D facies-based resistivity models (Figure 3, bottom) and (optionally) distributions of the resistivity mean and variance for each facies as well as the number of facies. From this distribution, inferences can be made about the probability of any given location falling in a particular facies (Figure 3, bottom).

CONCLUSIONS

Bayesian MCMC methods provide a useful framework for model assessment and uncertainty analysis of airborne frequency-domain electromagnetic data. Incorporation of a trans-dimensional algorithm, where the number of unknown parameters is allowed to vary, provides substantial flexibility in model parameterization. A 1D analysis can be effectively utilized to investigate uncertainty in different characteristic regions of a survey area, but is limited to small portions of the dataset. We introduce a 2D approach to the MCMC analysis that uses a geostatistical parameterization in order to reduce the number of underlying parameters. This parameterization naturally incorporates lateral constraints in the proposed models, which cannot be accomplished with individual 1D soundings. Calculating the forward response to a proposed model, which is the most computationally expensive part of the algorithm, is readily parallelized because of the limited footprint of the airborne system.

ACKNOWLEDGMENTS

Funding for this work was provided by the U.S. Geological Survey Mineral Resources Program, though the Integrated Methods Development Project led by Jeff Phillips. The Nebraska Environmental Trust Fund and the North and South Platte Natural Resource Districts provided support for this work.

REFERENCES

Ball, L.B., Smith, B.D., Minsley, B.J., Abraham, J.D., Voss, C.I., Deszcz-Pan, Maria, and Cannia, J.C., 2011, Airborne electromagnetic and magnetic survey data of the Yukon Flats and Ft. Wainwright areas, central Alaska, June 2010: U.S. Geological Survey Open-File Report, in review,.

Bodin, T., and Sambridge, M., 2009, Seismic tomography with the reversible jump algorithm: Geophysical Journal International, v. 178, no. 3, p. 1411-1436.

Cox, L.H., and Zhdanov, M.S., 2008, Advanced computational methods of rapid and rigorous 3-D inversion of airborne electromagnetic data: Communications in Computational Physics, v. 3, no. 1, p. 160-179.

Deutsch, C.V., and Journel, A.G., 1997, GSLIB: Geostatistical Software Library and User's Guide: Oxford University Press, USA.

Hastings, W.K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, v. 57, no. 1, p. 97-109.

Hopcroft, P.O., Gallagher, K., and Pain, C.C., 2007, Inference of past climate from borehole temperature data using Bayesian Reversible Jump Markov chain Monte Carlo: Geophysical Journal International, v. 171, no. 3, p. 1430-1439.

Irving, J., and Singha, K., 2010, Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities: Water Resources Research, v. 46, p. 16 PP.

Malinverno, A., 2002, Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem: Geophysical Journal International, v. 151, no. 3, p. 675-688.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Equation of State Calculations by Fast Computing Machines: The Journal of Chemical Physics, v. 21, no. 6, p. 1087-1092.

Minsley, B.J., 2011, A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data: Geophysical Journal International, v. 187, no. 1, p. 21.

Sambridge, M., Gallagher, K., Jackson, A., and Rickwood, P., 2006, Trans-dimensional inverse problems, model comparison and the evidence: Geophysical Journal International, v. 167, no. 2, p. 528-542.

Smith, B.D., Abraham, J.D., Cannia, J.C., Minsley, B.J., Deszcz-Pan, M., and Ball, L.B., 2010, Helicopter Electromagnetic and Magnetic Geophysical Survey Data, Portions of the North Platte and South Platte Natural Resources Districts, Western Nebraska, May 2009: U.S. Geological Survey Open-File Report 2010-1259.

Tarantola, A., and Valette, B., 1982, Inverse Problems = Quest for Information: Geophysics, v. 50, p. 159-170.