Transdimensional Monte Carlo Inversion of AEM Data

Ross C Brodie
Geoscience Australia
GPO Box 378
Canberra, ACT 2601
ross.c.brodie@ga.gov.au

Malcolm Sambridge
Research School of Earth Sciences
Australian National University
Canberra, ACT 0200
malcolm.sambridge@anu.edu.au

SUMMARY

A new approach for the 1D inversion of AEM data has been developed. We use a reversible jump Markov Chain Monte Carlo method to perform Bayesian inference. The Earth is partitioned by a variable number of non-overlapping cells defined by a 1D Voronoi tessellation. A cell is equivalent to a layer in conventional AEM inversion and has a corresponding conductivity value. The number and the position of the cells defining the geometry of the structure with depth, as well as their conductivities, are unknowns in the inversion.

The inversion is carried out with a fully non-linear parameter search method based on a transdimensional Markov chain. Many conductivity models, with variable numbers of layers, are generated via the Markov chain and information is extracted from the ensemble as a whole. The variability of the individual models in the ensemble represents the posterior distribution. Spatially averaging results is a form of 'data-driven' smoothing, without the need to impose a specific number of layers, an explicit smoothing function, or choose regularization parameters. The ensemble can also be examined to ascertain the most probable depths of the layer interfaces in the vertical structure.

The method is demonstrated with synthetic time-domain AEM data. The results show that an attractive feature of this method over conventional approaches is that rigorous information about the non-uniqueness and uncertainty of the solution is obtained. We also conclude that the method will also have utility for AEM system selection and investigation of calibration problems.

Key words: transdimensional, AEM, Bayesian, inversion, Monte Carlo.

INTRODUCTION

The majority of existing methods used for the inversion of airborne electromagnetic (AEM) data use what are generally called gradient-based optimization techniques. They typically minimize an objective function comprised of data misfit (e.g. least squares) and model regularization (e.g. roughness) terms. Since the problem is non-linear, an iterative search involving the matrix solution of equations, linearized about the current model, is performed. The final solution is a single model, that fits the data within the noise levels but also conforms as closely as possible to the constraints imposed by the model regularization.

Due to non-uniqueness and data errors, the single model is just one of an infinite suite of models that could possibly fit the data within the noise levels. On its own, the single solution provides no information about the non-uniqueness or uncertainty in the solution. The lack of uncertainty information is widely recognized as a drawback of the single-solution gradient-based inversions. Some methods make use of the posterior model covariance matrix (Menke, 1989) to estimate model parameter uncertainties. However, strictly speaking, such estimates are accurate only for linear problems and they cannot take account of the non-linearity or non-uniqueness of the AEM inversion problem. They also often reflect the particular choice of regularization parameters.

We present an example of a new approach to the 1D inversion of time-domain AEM data that provides not only a best fit model, but also a wealth of information about the uncertainty and non-uniqueness. We use a reversible jump Markov Chain Monte Carlo (RJ-McMC) method to perform Bayesian inference. Simulated annealing, another Monte Carlo method, was used by Yin and Hodges (2007) for AEM inversion. Rather than addressing uncertainty, their motivation was to avoid the solution being trapped in a local minima and being sensitive to the starting model. Our approach is however very similar in both motivation and method to just recently published work by Minsley (2011). Minsley demonstrates the extent of uncertainty information that can be attained from frequency-domain AEM data, and his method would apply equally well to time-domain data.

METHOD

This work has evolved from applications developed for seismic tomography by Bodin and Sambridge (2009) and Bodin *et al*. (2009), to which the reader should refer for a complete mathematical description. Here we only give an outline and a synthetic example to show how it has been adapted for AEM.

The Earth is parameterized by a variable number of non-overlapping cells defined by a 1D Voronoi tessellation. A cell is equivalent to a layer in conventional AEM inversion and has a corresponding conductivity value. The layer interfaces are positioned midway between the Voronoi cell nuclei, thereby defining the layer thicknesses and depths. We may choose to work in linear or logarithmic depth and/or conductivity units. The number of cells n, the cell conductivities c_k, and the depth position of their nuclei z_k, are all unknowns in the inversion. The parameters are confined to finite ranges (i.e. $n_{min} \le n \le n_{max}$, $c_{min} \le c_k \le c_{max}$, and $z_{min} \le z_k \le z_{max}$).

The RJ-McMC sampling is carried out on multiple independent Markov chains in parallel. The sampling for each chain begins by initialising with a random model. Its number of layers is randomly chosen, from the allowable number range, and with uniform probability. Then, in-turn, the cell conductivities and nuclei depths are similarly set. Once initialised, the main sampling loop begins generating models.

In each loop a new model ($\boldsymbol{m}'$) is proposed by altering the current model ($\boldsymbol{m}$) using one of the following propositions; (i) value-change, (ii) nucleus-move, (iii) birth, or (iv) death. The value change proposition occurs on every second loop, and on alternate loops, with equal probability, one of the other three propositions is used. For the value change proposition a cell is randomly chosen and its conductivity (c) is perturbed to a new value (c') according to a Gaussian proposal distribution with standard deviation σ_c (i.e. $c' = c+\eta\times\sigma_c$, where η is a Normal random deviate). For the move proposition a nucleus from $\mathbf{m}$ is randomly chosen and its original depth (z) is perturbed according to a Gaussian proposal distribution with standard deviation σ_m (i.e. $z' = z + \eta\times\sigma_m$).

In the birth proposition a new nucleus is inserted at a random position in the allowable depth range. Its conductivity is assigned by a perturbation to the conductivity of the original cell (c) in the current model $\mathbf{m}$, at the proposed new nucleus depth, according to a Gaussian proposal distribution with standard deviation σ_{bd} (i.e. $c' = c + \eta\times\sigma_{bd}$). The death proposition is the exact opposite of the birth step.

Once $\boldsymbol{m}'$ is generated it is tested against the bounds to make sure it is a permissible model. If not it is immediately rejected. If $\boldsymbol{m}'$ is permissible, a layered earth is constructed (and converted to linear-space if necessary), and the forward model response $\boldsymbol{g}(\boldsymbol{m}')$ is calculated. The data misfit $\Phi(\boldsymbol{m}') = \sum[(d_k - g_k(\boldsymbol{m}'))/e_k)^2]$, is computed, where $\boldsymbol{d}$ and $\boldsymbol{e}$ are the observed data and errors respectively. Then the likelihood of the data given the model $p(\boldsymbol{d}|\boldsymbol{m}') = exp\{-0.5\times\Phi(\boldsymbol{m}')\}$ is calculated.

A proposed model may be accepted into, or rejected from, the Markov chain according to an acceptance probability $\alpha(\boldsymbol{m}'|\boldsymbol{m})$. The acceptance probability is the key to ensuring that the samples will be generated according to the desired target posterior probability density $p(\boldsymbol{m}|\boldsymbol{d})$. Using Bayes' Theorem, it can be shown that the posterior $p(\boldsymbol{m}|\boldsymbol{d}) \propto p(\boldsymbol{m})\times p(\boldsymbol{d}|\boldsymbol{m})$, where $p(\boldsymbol{m})$ is the prior probability of the model $\boldsymbol{m}$. It turns out that the chain will converge if,

$$\alpha(\boldsymbol{m}' \mid \boldsymbol{m}) = \min\left[1, \frac{p(\boldsymbol{m}')}{p(\boldsymbol{m})} \times \frac{p(\boldsymbol{d} \mid \boldsymbol{m}')}{p(\boldsymbol{d} \mid \boldsymbol{m})} \times \frac{q(\boldsymbol{m} \mid \boldsymbol{m}')}{q(\boldsymbol{m}' \mid \boldsymbol{m})} \times |\boldsymbol{J}|\right],$$

where $q(\boldsymbol{m}|\boldsymbol{m}')$ is the proposal distribution and the term $|\boldsymbol{J}|$ is unity for the case at hand (Bodin and Sambridge, 2009). The expressions for $\alpha(\boldsymbol{m}'|\boldsymbol{m})$ differ depending on the type of proposal and are too cumbersome to reproduce here.

However, summarizing from Bodin and Sambridge (2009), we can say that for the value-change and nucleus-move propositions better fitting models are always accepted, and worse fitting models are accepted according to the ratio $p(\boldsymbol{m}'|\boldsymbol{d})/p(\boldsymbol{m}|\boldsymbol{d})$. For the birth and death propositions the acceptance probability is a balance between the proposal probability, which encourages conductivity changes, and the difference in data misfit, which penalizes conductivity changes if they degrade data fit. Also, given similar data fits, a proposed model has more chance of being accepted if $\boldsymbol{m}'$ has fewer layers than $\boldsymbol{m}$, giving the algorithm a form of natural parsimony. The process of accepting or rejecting moves in this way controls the sampling of the Markov chain so that it preferentially samples regions of parameter space with high values of the target probability density, $p(\boldsymbol{m}|\boldsymbol{d})$, or more precisely the density of the chain will asymptotically converge to that of the target density.

Each chain generates N_s models, including the burn-in period N_b that gives time for an acceptable misfit to be achieved. After the burn-in, new models are added into a discretized 2D posterior histogram. That is, for each discrete histogram depth-bin, the model conductivity is determined and the corresponding histogram conductivity-bin count is incremented. This builds up an image representation of the desired probability density function $p(\boldsymbol{m}|\boldsymbol{d})$. Similarly, a 1D changepoint histogram is built up by incrementing all depth-bins of the 1D histogram in which a layer interface falls.

EXAMPLE

We have inverted synthetically generated AEM data from a three-layered earth model. The secondary-field data were synthesized for the 15 X- and Z-component windows of the TEMPEST system (Lane *et al.*, 2000). Noise was not added to the synthetic data, but since it is required in the inversion to calculate the data misfit (Φ_d), it was estimated via the noise model of Green and Lane (2003). We used a 3.0% multiplicative noise and typical additive noise estimated from the standard deviation of high-altitude data.

Eight independent Markov chains were used in the inversions. The chains were sampled in parallel on eight CPUs of a cluster computer using the Message Passing Interface standard (Message-Passing-Interface-Forum, 1994). In each chain $N_s=10^6$ samples were acquired. Models were accumulated into the posterior probability and change-point histograms after the burn-in period of $N_b=10^4$ samples.

A minimum of 1 and maximum of 10 partitions were allowed in the partition model. The model was parameterized in linear-depth and base ten logarithmic-conductivity space. The model bounds were 0–200 m in depth and 10^{-4}–5.0 S/m in conductivity. For the storage of the posterior histograms the model space was discretized into 200 depth-cells and 100 conductivity-cells. We set the proposal distributions standard deviations to $\sigma_c=0.3$ logarithmic decades for the conductivity value change proposition, $\sigma_{bd}=1.0$ decade for the birth-death conductivity value change proposition, and $\sigma_m=5.0$ m for the nucleus move proposition.

Given the diminishing sensitivity of AEM with depth, one might question our decision to choose a linear-depth parameterization. However the choice should be based on prior geological expectation rather than sensitivity considerations. Accordingly, since there is typically no reason to presume that geological interfaces are more likely to be closer to the surface than they are likely to be at greater depth, a uniform dependence is appropriate. The choice to use a logarithmic-conductivity parameterization is based on the realization that conductivity is usually distributed logarithmically in nature.

Figure 1 shows a summary of the results for the inversion of the synthetic data. The top left panel shows the convergence of the data misfit (Φ_d) for each of the 8 chains. The horizontal dotted line shows the level of data misfit ($\Phi_d=N_d=30$) that

would indicate the data has been satisfactorily fitted. The vertical dotted line indicates the extent of the burn-in period, after which models are accumulated into the posterior histograms. Convergence was generally achieved by the end of the burn-in.

The top right panel shows a histogram of the number of partitions in the accumulated models. It indicates that a four layer model, one more than the actual number of layers in the synthetic model, is most likely. It also indicates that a one or two layer model is much less likely to be able to explain the data than three or more layers.

The bottom left panel shows the true synthetic three layered model (blue line) and information that summarizes the ensemble of models collected after the burn-in period. The greyscale shading represents the discretized 2D posterior histograms. The darker the shading, the more models (counts) in the accumulated ensemble that had that particular conductivity at that particular depth.

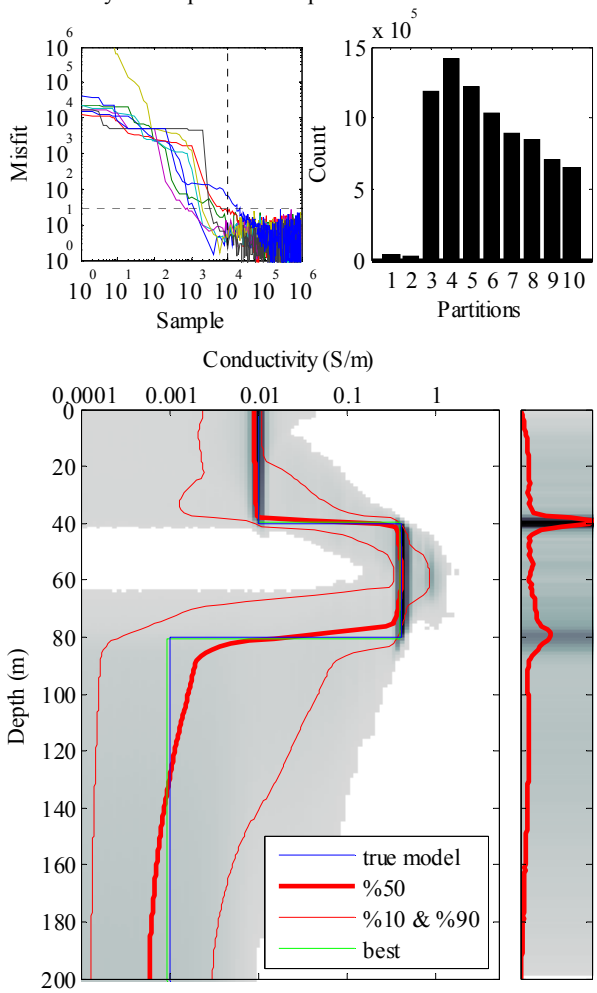

Figure 1. Results summary of the inversion of a synthetic TEMPEST response for a three layered earth model shown by the blue line in the bottom left panel.

To assist in the understanding of the shading, Figure 2 shows rows (profiles) from the histogram at specific depths of 10, 60 and 120 m. The width of the profile peaks can be considered a measure of the certainty. This is analogous to how the standard deviation of a normal distribution can be used to quantify certainty. The shape and skewness of the profiles also contains information.

The narrow peak for the 10 m depth trace (red) indicates that the conductivity at 10 m depth is well resolved. The width of the peak for the 60 m depth trace (green) is suggesting wider range of conductivities is likely at that depth. It peaks around the true conductivity of 0.4 S/m, but also shows that conductivities of up to a little over 1 S/m are quite probable. However the histogram shading around 60 m depth indicates the layer must be thinner if its conductivity is 1 S/m, which is simply a reflection of the well known principle of electrical equivalence (Mallick and Verma, 1979).

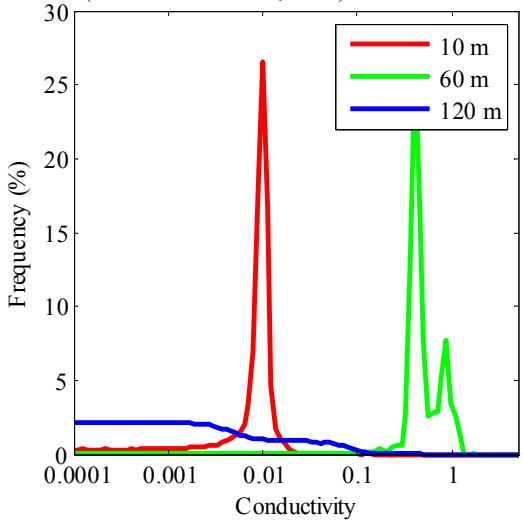

Figure 2. Profiles (or rows) at three different depths from the 2D posterior histogram represented by the greyscale shading in Figure 1.

The lack of a peak at 120 m depth trace (blue) in Figure 2 indicates the conductivity is unresolved at that depth. It rules out a conductive layer (>0.1 S/m) at that depth, since the histogram shading disappears at high conductivity, but discrimination between the lower conductivities is quite poor. This was also the case for 10 m depth. Again, this is really just a reflection of the poor resolving power of AEM, and inductive techniques more generally, at the lower end of the conductivity scale.

Also shown on the bottom left panel of Figure 1 are the best fit model (green) and the 10, 50 and 90 percentiles (red). The best fit model is the model from the accumulated ensemble that had the lowest data misfit. In this case, where noise was not added to the synthetic data, it happens to be very close to the true synthetic model. It is important to understand, that this would not necessarily be the case for real data containing noise.

The percentile models are calculated from the posterior histogram such that, at any given depth, *p*% of the models in the ensemble have conductivities less than the *p*th percentile model. The 50th percentile model could therefore be described as a 'median model'. We would most likely use the 50th percentile model for generating flight-line conductivity-depth sections. The 10th and 90th percentile models provide some indication of the probable lower and upper bounds of the conductivity range.

The distance between the 10th and 90th percentiles, also called the 80% credible interval, gives a measure of (un)certainty in the conductivity estimates. It could also be used to quantify the depth of investigation (DOI). We might

define that the DOI as the depth at which the 80% credible interval is greater than, for example, three decades of conductivity, which would equate to about 80 m for the example above. The choice of credible interval would be somewhat arbitrary, as in the case of other DOI estimation methods.

The bottom right panel of Figure 1 shows the changepoint information (e.g. Gallagher, *et al.*, 2011). The changepoint information is a 1D histogram, represented with both shading and a trace, of the depths at which any layer interfaces occur in the accumulated ensemble of models. It gives a very strong indication in this case of the probable depths of layer boundaries.

DISCUSSION AND CONCLUSION

The method we have demonstrated shows great promise as a method for 1D inversion of AEM data. The advantage of the method over conventional gradient (matrix) based inversion methods is that rigorous information about the non-uniqueness and uncertainty of the solution is obtained. Furthermore, the ability to extract DOI and changepoint parameters is also useful. The method is not subject to starting model sensitivity and instability issues that are sometimes encountered in gradient methods. Another appealing feature is not having to pre-specify the number of layers in the model or the amount of smoothing. This is especially attractive for green-field exploration areas where prior knowledge of the geology may not be adequate. We also foresee potential for this method in detailed studies on comparing the ability of different AEM systems to resolve specific targets, and thus to inform decision on the optimal AEM system to use for a particular survey.

At the time of writing (August 2011) we have not yet fully deployed the method for routine inversion of AEM surveys at Geoscience Australia. However we have inverted every 20th sample (every ~60 m along flight lines) of a complete 8 800 line kilometre VTEM™ survey. This required 70,000 CPU hours, which at a commercial cluster compute rate of $0.12/hour (e.g. Amazon Elastic) would come to a cost of just under $1.00/line kilometre, which is negligible in comparison to per line kilometre AEM data acquisition costs. Our assessment is that this will be a practically viable inversion method for routine deployment in the near future.

Further developments are required to finalize our implementation for routine deployment on typical AEM datasets. For example when inverting fixed-wing AEM data we usually have to invert for the transmitter-receiver offsets, receiver pitch angle and possibly transmitter height (Brodie, 2010). It is intended to include the capability of inverting for these geometric parameters as well (e.g. Minsley, 2011). We also intend to make provision for different kinds of prior conductivity distributions. Further experimentation will also be carried out to determine optimal combinations of samples and number of Markov chains for the most efficient convergence.

ACKNOWLEDGEMENTS

We would like to acknowledge the contribution of Mica Hartley and Thomas Bodin, Research School of Earth Sciences, Australian National University, who were responsible for the original prototyping of this application and are co-authors of our more detailed paper which is currently in preparation.

This computational work was carried out on the National Computational Infrastructure (NCI, http://www.nci.org.au). NCI is supported by the Australian Government through the Department of Innovation, Industry, Science and Research and by a number of major co-investing partner organisations: the Australian National University, CSIRO, the Australian Bureau of Meteorology and Geoscience Australia.

Brodie publishes with the permission of the CEO of Geoscience Australia.

REFERENCES

Bodin, T., and Sambridge, M., 2009, Seismic tomography with the reversible jump algorithm. Geophysical Journal International 178, 1411-36.

Bodin, T., Sambridge, M., and Gallagher, K., 2009, A self-parametrizing partition model approach to tomographic inverse problems. Inverse Problems 25.

Brodie, R.C., 2010, Holistic inversion of airborne electromagnetic data. Ph.D. Thesis, Australian National University (http://hdl.handle.net/1885/49403).

Gallagher, K., Bodin, T., Sambridge, M., Weiss, D., Kylander, M., and Large, D., 2011, Inference of abrupt changes in noisy geochemical records using Bayesian Transdimensional changepoint models. Earth and Planetary Science Letters (*in press*).

Green, A., and Lane, R., 2003, Estimating noise levels in AEM data: 16th Geophysical Conference and Exhibition, Australian Society of Exploration Geophysicists, Extended Abstracts.

Lane, R., Green, A., Golding, C., Owers, M., Pik, P., Plunkett, C., Sattel, D., and Thorn, B., 2000, An example of 3D conductivity mapping using the TEMPEST airborne electromagnetic system. Exploration Geophysics 31, 162-72.

Mallick, K., and Verma, R.K., 1979, Time-domain electromagnetic sounding - computation of multi-layer response and the problem of equivalence in interpretation. Geophysical Prospecting 27, 137-54.

Menke, W., 1989, Geophysical data analysis: discrete inverse theory. Academic Press.

Message-Passing-Interface-Forum, 1994, MPI: A message-passing interface standard. International Journal of Supercomputing Applications 8, 165-416.

Minsley, B.J., 2011, A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data. Geophysical Journal International Early View Online (doi: 10.1111/j.1365-246X.2011.05165.x).

Yin, C., and Hodges, G., 2007, Simulated annealing for airborne EM inversion. Geophysics 72, F189-F195.