ASEG-PESA 2013
'THE EUREKA MOMENT'
11-14 AUGUST 2013 • MELBOURNE, AUSTRALIA

Bootstrapping as a Means of Solution Ensemble Based Uncertainty Analysis in Geophysical Inversion Modelling

Sebastian Schnaidt
Deep Exploration Technologies CRC, The University of Adelaide
The University of Adelaide, SA 5005, AUSTRALIA
sebastian.schnaidt@adelaide.edu.au

Graham Heinson
Deep Exploration Technologies CRC, The University of Adelaide
The University of Adelaide, SA 5005, AUSTRALIA
graham.heinson@adelaide.edu.au

SUMMARY

Many geophysical models are created without satisfactory uncertainty analysis. Most geophysicists are aware of their model's limitations, but if the model is passed on to a third party, this information is lost and the risk of misinterpretation arises.
This project develops multi-solution inversion techniques to improve inversion and joint inversion modelling of geophysical data in mineral exploration. The main focus is the advancement of the probability and uncertainty analysis of inversion models to increase their reliability.
To create solution ensembles, a bootstrapping resampling approach is taken, which produces reduced data sets from a base data set by random omission of data points. Each of these new data sets is run through a conventional inversion process to produce a variety of solutions with minor variations.
In the appraisal stage the solution ensemble is statistically analysed to infer model uncertainties, which are then visualised to allow easy communication of the results. The process yields a clear and easy to interpret uncertainty map for the connected model and we demonstrate its effectiveness with several case studies.
Furthermore, we are currently investigating swarm intelligence based global search algorithms as a second approach to solution ensemble creation.

Key words: solution ensemble, bootstrapping, inversion modelling, uncertainty analysis, magnetotellurics

INTRODUCTION

Geophysical inversion modelling has come a long way since its introduction in the 1960s. Today inversion modelling is a widely used tool to interpret 1-dimensional, 2-dimensional and 3-dimensional geophysical data and inversion models are used in a variety of areas, such as research and the exploration industry. Models are a widely traded commodity, which requires them to be highly reliable to ensure correct interpretation. Highly reliable models lower the risk of the exploration process, thus reducing costs and environmental impact.

The problem is that most of today's inversion schemes produce only a single best fit solution and an uncertainty analysis based on a single model has only limited information value. Even if the person who created the model is aware of its limitations, as soon as the results are handed to a third party, overconfidence in the model bears the risk of misinterpretation.

A way to improve uncertainty analysis is to use multiple solutions to an inversion problem. Some of the first attempts on ensemble creation have been made by Press (1968), as well as Wiggins (1969), but it is only now that those techniques have become feasible as the computing power necessary for the ensemble creation is regularly available. Thus, multi-solution approaches have recently become more widely used, for example by Sambridge & Drijkoningen (1992), Lomax & Snieder (1995), Moorkamp *et al.* (2010).

In this project we develop methods to create solution ensembles and advanced uncertainty analysis techniques based on these ensembles to increase model reliability. The main method we present here is based on bootstrapping resampling.

METHOD

Bootstrapping is a resampling method used in statistics to calculate sample estimates. It was first described by Efron (1979) and is based on statistics calculated from random samples $x^* = (x_1, \dots, x_m)$, repeatedly drawn from a base data set X ($x^* \subset X$). We applied this principle to magnetotelluric (MT) data.

The behaviour of the electric field $\boldsymbol{E}$ and the magnetic induction $\boldsymbol{B}$ involved in MT are governed by diffusion equations (see eq. 1 & 2) (Simpson & Bahr, 2005).

$$\nabla^2 \boldsymbol{E} = \mu_0 \sigma \frac{\partial \boldsymbol{E}}{\partial t} \quad (1)$$

$$\nabla^2 \boldsymbol{B} = \mu_0 \sigma \frac{\partial \boldsymbol{B}}{\partial t} \quad (2)$$

This diffusive nature leads to MT measurements representing averages over the whole volume of medium permeated by the electromagnetic waves. Thus, data from different sites and different sounding periods T contain, depending on the penetration depth p (see eq. 3), overlapping information.

$$p = \sqrt{\frac{T\bar{\rho}}{\pi\mu}} \quad (3)$$

Under this assumption the random omission of some of the data points should not change the result of an inversion, as the information is still contained in the remaining data points. Hence, variations in the results, based on resampled sets of data, give an estimate of the uncertainty inherent in the data.

RESULTS

A first test of the bootstrapping approach was conducted with a data set comprised of broadband MT measurements from 68 sites along the Southern Delamerian transect from Victoria, Australia (Robertson, 2012), with a frequency range of 156.25-0.012 Hz and a total number of 6668 data points (for a complete geophysical and geological interpretation of the data set, see Robertson (2012)).

The first step is to perform a conventional inversion of the entire data set, the so called master inversion. All following inversions are then based on that data set and use the same model grid, as well as the same starting model, as the master inversion. The code used for all inversions was the OCCAM (Constable *et al*., 1987) 2D smooth inversion code.

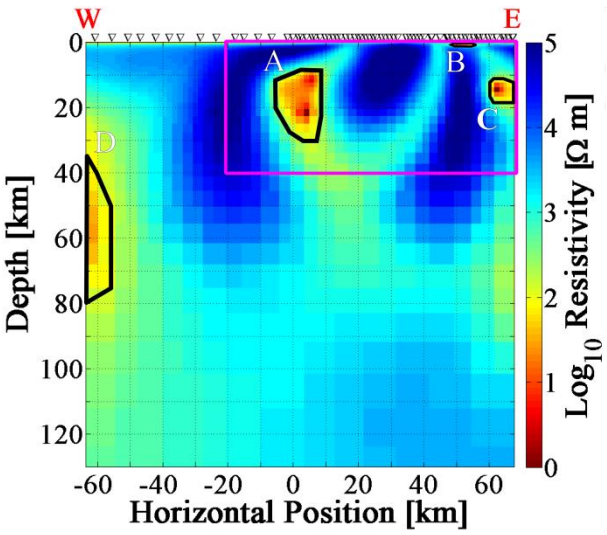

Figure 1: Inversion result of the complete data set of the Southern Delamerian MT transect ($RMS = 1.84$). The main anomalies (A, B, C, D) are delineated in black. A larger version of the zone delineated in magenta is available in Figure 3.

Figure 1 shows the result of the standard inversion of the complete data set. The model features three zones of low resistivity A, B and C in the crust at various depth down to 30 km, as well as a low resistivity zone D at a depth of 50-60 km in the very west of the survey line. All these features exhibit resistivities in the range of 1-50 Ωm.

Appraisal of the Ensemble

So far we have calculated 30 bootstrapped inversion models with a random data omission of 30 % for each model.

The first step in the statistical evaluation of the solution ensemble was to assess the type of value distribution of each model cell to determine a valid set of statistical tools to be used. Tests show that the distributions are relatively close to a normal distribution, but exhibit a non-zero skew as well as a non-zero excess kurtosis, which vary considerably for the different model cells. Therefore, the distributions cannot be considered normal distributions.

Consequently, the commonly used arithmetic mean and associated variance are not valid in this case, and we chose the more general approach of statistical moments.

The first raw moment corresponds to the expected value $\mathrm{E}[X]$ or mean $\bar{x}$ and is, for a discrete random variable X (in our case the model cell resistivities), defined as

$$\mathrm{E}[X] = \bar{x} = \frac{1}{\sum_{i=1}^{n}(p_i)} \sum_{i=1}^{n}(p_i x_i) \tag{4}$$

with the probability mass function $x_1 \rightarrow p_1, \dots, x_n \rightarrow p_n$. Furthermore, the second central moment equates to the variance $\mathrm{Var}(X)$

$$\mathrm{Var}(X) = \frac{1}{\sum_{i=1}^{n}(p_i)} \sum_{i=1}^{n}(p_i (x_i - \bar{x})^2) \tag{5}$$

which defines the standard deviation σ as

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{1}{\sum_{i=1}^{n}(p_i)} \sum_{i=1}^{n}(p_i (x_i - \bar{x})^2)} \tag{6}$$

We estimate the probabilities p_i by classifying all values and determining the proportion of each class. The class width h, and therefore the number of classes, is determined via the Freedman-Diaconis rule (Freedman & Diaconis, 1981)

$$h = 2\frac{\mathrm{IQR}(X)}{n^{1/3}} \tag{7}$$

with the interquartile range $\mathrm{IQR}(X)$ and the number of samples n.

To evaluate the solution ensemble, we calculate the absolute standard deviation σ (see eq. 6) of the resistivity value $log_{10}(\rho)$ for each model cell and then scaled to a relative standard deviation $\sigma_{\mathrm{rel.}} = \sigma/\bar{x}$ to allow for a comparison of the standard deviation of different cells as shown in Figure 2.

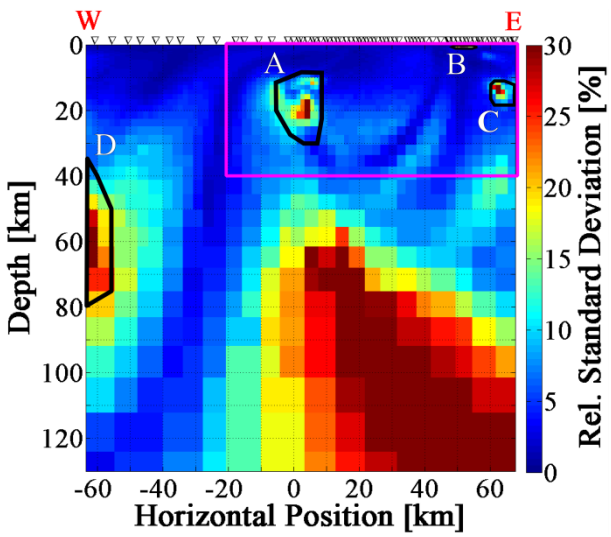

Figure 2: Relative standard deviation $\sigma_{\mathrm{rel.}}$ of the resistivity values $log_{10}(\rho)$ of all model cells of 30 bootstrap models with 30 % data point omission. Scaled to $0-30$ % for higher clarity ($\bar{\sigma}_{\mathrm{rel.}} = 7.22$ %). The main anomalies (A, B, C, D), as identified in Figure 1, are delineated in black. A larger version of the zone delineated in magenta is available in Figure 4.

Figure 2 clearly shows that the areas of highest uncertainty are mostly correlated to areas of low resistivity, with exception of the extended area of elevated uncertainty in the eastern half of the model region, at a depth of 60-130 km.

CONCLUSIONS

First results are very promising. As shown in Figure 2 and 4 the process produces a clear and easy to interpret uncertainty map for the related model (see Figure 1 & 3).

As to be expected, the areas of highest uncertainty are connected to the areas of low resistivity, as low resistivity structures cause a strong attenuation of the electromagnetic fields and, hence, lower the achievable resolution. This effect occurs especially at the lower parts of low resistivity structure, which is clearly picked up by our uncertainty analysis, as can be seen at features A and D. This shielding effect causes the smearing-out of low resistivity regions to greater depth often encountered in inversion modelling.

The extended area of high uncertainty in the eastern half of the model region, is likely due to areas of high resistivity at great depth not contributing as much to the responses measured at the surface, and the measured data containing less information about those regions. Thus, variations in those areas during the modelling process do not affect the modelling response and misfit of the model considerably and these variations are more likely to occur without the inversion code compensating. In addition, the low resistivity structures above that region are likely to add a shielding effect, lowering the achievable resolution in that area even further.

We consider bootstrapping to be superior to the standard sensitivity analysis. A sensitivity analysis only tests the effect of changes to the model parameters on the model response, which generally just highlights areas of low resistivity, as those areas are affecting the model response the most. In contrast, bootstrapping directly tests the impact of the data on the model, and flags areas of the model that are poorly constrained.

The sheer number of models to be calculated requires the use of multi core machines and is the main reason that this approach is currently only feasible for 2D inversions. The general concept of multi solution methods is very much applicable to 3D inversions as well and will yield similarly good results, when rendered possible by the availability of more computing power in the next few years.

Note that, even though we only tested the method on MT data so far, the bootstrapping approach is generally applicable to all geophysical methods that sample volumes, e.g. potential methods like gravity or magnetics.

OUTLOOK

The project is clearly a work in progress. The next step is to test the effect of different omission percentages and different numbers of bootstrap models. In addition, we will test the random omission of whole sites instead of random data points and will conduct a second case study with a different high quality data set and a test using a synthetic data set for more control over the results.

Furthermore, we will expand the idea of solution ensemble based uncertainty analysis to geophysical methods that do not necessarily have to address volumes by developing an inversion (and later joint inversion) code based on metaheuristic global search methods.

ACKNOWLEDGMENTS

The authors thank Kate Robertson for providing us with the necessary data and Lars Krieger for his technical advice. Furthermore, we thank the Deep Exploration Cooperative Research Centre (DET CRC) and its participants for supporting this project.

REFERENCES

Constable, S. C., Parker, R. L., & Constable, C. G., 1987, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, 52, Nr. 3, 289-300.

Efron, B., 1979, Bootstrap methods: Another look at the Jackknife: The Annals of Statistics, 7, Nr. 1, 1-26.

Freedman, D.; Diaconis, P., 1981, On the histogram as a density estimator: L_2 theory, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 57 (4): 453-476.

Robertson, K. E., 2012, An electrical resistivity model of the southeast Australian lithosphere and asthenosphere: Honours thesis, The University of Adelaide.

Lomax, A. & Snieder, R., 1995, Identifying sets of acceptable solutions to non-linear, geophysical inverse problems which have complicated mist functions: Nonlinear Processes in Geophysics 2, 222-227.

Moorkamp, M., Jones, A. G. & Fishwick, S., 2010, Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data: Journal of Geophysical Research 115, B04318.

Sambridge, M. & Drijkoningen, G., 1992, Genetic algorithms in seismic waveform inversion: Geophysical Journal International 109, 323-342.

Simpson, F., & Bahr, K., 2005, Practical magnetotellurics: Cambridge University Press.

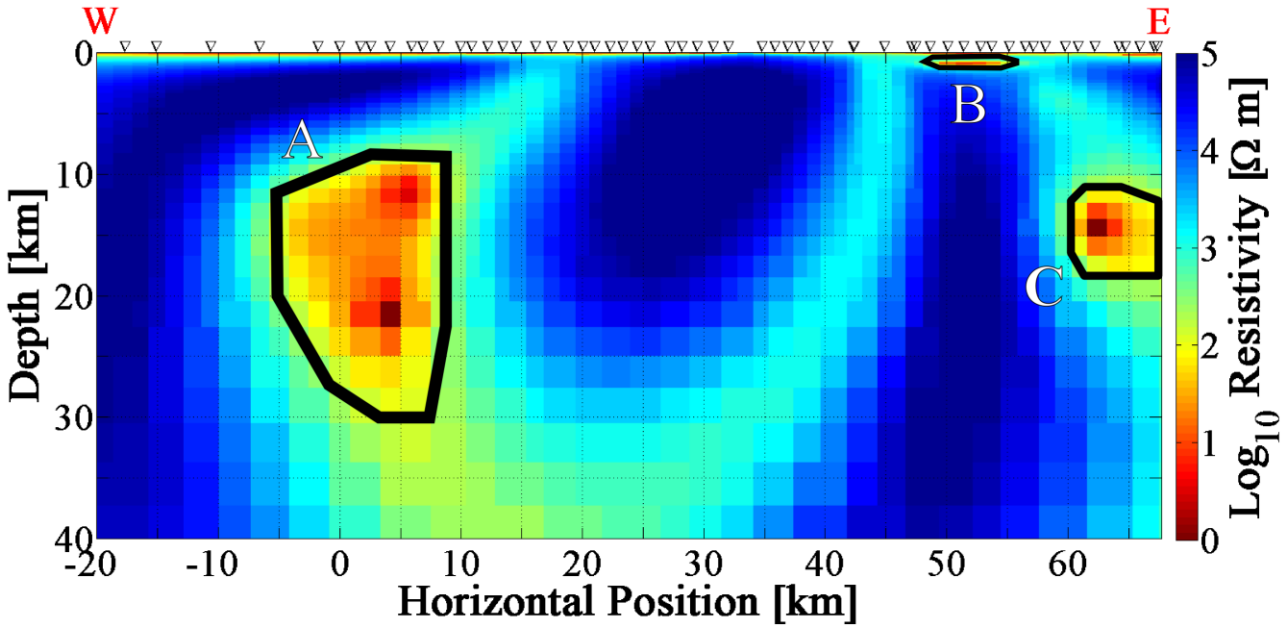

Figure 3: Close up of the region marked in magenta in Figure 1. Inversion result of the complete data set. The main anomalies (A, B, C) are delineated in black.

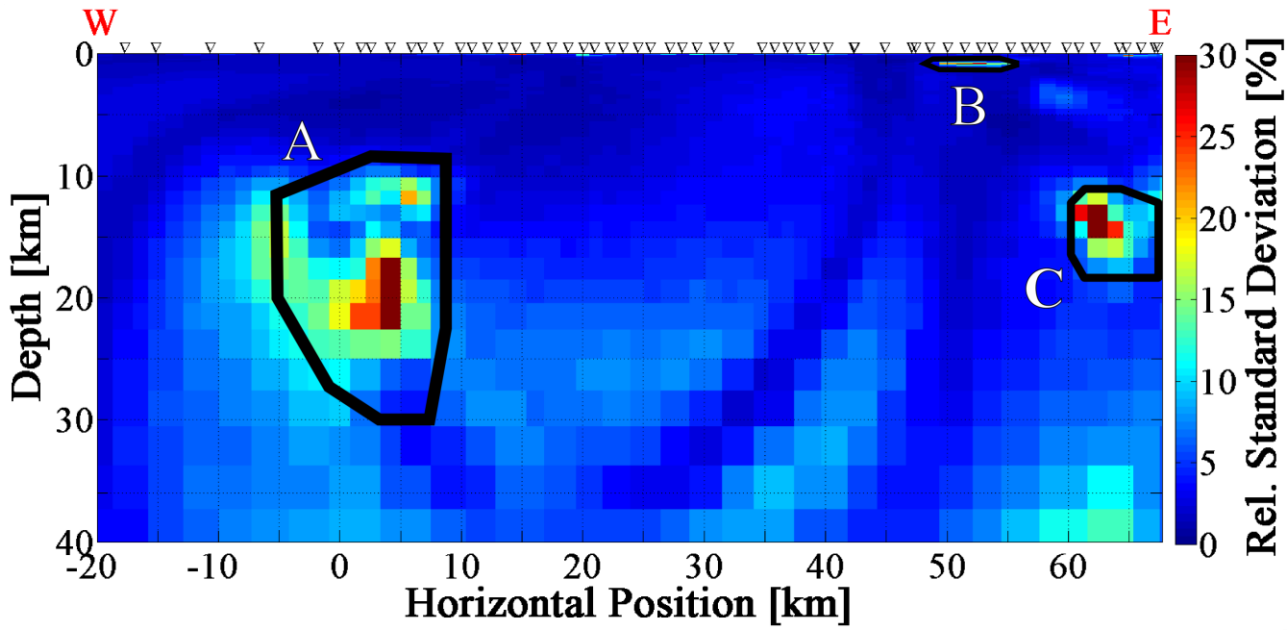

Figure 4: Close up of the region marked in magenta in Figure 2. Relative standard deviation $\sigma_{\text{rel.}}$ of the resistivity values $log_{10}(\rho)$. The main anomalies (A, B, C), as identified in Figure 1, are delineated in black.