Down the borehole but outside the box: innovative approaches to wireline log data interpretation

Anya M. Reading
Sch. Earth Sciences and CODES
Private Bag 79
UTAS, Hobart, TAS, 7001
anya.reading@utas.edu.au

Thomas Bodin
RSES,
ANU, Canberra
ACT, 0200
thomas.bodin@anu.edu.au

Malcolm Sambridge
RSES
ANU, Canberra
ACT, 0200
malcolm.sambridge@anu.edu.au

Stephanie Howe
School of Earth Sciences
Private Bag 79
UTAS, Hobart, TAS, 7001
howes1.1@gmail.com

Michael Roach
Sch. Earth Sciences and CODES
Private Bag 79
UTAS, Hobart, TAS, 7001
michael.roach@utas.edu.au

SUMMARY

Wireline logs record the variation in a number of physical measurements, sometimes 20 or more different properties, with depth down a borehole. They are routinely correlated with, and/or interpreted in terms of, the rock stratigraphic record. Logs are also interpreted with the aim of inferring other useful physical properties not directly measured. In some data-rich exploration industries, such as oil and gas, wireline log interpretation is highly developed. In other industries, such as geothermal and minerals exploration, wireline information is often incomplete and may be inaccurate or inconsistent. The aim of this work is to explore an innovative approach to the analysis of wireline logs.

We use an innovative, flexible approach to the identification of 'change points', which may indicate boundaries between lithologies or significant intra-formation structure. The number of boundaries/classes is not fixed in advance, being solved for as part of the modelling process. The approaches are applicable in data-rich environments with relatively well-know stratigraphy where they will add to physics and more conventional statistics-based inference. They may also find particular utility in situations with less than ideal data and diverse stratigraphy as they naturally incorporate ways of handling uncertainty. Unforeseen relationships are allowed to emerge and, hence, inform future predictive analysis.

Key words: data inference, inversion, petrophysics, thermal properties, wireline log

INTRODUCTION

Many innovative approaches to using data to make inferences about aspects of the natural world necessitate the calculation and appraisal of large numbers of alternative models or the repeated sampling of a multi-parameter solution space. While this is not new to science, it has hitherto been restricted to either very simple models, or to institutions with powerful computing infrastructure. Such institutions include large commercial companies and publicly funded computing initiatives, generally housed at universities or government research establishments. Advances in desktop and notebook computer processing and data storage capabilities now enable a variety of multiple-model approaches to geophysical data inference to be carried out on personal computers. It is therefore timely reconsider such approaches as innovative in recognition of their newly increased potential for widespread use. We also illustrate an innovative algorithm that is highly applicable to many aspects of geoscience.

An example of a successful borehole wireline interpretation program is LogTrans (Zhou and Fullagar, 2009) which uses a statistical characterisation of a representative control dataset to predict lithology from wireline data. Our work aims to add to such practical approaches by using the data itself to suggest change points, without using a control dataset. This may enable further insights to be made in data rich environments. For regions with sparse data and no control datasets, it may enable interpretations to be made which would not have been possible before.

Making an inference from data in a quantitative sense is a significant branch of mathematics often termed 'inverse theory'. Essential definitions, as applied to geophysical data analysis are reviewed in the widely read texts by Menke (1989) and Aster et al. (2005). The latter text makes clear the distinction between the goal of model parameter estimation and that of understanding the full scope of the inverse problem in question. In this paper, we apply the term 'traditional' to data inference problems in which a single model is derived from data, and a reasonable starting model, possibly over several iterations. Of course there are still many applications for which this approach is well-founded and there is certainly scope for innovation in this area (not least in the formation of the forward problem). We do not consider Monte Carlo-style direct search algorithms (e.g. genetic algorithms) in which multiple models are calculated and subject to assessment and refinement, although we also recognise the value of these techniques (Sambridge, 1999). Increased computer power does make these approaches more realistic for models of several tens of parameters and there are certainly some philosophical similarities between some methods of assessing the resulting ensemble of models and the methods described below.

The innovations we focus on in this paper concern the nature of the solution. Instead of there being an unknown model which we would like to determine, *the solution is a probability distribution for the model parameters and the number of model parameters is allowed to vary*. We explore these probability distributions by sampling the multi-dimensional *posterior* model space. This is done using a Monte Carlo approach, with sampling being guided by previous samples using a Markov Chain. The work falls in the category of 'Bayesian' techniques. Bayesian techniques in general have met with criticism and failure in usage in the past owing to the lack of understanding of underlying fundamentals, but used with intelligence, they facilitate extremely well-founded, flexible algorithms which are sufficiently efficient to be of practical use in a typical desk-top computing environment. Importantly, they make use of the information inherent in a well-sampled probability distribution for the model parameters, and may now be viewed as an alternative, with many advantageous features, to traditional forms of geophysical modelling.

One particular advantage is flexibility in the number of parameters describing the model. This can be explored and changed as the model parameter sampling progresses. Making use of such advantages requires working geophysicists to be familiar with a wider range of modelling philosophies than is necessary with commercially available 'traditional' software. Data that have been expensive to acquire can be made use of in a much more complete way. There is often additional insight to be gained through the reinterpretation of existing data in innovative ways. In this paper, we use wireline log data to illustrate some of the advantages of such approaches.

APPLICABILITY

Petroleum exploration has been carried out in on-land regions of Australia for some decades. There is a large amount of existing archive wireline log data residing in the archives of state and federal government sources, much of which is now available online. Many of the regions formally explored for petroleum are now the subject of renewed interest in the form of exploration for various types of geothermal energy generation potential. These include Enhanced Geothermal Systems (heat producing basement rock overlain by thermally insulating sediments) and Hot Sedimentary Aquifers. The geothermal industry in Australia is a yet unproven investment risk. Raising the necessary capital for a comprehensive exploration drilling program is difficult and hence there is much to be gained from making good use of pre-existing data. Junior geothermal exploration companies are now beginning to drill exploratory deep holes, and there is also much to be gained from optimal interpretation of new wireline data for this developing industry. The technique outlined below is not, however, confined to a particular industry and may be applied to any sequence of noisy data where 'change points' which change the underlying relations between the data are suspected. Importantly, the physics, or petrophysics, underlying the change points is not required in calculating the forward problem. It is the statistical relationships between the data that are explored. Change points are then assigned with appropriate probabilities.

Change points in wireline log data may well correspond to lithological boundaries. Importantly for the geothermal industry, they may also indicate regions within lithological units with varying thermal properties but no obvious change in lithology (even if an accurately logged core is available). For example, sandstone containing a large proportion of quartz is likely to be much less thermally conductive than sandstone containing a larger proportion of feldspar. These proportions may change within an apparently uniform formation and would affect the results of thermal modelling in a system dominated by conductive heat flow.

DATA

Wireline log data for this analysis come from two locations in Eastern Australia, Boyne River in Queensland (shown in this paper) and Singleton in New South Wales. They were chosen because they also have a reasonably complete core which was available for sampling and complementary laboratory analysis of petrophysical and rock thermal properties although this is not generally the case with archive wireline log data. Sedimentary cores in particular may be dried out and not suitable for laboratory testing, if indeed they exist at all (Howe, 2009).

The logs used in this example are Density (LSD, counts per sec), Neutron (Porosity, instrumental units), Interval time (DT, us/ft) and Electric Array (EAL, ohm.m). This information is not essential to the method, which normalises the data prior to inversion and converts back to the original scale values for output. Knowledge of the uncertainty in measurement values is not required as this is solved for as part of the inversion. Four logs were used in this case, although typically 12-20 different logs are collected (Dewan, 1984, Serra 1988) and more could be used if required. A 75 m section of core / wireline log was used to illustrate the method. It was simplified considerably by subsampling the wireline values for testing although this is not a general requirement of the method.

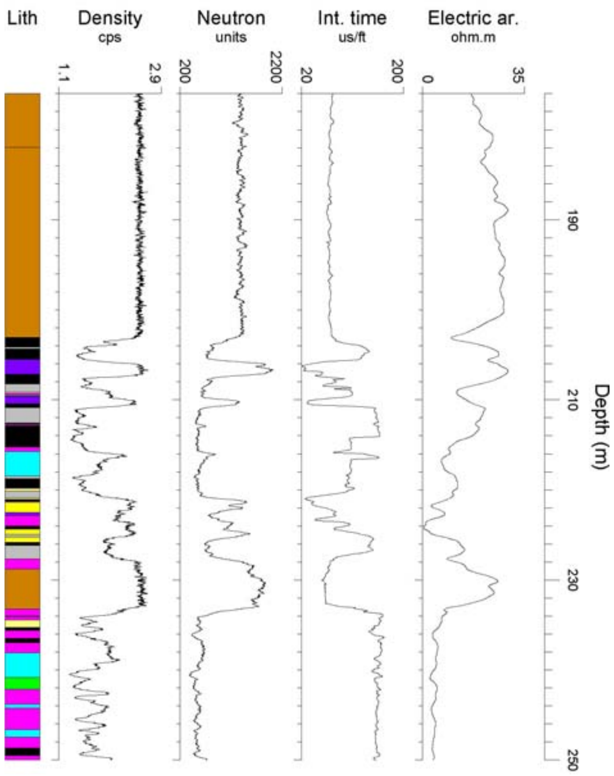

Figure 1. Summary lithology and wireline logs. The thick orange/brown units are dolerite, the other units interbedded shale and conglomerate, including some coal. This is an approximately 75 m long section of a deeper core / wireline log.

METHOD

We investigate 'change points' in the wireline log using a new implementation which does not require prior knowledge of the uncertainty in the data (Bodin et al., 2010). This is a particular advantage in the analysis of wireline data as noted in the previous section.

The reversible jump algorithm (Geyer and Møller, 1994; Green, 1995) represents a Bayesian framework for the construction of reversible Markov chain samplers that jump between parameter subspaces of differing dimensionality. In practical terms, this enables models with differing numbers of change points to be tested and the probability of each change point to be calculated. Such 'transdimensional' inversion methods in geophysics have been used in geostatistics (Stephenson et al. 2004), thermochronology (e.g. Stephenson et al. 2006), palaeoclimate inference (e.g. Hopcroft et al. 2007) and tomographic inversion (Bodin et al., 2009). While this approach is a very promising addition to the geophysicist / well-log analyst's computational toolbox, a problem arises given the lack of knowledge of the level of uncertainty in the data. Allowing the number of change points, model parameters, to vary, effectively allows freedom in the level of complexity in the model. This could be problematic, in that sampling could favour very complex models that are effectively trying to fit noise.

In the new implementation (Bodin et al., 2010), the formulation is improved by parameterising the data uncertainty in the form of one value (a hyperparameter), corresponding to the standard deviation, for each dataset. The algorithm solves for these values in the same way as other model parameters, i.e. providing a posterior distribution for each. This is made possible through a Hierarchical Bayes ('HB') regression formulation.

In general, it is possible to improve the fit of a given model to a dataset by making the model more complicated. At some point, the model starts to fit noise in the data, rather than the underlying features that are the target of the analysis. This is sometimes referred to as 'over-fitting'. In estimating model parameters (or optimization based inversions), the aim is to find a model that minimises the misfit between the data and model. Assuming that the data have normally distributed independent errors, with an expected value of zero, it is common to minimise a misfit function such as the chi-square measure and the most complex model will always be preferred.

Contrary to optimization based inversions, the HB algorithm is able to infer the data noise uncertainty for each data type and to generate models with the required complexity. As data noise increases, the level of acceptable misfit becomes increasingly large and we should accept poorer fitting models. As the estimated noise is decreased, the required data fit becomes more stringent and the Markov chain adds more parameters to provide more closely fitting models. The HB takes into account the lack of knowledge about data errors. Instead of being fixed, the variance of the measurement errors can be assigned a broad prior uncertainty and the posterior distribution sampled accordingly. More details about HB methods and hyperparameter formulations in general may be found in the general text by Gelman et al., (2004) and HB methods are noted in recent overview of uncertainty quantification with geophysical problems in mind by Malinverno and Briggs (2004).

In summary, the new HB implementation that we use allows the inference of a change point structure that is common to all the different input logs, adapts to the complexity defined by the data itself and allows for the uncertainties in each dataset.

RESULTS

The results of HB regression analysis to determine change points in the wireline log data are shown below.

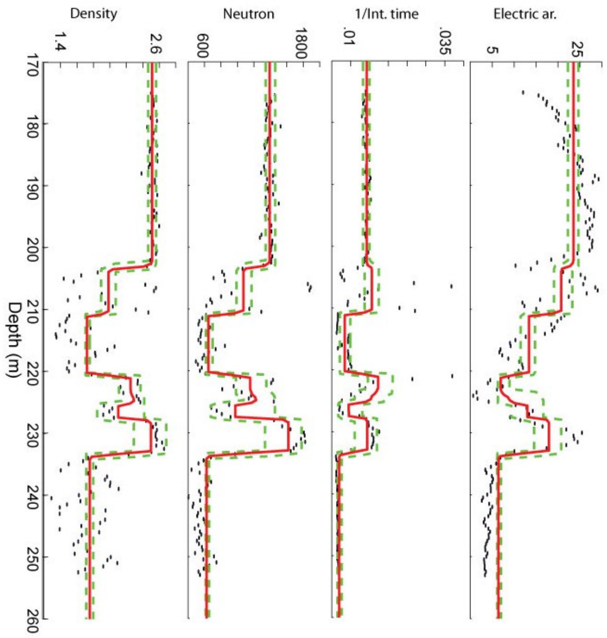

Figure 2. Change points in the test section of wireline log data solved for using the new HB implementation simultaneously over 4 wireline logs. Blue dots = input data, red solid line = final result, green dotted lines = confidence limits on final result. (The over-simplification due to subsampling of the original data is being addressed in ongoing work and is not an inherent limitation of the method.)

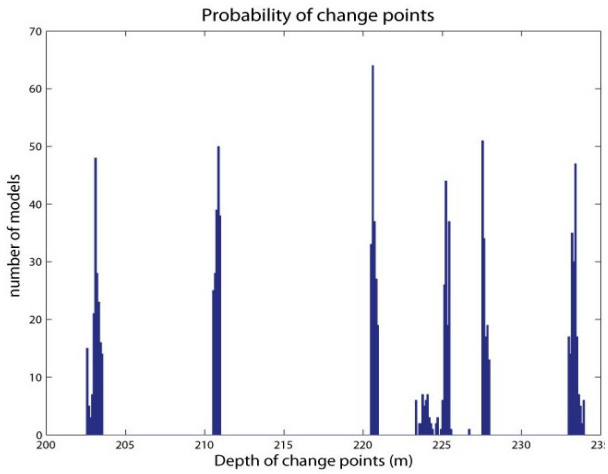

Figure 3. The relative probability of change points existing at different depths is indicated by the number of models showing a change point at that depth.

Figures 2 and 3 show that the HB algorithm has successfully identified change points in the wireline data with no prior information regarding the number of changes in the model or the noise in the data. Figure 3 shows that some of the change points are extremely clear cut (e.g. those at 211 and 221 metres deep), where as others are less well defined (e.g. that at 203 m) or just suggested (around 224 m).

The change points can be used to form a framework for the prediction of other rock properties, such as thermal conductivity, which is the subject of related work.

COMMENT

The classification of inverse problems in geophysics, and the description of approaches to their solution, is becoming less straightforward as greater use is made of algorithms developed in other fields. For example, inverse problems in geophysics are classified as 'linear' or 'nonlinear' according to the relationship between the data and model (i.e. not the form of the function describing the model). This makes good sense as the approaches used to tackle linear or linearised problems form a particular branch of inverse theory (Aster, 2005). Linear approaches are not suitable for nonlinear problems. However, model search algorithms that are suitable for non-linear problems (e.g. Monte Carlo-style approaches such as genetic algorithms) may, in theory at least, be applied to linear inverse problems. Provided sufficient computer power is available, the distinction becomes much less useful. In contrast, material in the text by Denison et al. (2002) uses the term 'nonlinear' to refer to the nature of the function describing the model. The reader wishing to explore different approaches must be aware of this and many other inconsistencies in language usage.

CONCLUSION

Transdimensional inversion methods are an extremely promising area of inverse theory and should find wide use in the near future. The HB implementation illustrated in this example using wireline data should prove particularly appropriate given the frequent lack of knowledge of levels of data noise in geophysical applications.

ACKNOWLEDGMENTS

Data sources include holdings of the Queensland Department of Mines and Energy QDEX system (Boyne River 2C wireline log) and the New South Wales Department of Primary Industries DIGS system (Singleton wireline log and core samples).

Granite Power Ltd is thanked for their collaborative involvement, assistance with core sampling, and in particular their support of Honours student projects at the University of Tasmania including that of S. Howe.

REFERENCES

Aster, R., Borchers, B., & Thurber, C., 2005. Parameter Estimation and Inverse Problems: International Geophysics Series, Elsevier Academic Press, USA.

Bodin, T., Sambridge, M. and Gallagher, K., 2009. A self-parameterising partition model approach to tomographic inverse problems: Inverse Problems, 25, 055009, doi: 10.1088/0266-5611/2/5/055009.

Bodin, T., Sambridge, M. and Gallagher, K., 2010. Accounting for data noise uncertainty in nonlinear regression: Theory and application to palaeoclimate data: Geophysical Journal International (in preparation).

Denison, D.G.T., Holmes, C.C., Mallick, B.K. and Smith, A.F.M., 2002. Bayesian Methods for Nonlinear Classification and Regression: John Wiley and Sons Ltd, England.

Dewan, J.T., 1983. Modern Open-Hole Log Interpretation. Pennwell Books, USA.

Gelman, A., Carlin, J., Stern, H. and Rubin, D., 2004. Bayesian Data Analysis 2nd edition: Chapman and Hall, USA

Geyer C.J. and Møller, J., 1994. Simulation procedures and likelihood inference for spatial point processes: Scandanavian Journal of Statistics, 21, 369-373.

Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Baysian model selection, Biometrika: 82, 711-732.

Hopcroft, P., Gallagher, K. & Pain, C., 2007. Inference of past climate from borehole temperature data using Bayesian Reversible Jump Markov chain Monte Carlo: Geophysical Journal International, 171, 1430–1439.

Howe, S., 2009. Deriving rock thermal properties from wireline well logs: University of Tasmania, School of Earth Sciences, Honours Thesis.

Maliverno, A. and Briggs, V., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and Empiracal Bayes, Geophysics, 69, 1005.

Menke, W., 1989. Geophysical Data Analysis: Discrete Inverse Theory: International Geophysics Series, Elsevier Academic Press.

Sambridge, M., 1999. Geophysical Inversion with a Neighbourhood Algorithm -II. Appraising the ensemble: Geophysical Journal International, 138, 727-746, 1999.

Serra, O., 1988. Fundamental of well-log interpretation. Developments in petroleum science: Elsevier Science Publishers, The Netherlands.

Stephenson, J., Gallagher, K. & Holmes, C., 2004. Beyond kriging: dealing with discontinuous spatial data fields using adaptive prior information and Bayesian partition modelling: Geological Society of London Special Publication 239(1), 195.

Stephenson, J., Gallagher, K. & Holmes, C., 2006. Low temperature thermochronology and strategies for multiple samples 2: partition modelling for 2D/3D distributions with discontinuities: Earth and Planetary Sci. Letters, 241(3-4), 557–570.

Zhou, B. and Fullagar, P., 2009. LogTrans: Automated interpretation of borehole data: Software download from http://fullagargeophysics.com