# Inference from noisy data with an unknown number of discontinuities: ideas from outside the box.


**Malcolm Sambridge**
*Research School of Earth Sciences,*
*Australian National University,*
*ACT 0200*

*Malcolm.Sambridge@anu.edu.au*

**Thomas Bodin**
*Research School of Earth Sciences*
*Australian National University,*
*ACT 0200*

*Thomas.Bodin@anu.edu.au*

**Anya Reading**
*School of Earth Sciences &*
*CODES Centre of*
*Excellence for Ore Deposits*

*Anya.Reading@utas.edu.au*

**Kerry Gallagher**
*Géosciences Rennes,*
*Université de Rennes 1,*
*Rennes, France, 35042.*

*kerry.gallagher@univ-rennes1.fr*



## SUMMARY

The focus of this presentation is on new ways to extract information from complex datasets in situations where direct measurement is not possible. Such inverse problems are ubiquitous across the physical and mathematical sciences and are central to discovery of resources within the Earth upon which Australian society is dependent.

A recurring problem is how to choose the number of unknowns with which to fit noisy data. If too few are chosen the data cannot be fit and if too many the inversion results contain unwarranted detail. Statistical methods are often used to find optimal numbers of unknowns, but these are based on simplistic assumptions and typically require multiple trial inversions to be performed with different numbers of variables. A new general approach recently applied to geophysical problems is to ask the data itself `How many unknowns should be used ?' While this may seem counter-intuitive at first sight it turns out to be entirely feasible. In effect the number of unknowns itself becomes an unknown. An extension of the basic approach also allows the level of noise on the data to also be included as an unknown.

In this presentation we outline the central ideas, and illustrated through an example where a geophysical property varies only in 1-D (usually depth or time) and is constrained from surface measurements. Applications of the general approach are to airborne EM data, borehole geophysics, seismic interpretation and also palaeoclimate reconstructions.



**Key words:** Geophysical inversion methods, disocontinuous fields, seismic, wireline, AEM.


## INTRODUCTION

Exploration geophysics often gives rise to situations where subsurface properties vary spatially. For example in 1-D, this may be a function of depth within the outer layers of the Earth or down a borehole In each case the physical property of interest may vary abruptly due to discontinuities in the medium. An example is airborne EM where flight line data can be inverted as a series of 1-D layered models of subsurface conductivity at each point along a profile which is subsequently stitched together to form a 2-D image. For a full discussion of this problem see Brodie & Sambridge (2006). A well known feature of AEM inversion is the presence of non-uniqueness, i.e. different conductivity models give rise to the same surface response. This is illustrated in Figure 1 where the surface response of a three layer conductivity profile is reproduced by multiple three layer models of conductivity. Reassuringly all three layer models are similar. With 15 layers many profiles are again found that also fit the data, but these bear no resemblance to each other. Moreover, the best fitting model bears no resemblance to the truth. Clearly then knowledge of the number of layers is crucial for a reliable interpretation. Without prior knowledge of the number of unknowns to include how is the practitioner to decide which if any should be the preferred interpretation?

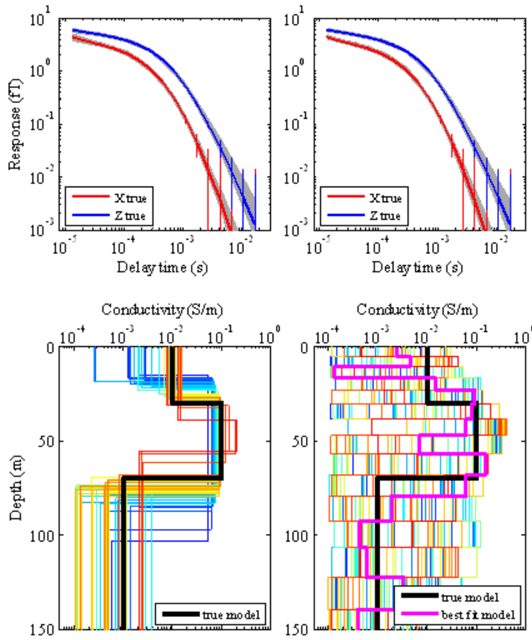


**Figure 1. An illustration of non-uniqueness in the inversion of AEM data: The bottom left panel shows a synthetic three layer model (thick black line) and several other three layer models (coloured lines) all of which fit the data to within observational error. Plotted in the top left panel is the true model's forward response (red and blue curves with error bars) and the forward responses of models, which fit the true model response within noise levels. The right hand panels show the same information, in a parametrization involving 15 fixed-thickness layers. The best fitting model is plotted in bold magenta to improve clarity.**

A common solution is to select the number of unknowns using statistical methods based on the level of data fit achieved after repeated inversions specifying the number of unknowns each time. Such approaches have their limitations (See Sambridge et al. 2006) and in any case are always retrospective. Another approach is to deliberately choose a large number of unknowns (e.g. as in the 15 layer example in Figure 1) beyond what we believe the data will resolve and then impose some form of smoothing or damping to restrict extravagant behaviour. In the example in Figure 1 this was not done and one ends up with the extravagant best fitting conductivity profile in purple. Smooth models have their drawbacks too. Typically, physical properties in the subsurface are not expected to be smooth, as a consequence of discontinuities in lithologies (bedding, facies variations, faults)**.**

Recently a new approach, initially developed in the field of computational statistics, overcomes many of these problems in a more satisfactory manner. Bayesian partition modelling (hereafter referred to as BPM) (Denison et al. 2002a, 2002b) is a Monte Carlo technique in which the number of unknowns is itself an unknown to be constrained by the data. In contrast to the usual optimisation based techniques commonly employed in exploration problems, BPM is an ensemble based approach where many models of the subsurface are generated in a pseudo random fashion biased toward those that fit the observations. The Bayesian nature of this algorithm means that it is `naturally parsimonious' in that subsurface models with fewer unknowns are automatically favoured over those with more unknowns, without explicitly imposing this fact externally. As a result the ensemble of candidate solutions favours simple models, i.e. less extravagant (fitting our preconceptions) but this is achieved without imposing smoothness. Indeed a characteristic of the approach is that it is particularly good at estimating discontinuities in subsurface models when they are supported by the data. As such it is ideally suited to the 1-D inversion problem highlighted above and can also be extened to 2-3D problems.

In recent years the approach has been applied with success to various Earth Science problems including thermochronology (Stephenson et al. 2006), seismic tomography (Bodin et al. 2009, Bodin and Sambridge, 2009), inversion of stratigraphic data for environmental parameters (Charvin et al., 2009), and borehole temperature inversion (Hopcroft et al, 2009). In this paper we provide a brief introduction to the basic approach and illustrate it with a simple example. We believe there are significant potential applications of the methodology in several areas relevant to exploration geophysics, especially involving recovery of 1-D or 2-D spatial fields with discontinuities.

## BAYESIAN PARTITION MODELLING

The details of Bayesian Partition modelling including mathematics and computational algorithms can be found in Denison et al. (2002) and also the articles cited above in the context of geophysical applications. Here we provide a basic outline for the 1-D case illustrated through a simple synthetic example involving regression of a discontinuous field.

### 1-D regression example

We consider the 1-D regression problem in Figure 2. Here the grey piecewise continuous line is the unknown signal that we seek to determine using the noisy measurements represented by red dots. These were generated from the grey line with Gaussian random noise added using a standard deviation, $\sigma$, of 10 vertical units. The data therefore consist of values of (d,x) where d is a noisy estimate and x a noiseless control variable. We seek to estimate the underlying signal (grey line) and prefer not to smooth over the discontinuities in the signal. This is a classic problem ripe for application of 1-D Bayesian Partition modelling.

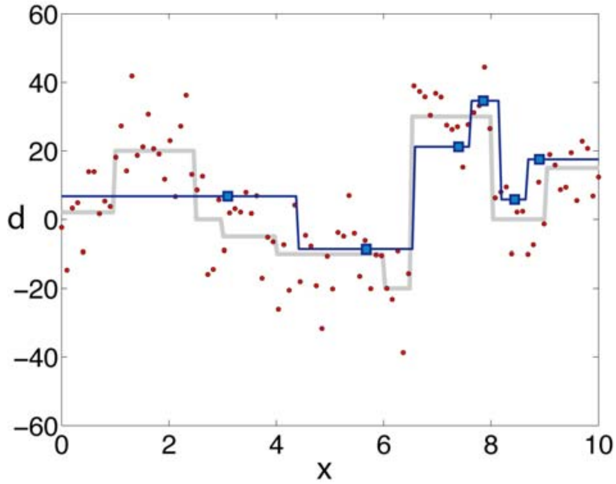


**Figure 2. Example of the parametrization used in BPM. Red dots are the data, grey line is the true model and blue line represents a partition model controlled by the heights, $d_j$, in each partition and the position of control points blue squares at $x_i$, which are equi-distant to interfaces.**

The real grey signal here contains nine separate partitions between which the function is discontinuous. The parametrization for the inverse problem is in terms of a finite set of partitions represented mathematically by the expression

$$d(x) = \sum_{i=1}^{k} m_i \phi_i(x) \qquad (1)$$

where $m_i$ is the height of the function in the ith partition, $\phi_t(x)$ is a basis function equal to 1 when x is inside partition i and zero otherwise, and k is the number of partitions. In this case the unknowns of the problem are the signal values, $m_i$ (i=1,…,k) the position of the partition boundaries and the number of partitions, k. Figure 2 shows an example of a d(x) with six partitions (blue line). The blue squares represent the locations of control points, $x_i$ (i=1,…,k) that are used to define the partition boundaries, which are equidistant between successive pairs of control points (see Figure 2). By varying ($m_{i,}$, $x_{i,}$, k) a wide range of possible piecewise functions can be generated with variable numbers of partitions. The inference problem simply stated is then given N noisy observations, $d_{j,}$, (j=1,…,N) we wish to reconstruct the underlying grey signal using the variables ($m_{i,}$, $x_{i,}$, k).

To measure the discrepancy between model predictions and data, the least squares misfit is calculated as the sum of squares of the differences between the heights of the red dots and the blue line, normalised by the data error. One way to solve the problem would be to seek a best fit solution by minimizing this misfit function. Irrespective of the data error, the best fit would be achieved by setting the number of partitions equal to the number of data, k = N, arrange for each datum to fall in its own partition and setting $m_i = d_j$ (i=1,…,k). This would produce a perfect fit to the data but clearly a poor estimate of the true signal. Such approaches are often quite ad hoc in that the type of smoothing will dictate the character of the final solution.

The alternative approach adopted here is to use a Bayesian sampling algorithm. In Bayesian inference all information is represented by probability density functions. Rather than just seeking a single best fit solution we generate many candidate solutions whose density follows the *a posterioi* probability density function p(**m** | **d**) where

$$p(m|d) \propto p(d|m)p(m). \qquad (2)$$

Here we collect all unknowns into the vector **m** and all data into the vector **d**. In this expression p(**m**|**d**) represents the probability density of the model, **m** given the data, **d**. It is proportional to the probability of obtaining the data, **d**, given the model, **m** (otherwise known as the likelihood) multiplied by the *a priori* probability of the model **m**. For Gaussian noise on the data the likelihood function is simply a Gaussian distribution measuring the discrepancy between $d_j$ and $m_j$ .

$$p(d\,|\,m) \propto \frac{1}{\sigma^N}\exp\left(-\tfrac{1}{2}\sum_{i=1}^{N}\frac{(d_i - g(m_{j(i)}))^2}{\sigma^2}\right) \qquad (3)$$

where j(i) indicates the model partition containing the ith datum, and g(.) is the forward model response. The prior PDF of the model parameters p(**m**) is represents our state of knowledge about the unknowns prior to receiving the data. Here we simply set p(**m**) to a uniform (constant) distribution which means it is a constant in (2). The object of the Bayesian approach is then to generate a large number of models **m** whose density follows p(**m** | **d**). For the regression problem each model consists of a set of values for ($m_i$, $x_i$, k) and corresponds to a piecewise constant curve like the blue line in Figure 2.

Algorithms for sampling complex distributions like (2) have been the subject of much research and have become practical with modern computing. The workhorse technique is the Markov chain Monte Carlo algorithm (McMC) (see Gallagher et al. (2009) for details). In recent years McMC algorithms have been extended to trans-dimensional problems (Geyer and Moller, 1994, Green, 1995, Sambridge et al. 2006).

In BPM the number of partitions and hence number of unknowns in the model are also variable and a transdimensional approach is needed. The results from a transdimensional McMC approach for our regression problem are shown in Figure 3b. The green curve is the average model in the ensemble generated by the Bayesian sampling algorithm. It is an excellent estimate of the underlying signal. The right panel of Figure 3b shows a histogram of the number of partitions in each model within the ensemble. Even though the prior PDF for this parameter was a uniform distribution between 1 and 50 the posterior distribution is peaked at 9 (the correct value). We see then that not only has BPM accurately recovered the true model it has also estimated the number of degrees of freedom adequately. This has been achieved without imposing any explicit form of regularization in terms of smoothness. Note also that the probability density of the number of partitions falls to zero by about 20 even though the prior distribution is finite up to 50. This feature has been driven by the data itself. The Bayesian approach is naturally parsimonious in that it prefers to choose fewer numbers of parameters to fit the data.

Implicit in all Bayesian procedures is an error estimate on the unknowns which can be represented by the variance of the histograms of each parameter such as those in Figure 3.

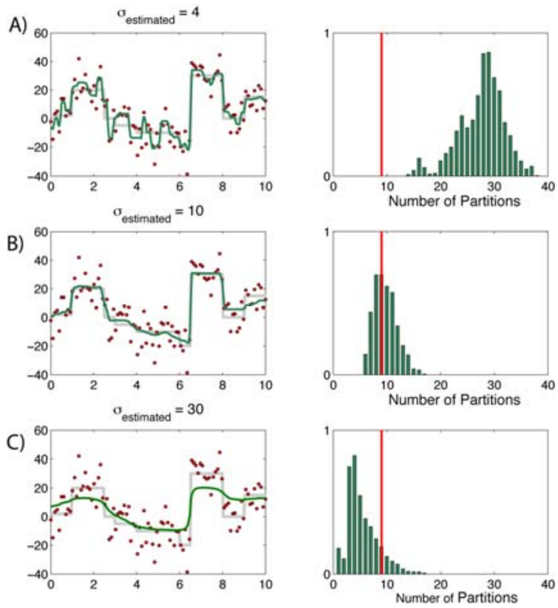


**Figure 3. Results of 1-D regression problem. Left hand panels show the data (red dots), true model (grey line) and average of the ensemble of curves produced by Bayesian Partition modelling (green line). The left hand panels show histograms of the number of partitions. Red line is the true value of 9. A) is for underestimated data noise level, B) for the correct noise level and C) for too high a noise level. In each case BPM finds the simplest model needed to fit the data.**

**Hierarchical Bayes – unknown data errors**

The results of the regression problem in Figure 3b were carried out using the correct level of noise on the data in the Likelihood function in (3). In most inversion problems one needs to know the level of data noise in order to calculate model uncertainty. Figure 3a and 3b show the results of the same BPM algorithm taking a error value which is too small (3a) and too large (3c). We see that the results are changed. When the error is too small we are in effect telling the algorithm to fit the data better, which it does by introducing a more complex model with too many partitions. Conversely when the errors are large we are saying that there is no need to fit as well. Hence the resulting model is less complex with fewer partitions. We see then that unlike the situation with optimisation of (3). which would be unaffected by the choice of data variance, the sampling algorithm is influenced by the choice of data noise. We need to know the level of data noise for the algorithm to perform well and estimate the true signal. In the common case where we do not know the data errors well, this raises the question `Can we let the data decide both the number of partitions in the model and the level of noise in the data ?' The answer is - yes it is possible to do just that in many cases. This is known as Hierarchical Bayesian inference (Malinverno and Briggs, 2004).

Figure 4 shows a repeat of the same regression experiment allowing both the number of partitions and the data variance to be an unknown. The ensemble average regression signal (Green line in Figure 4a) closely follows the true signal (grey). and the histogram of the number of partitions (Figure 4b) is again centred on the true value of 9. Figure 4c shows a histogram of the noise σ values in the ensemble, which is also centred near the correct value of 10 units.

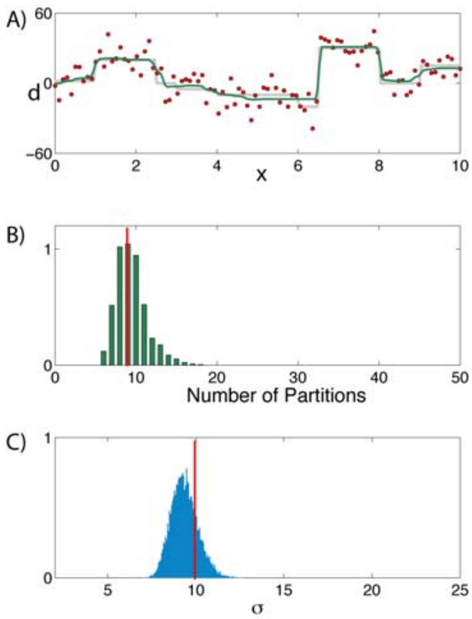


**Figure 4. Hierarchical Bayes applied to the regression problem a) shows the true (gray) and average recovered signal (green), b) & c) show histograms of the recovered number of partitions and standard deviation of the noise.**

## CONCLUSIONS

We present examples of a style of inversion where the number of unknowns and the level of noise in the data is unknown. We illustrate this with a simple regression problem where a piecewise constant function is recovered from noisy samples. The 1-D example is straightforward to extend to piecewise linear segments and to cases where the data consist of more complex physical properties. This style of 1-D inversion is readily adapted to several commonly encountered problems in exploration geophysics including the interpolation of discontinuous spatial fields (as an alternate to Kriging) (Stephenson et al., 2005), 1-D inversion of Airborne EM for subsurface conductivity (Brodie & Sambridge, 2006) and reservoir-related stratigraphic modelling (Charvin et al. 2009). In 2-D the same style of approach may be applied to inversion of ambient noise for near surface seismic structure (Bodin & Sambridge, 2009) and identifying spatial variations in palaeoclimate from borehole temperature inversion (Hopcroft et al. 2009).

The ability of the approach to infer discontinuities, while estimating both the number of degrees of freedom in the model and the noise level in the data make it highly novel in the geophysical context and a significant step beyond the commonly used least square optimisation framework.

## ACKNOWLEDGMENTS

We thank Ross Brodie of Geoscience Australia for providing Figure 1, and FAST (French-Australia Science and Technology exchange program) for their support during the preparation of this article. This project is supported by the Commonwealth of Australia under the *International Science Linkages program.*

## REFERENCES

Bodin, T. and Sambridge, M., 2009, Seismic tomography with the reversible jump algorithm, *Geophysical Journal International*, **178**, 1411-1436.

Bodin, T., Sambridge, M., and Gallagher, K., 2009, A self-parameterising partition model approach to tomographic inverse problems, *Inverse Problems*, **25**, 055009, doi:10.1088/0266-5611/25/5/055009, 17th March 2009.

Brodie, R., and Sambridge, M., A holistic approach to inversion of frequency domain airborne EM data**,** *Geophysics*, **71**, No. 6, pp G301-G312, 2006.

Charvin, K., Gallagher, K., Hampson, G. and Labourdette, R., 2009, A Bayesian approach to infer environmental parameters from stratigraphic data 1: Methodology, *Basin Research*, **21**, 5-25.

Denison, D., Adams, N., Holmes, C. and Hand, D., 2002a, Bayesian partition modelling, *Comput. Stat. Data Anal.*, **38**(4), 475–485.

Denison, D., Holmes, C., Mallick, B. and Smith, A., 2002b. *Bayesian Nonlinear Methods for Classification and Regression*, John Wiley, Chichester.

Gallagher, K., Charvin, Nielsen, S., Sambridge, M., and Stephenson, J., 2009, Markov chain Monte Carlo (McMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *Marine and Petroeum Geology*, **26**, 525-535.

Geyer, C. J. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes, *Scandinavian Journal of Statistics*, **21**, 369-373.

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika, **82**, 711-732.

Hopcroft, P., Gallagher, K. and Pain, C.C., 2009, A Bayesian Partition Modelling approach to resolve spatial variability in climate records from borehole temperature inversion, *Geophysical Journal International*, 178, 651-666.

Malinverno, A. and Briggs, V., 2004. Expanded uncertainty quanti_cation in inverse problems: Hierarchical Bayes and empirical Bayes, Geophysics, 69, 1005.

Sambridge, M., Gallagher, K, Jackson, A and Rickwood, P., 2006, Trans-dimensional inverse problems, Model Comparision and the Evidence, *Geophysical Journal International*, **167**, 528-542, doi: 10.1111/j.1365-246X2006.03155.x.

Stephenson, J., Gallagher, K. and Holmes, C., 2005, Beyond Kriging – dealing with discontinuous spatial data fields using adaptive prior information and Bayesian Partition modeling, in Curtis, A. and Wood, R. Geological Prior Information:

informing Science and Engineering, Geol. Soc. Lond. Spec. Publication 239, 195-209

Stephenson, J., Gallagher, K. and Holmes, C. (2006) Low temperature thermochronology and modelling strategies for multiple samples 2 : Partition modeling for 2D and 3D distributions with discontinuities, Earth Planet Sci. Letts. 241, 557-570.