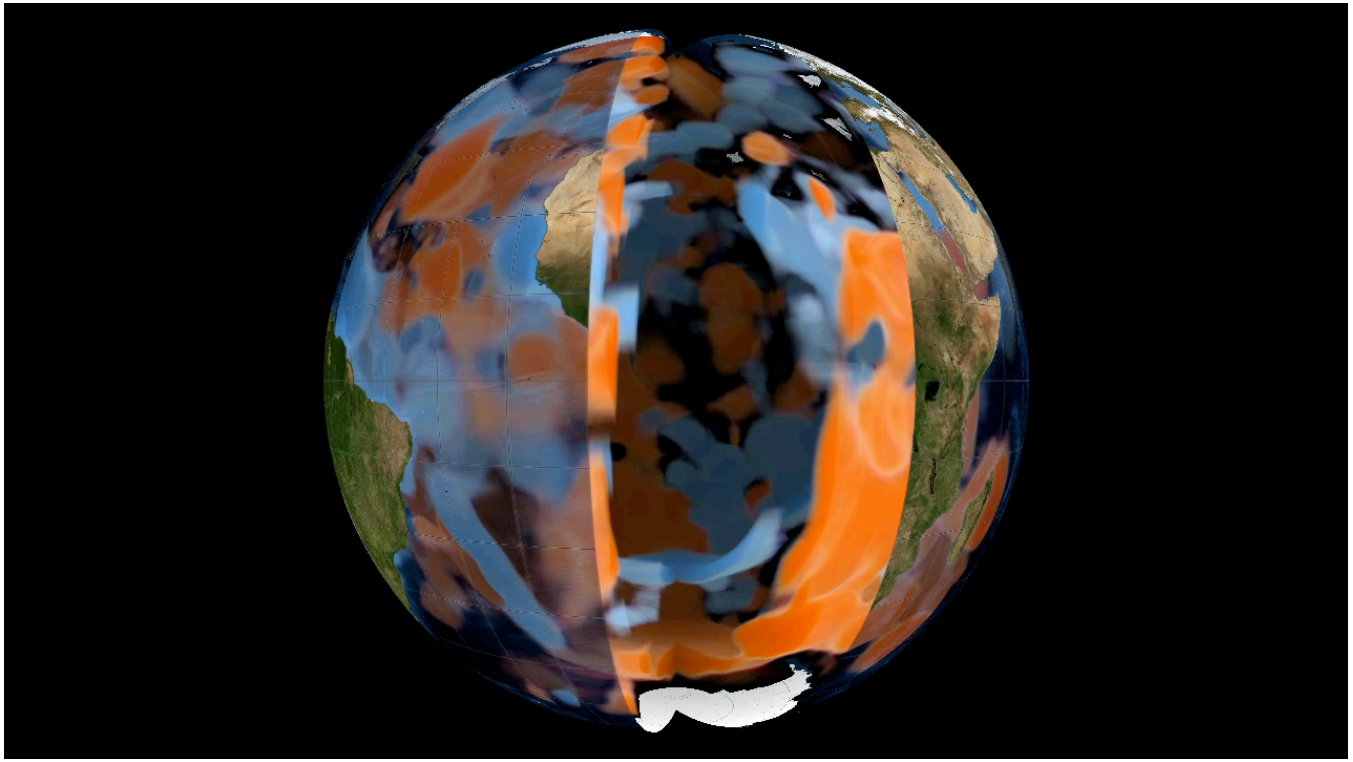
Transdimensional inverse problems in the geosciences

Malcolm Sambridge
Research School of Earth Sciences,
Australian National University
Canberra, ACT, Australia.

Signal processing and Inference for the physical sciences
26th March 2012, The Royal Society.

Collaborators: Thomas Bodin (ANU), Hrvoje Tkalcic (ANU)
Kerry Gallagher (Univ. de Rennes),
Ross Brodie (Geoscience Australia)

Geophysical imaging

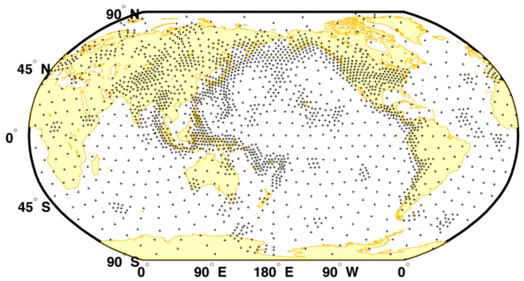


A 3-D image of the Earth's mantle using multi-frequency seismic shear waves

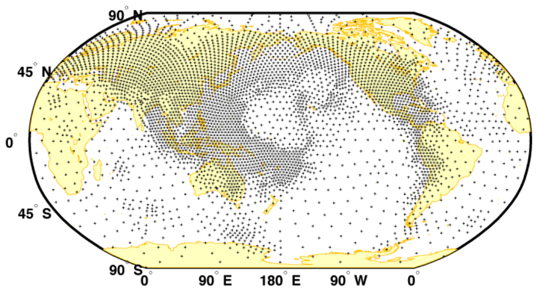
From Zaroli (2010)

Parametrization uncertainty

100–200 km

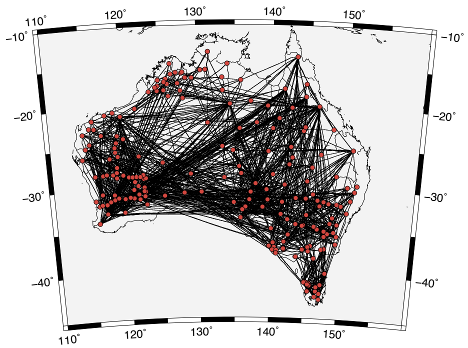


810–960 km

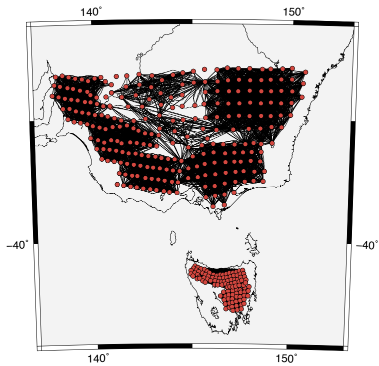


Raypth densities of 400,000 SH wave
multiple frequency travel times

From Zaroli (2010)



Ambient noise tomography in Australia
Saygin (2007)

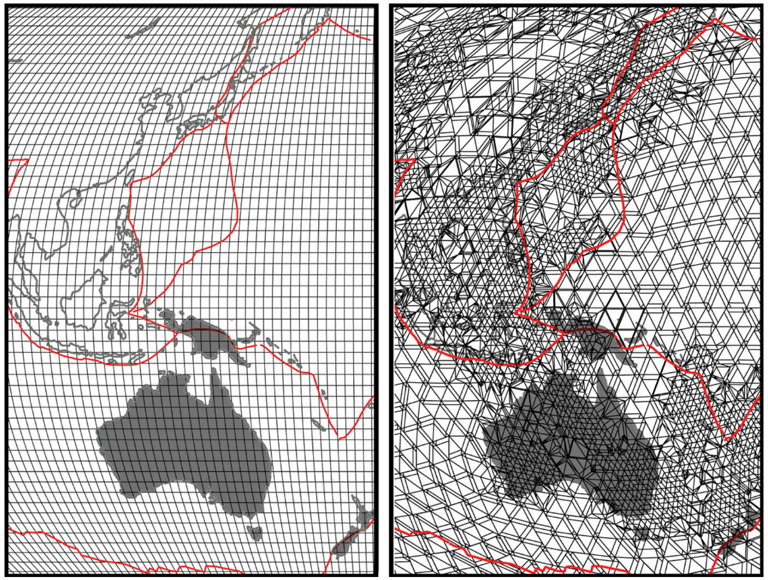


WOMBAT and TIGGER arrays
Arroucau et al (2010)

Irregular parametrizations

Static optimized meshes in regional and global studies have become popular

Zaroli et al. (2010)



Sambridge & Rawlinson (2005)

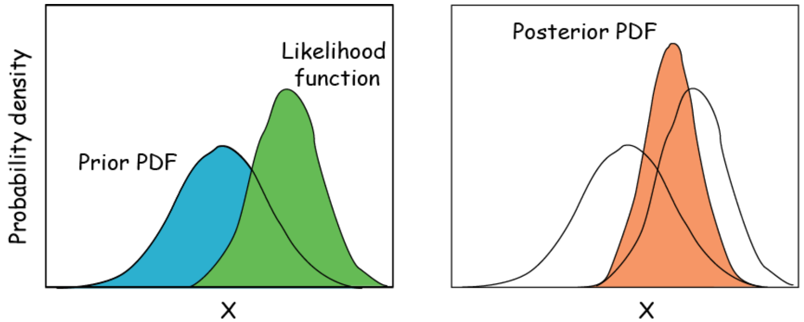
Is it possible to use the data to decide how to parametrize the model ?

Chou & Booker (1979); Tarantola & Nercessian (1984); Abers & Rocker (1991); Fukao et al. (1992); Zelt & Smith (1992); Michelini (1995); Vesnaver (1996); Curtis & Snieder (1997); Widiyantoro & van der Hilst (1998); Bijwaard et al. (1998); Bohm et al. (2000); Sambridge & Faletic (2003).

A probabilistic framework

Posterior PDF — Likelihood — Prior PDF

$$p(\mathbf{m}|\mathbf{d}) = C \times p(\mathbf{d}|\mathbf{m})\ p(\mathbf{m})$$



$$\mathbf{m}^T = (m_1, \ldots, m_k)$$

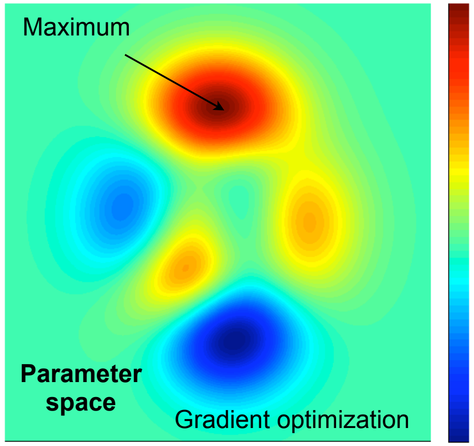
$$m(\mathbf{x}) = \sum_{j=1}^{k} m_j \phi_j(\mathbf{x})$$

Probabilities associated with inferences

model estimation <-> model uncertainty

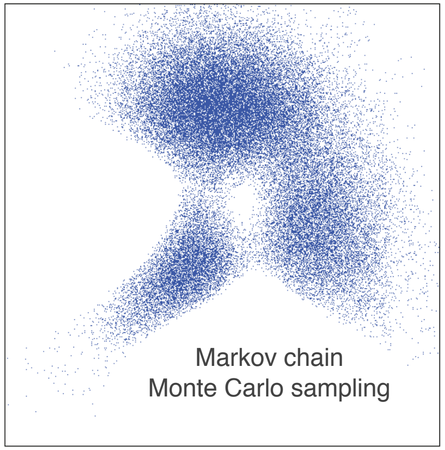
Sampling and optimization

Optimization framework



Data fit or Log Likelihood function

Probabilistic sampling framework



Samples from posterior PDF

$$\phi(\mathbf{m}) = ||\mathbf{d} - g(\mathbf{m})||_2^2 + \alpha^2||\mathbf{m}||_2^2$$

$$p(\mathbf{m}|\mathbf{d}) = C \times p(\mathbf{d}|\mathbf{m})\ p(\mathbf{m})$$

Transdimensional inversion

'with four parameters I can fit an elephant and with five I can make him wiggle his trunk'

J. von Neumann

Posterior PDF $p(\mathbf{m}|\mathbf{d}, I)$ I, assumptions

Assumption uncertainty

$$\mathbf{m}(\mathbf{x}) = \sum_{j=1}^{k} m_j \psi_j(\mathbf{x})$$

Uncertain parametrization

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{(\sigma\sqrt{2\pi})^N} \exp\left\{-\frac{1}{2}\sum_{i}^{N} \frac{|\mathbf{d} - g(\mathbf{m})|^2}{\sigma^2}\right\}$$

Uncertain Likelihood

Hierarchical Bayes

Let the data decide

Many applications...

Dension, Holmes, Mallick and Smith, 2002

Markov chain Monte Carlo

Target PDF

$$p(\mathbf{m}|\mathbf{d}) = C \times p(\mathbf{d}|\mathbf{m})\ p(\mathbf{m})$$

Proposal PDF

$\mathbf{m}_1 \rightarrow \mathbf{m}_2$

$$q(\mathbf{m}_2|\mathbf{m}_1)$$

M-H Acceptance probability

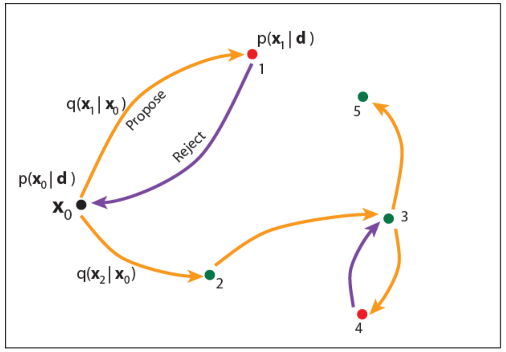
$$\alpha = 1 \wedge \left[\frac{p(\mathbf{m}_2|\mathbf{d})q(\mathbf{m}_1|\mathbf{m}_2)}{p(\mathbf{m}_1|\mathbf{d})q(\mathbf{m}_2|\mathbf{m}_1)}\right]$$

Issues:

Efficiency

Dimension

M-H algorithm



Example

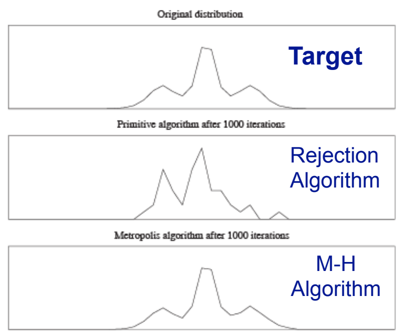
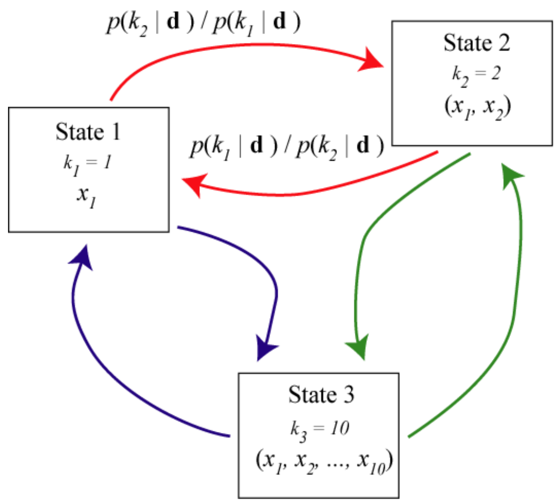


Figure 1. A probability distribution (top) were sampled 1000 times by the primitive algorithm (middle) and the Metropolis algorithm (bottom). The superiority of the Metropolis algorithm over the primitive algorithm is seen even in this one-dimensional example. In the Metropolis case all 1000 iterations yielded a sample contributing to the ?nal result, whereas for the primitive algorithm, only 42 samples were accepted for histogram building.

Transdimensional inversion

Variable dimension

$$p(\mathbf{m}, k|\mathbf{d}) = C \times p(\mathbf{d}|\mathbf{m}, k)\ p(\mathbf{m}|k)p(k)$$



$$\alpha = 1 \wedge \left[\frac{p(\mathbf{d}|\mathbf{m}_2, k_2)p(\mathbf{m}_2|k_2)p(k_2)q(\mathbf{m}_1, k_1|\mathbf{m}_2, k_2)}{p(\mathbf{d}|\mathbf{m}_1, k_1)p(\mathbf{m}_1|k_1)p(k_1)q(\mathbf{m}_2, k_2|\mathbf{m}_1, k_1)}|J|\right]$$

Birth-Death McMC algorithm,
Geyer & Moller (1994)

$$k_2 = k_1 \pm 1$$

RJ-McMC algorithm
Green (1995)

$$m(\mathbf{x}) = \sum_{j=1}^{k} m_j \phi_j(\mathbf{x})$$

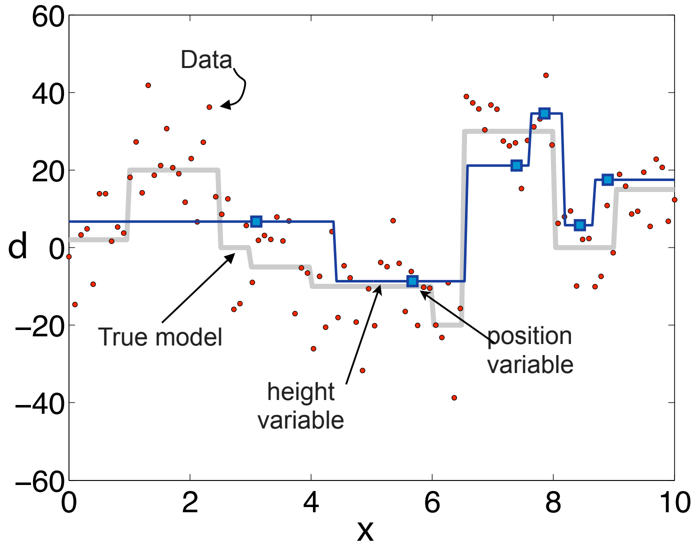
Example regression

' Make everything as simple as possible, but not simpler'

Einstein

Birth-Death example
1-D Regression

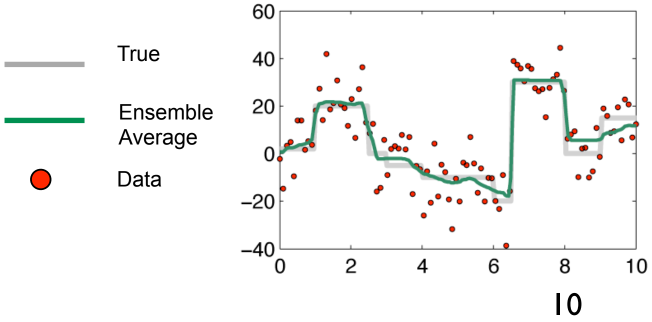
Recover the discontinuous function (gray) from the noisy data (red)

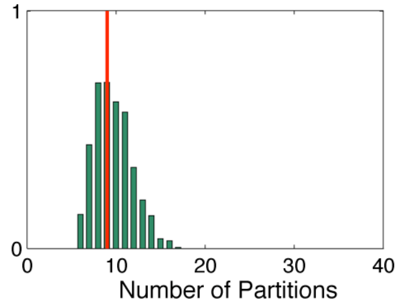


Partition modelling

$$\mathbf{m}^T = (\mathbf{f}, \mathbf{x}, k)^T$$

Properties of the ensemble may be examined





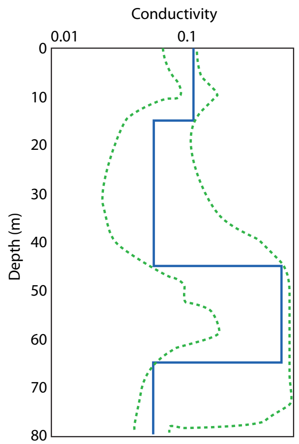
Uncertainty in model dimension

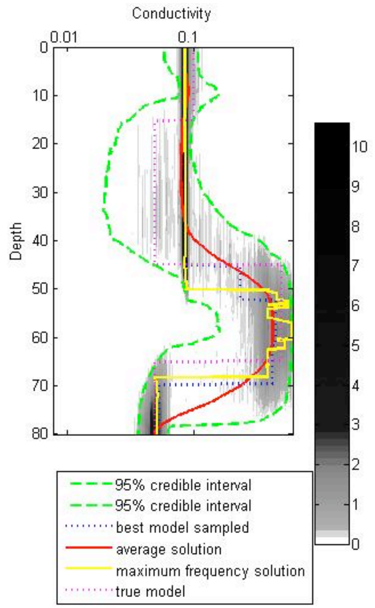
Inversion for 1-D earth models

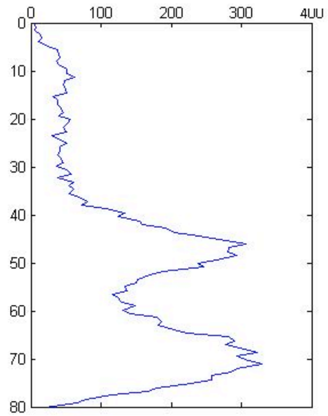
Using Airborne EM to constrain subsurface conductivity

$$\mathbf{m}^T = (\mathbf{c}, \mathbf{x}, k)^T$$

MEGATEM, Dash 7

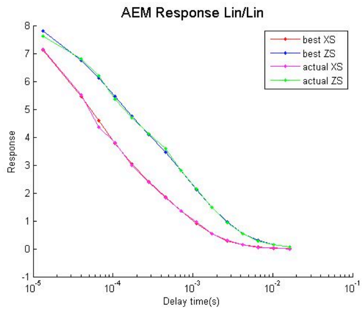


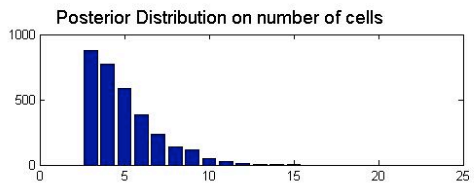




Marginals in interface position

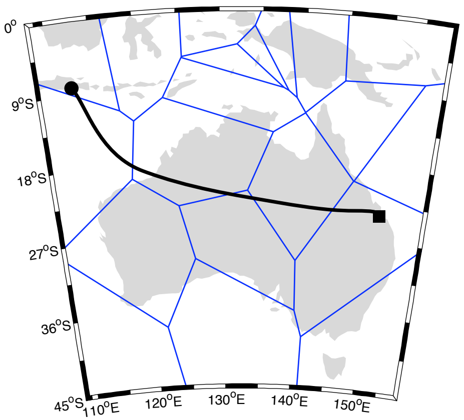
Uncertainty in discontinuity position



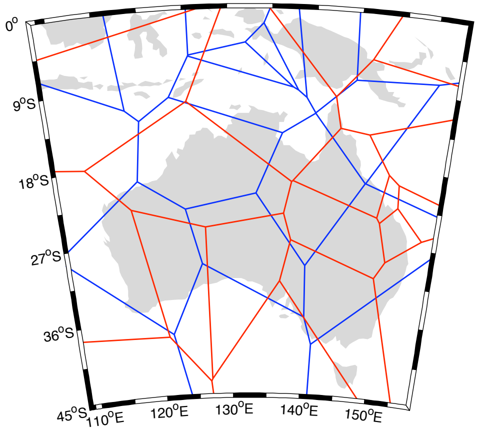


Courtesy R. Brodie, M. Hartley

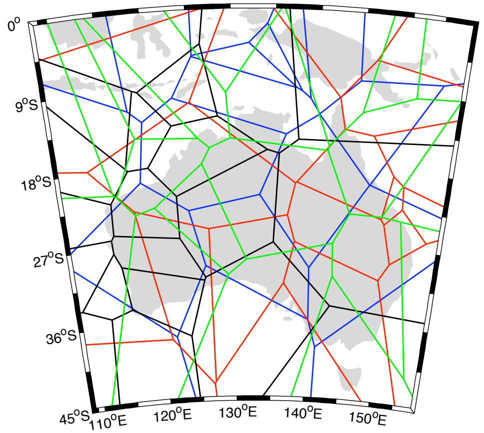
2-D spatial problems



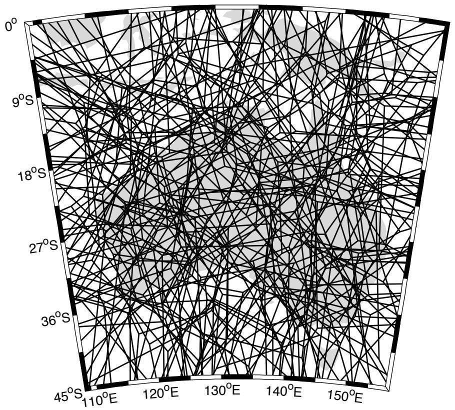
20 uniformly random Voronoi cells



2 overlapping Voronoi meshes



4 overlapping Voronoi meshes



20 Voronoi meshes

Birth-Death McMC with unknowns $\mathbf{m}^T = (\mathbf{f}, \mathbf{x}, k)^T$

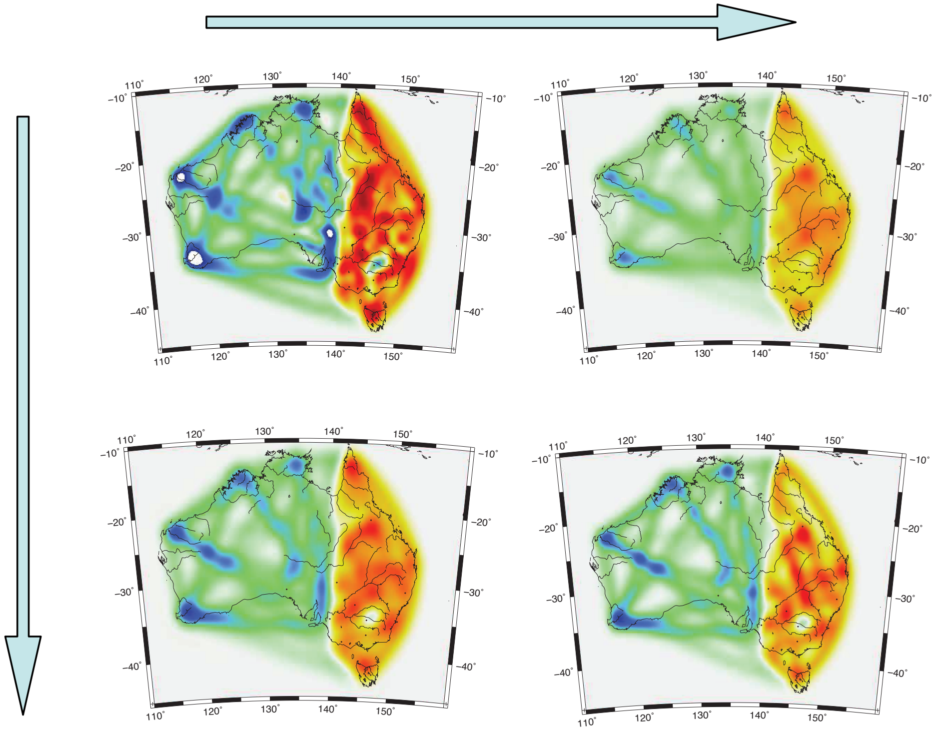
Point by point averaging of output gives an ensemble mean, covariance measures.

Effective mesh resolution increases with number of meshes -> super resolution.

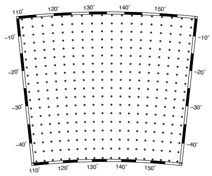
Classic tomography: single models with a fixed grid

Damping

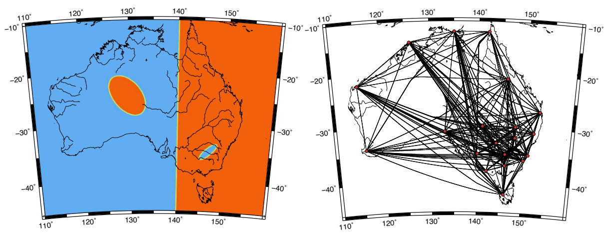
Smoothing

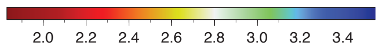


Tikhonov solutions



50 x 35
B-splines nodes





$$\min_{\mathbf{m}} \quad ||\delta\mathbf{d} - G\delta\mathbf{m}||^2 + \alpha||\mathbf{m} - \mathbf{m}_o||^2 + \beta||\nabla\mathbf{m}||^2$$

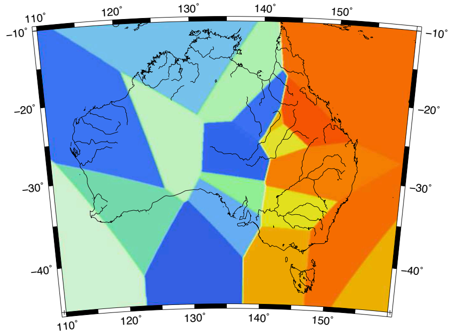
After Bodin et al. (2009)

2-D tomography

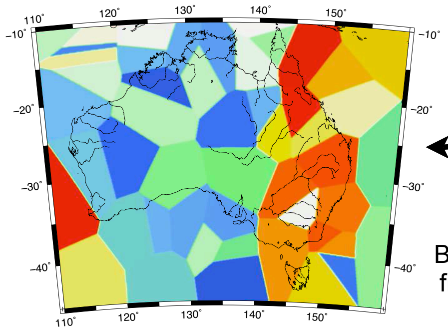
The wisdom of the crowd

$$\mathbf{m}^T = (\mathbf{s}, \mathbf{x}, k)^T$$

Individual models along the Markov chain

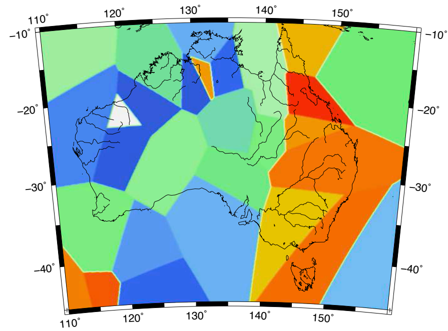


20 cells



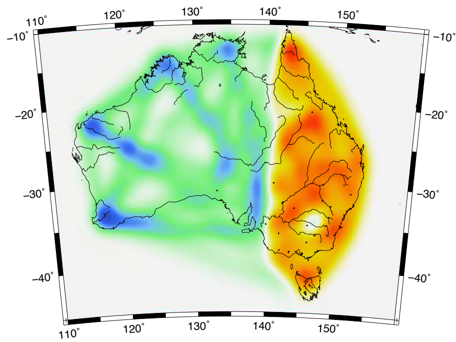
50 cells

Best data fit model

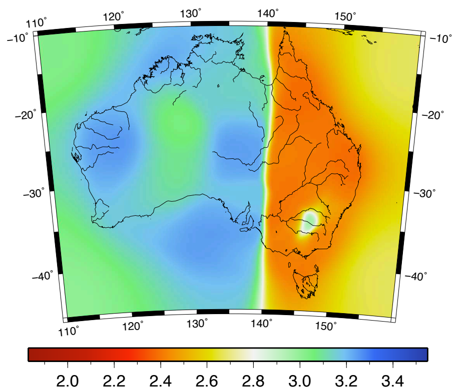


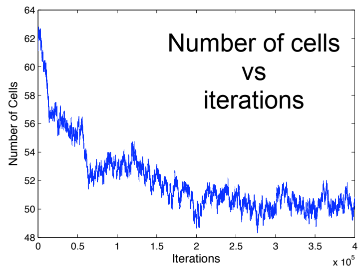
30 cells

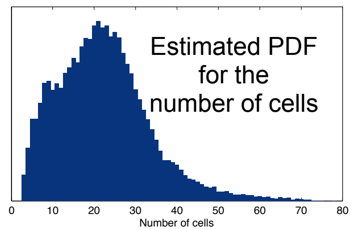
Optimization solution
1750 cells + regularization

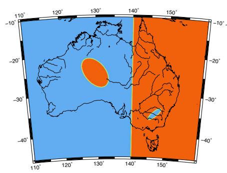


Ensemble mean

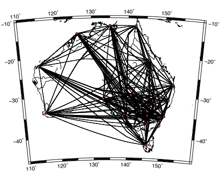








True

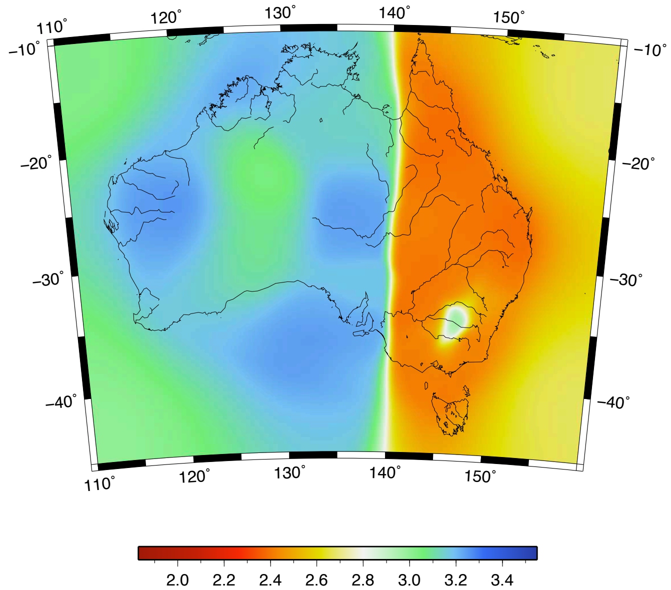


Data

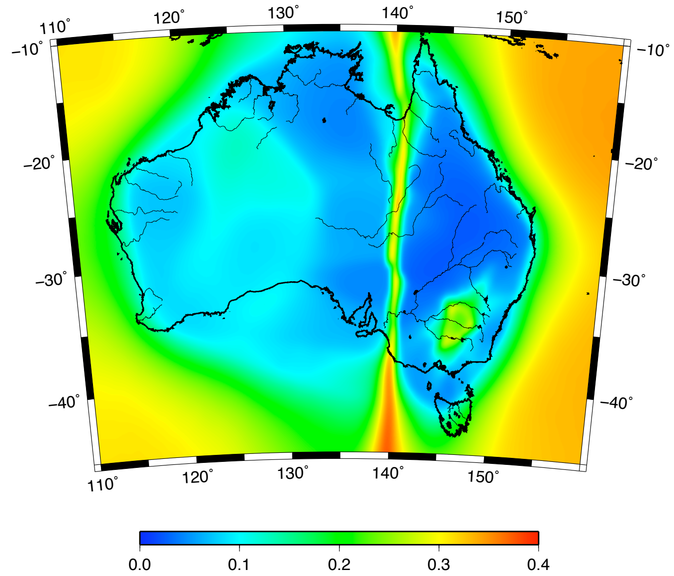
Super resolution & self regularization

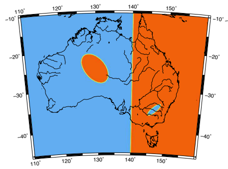
Uncertainty estimates for free

Ensemble average

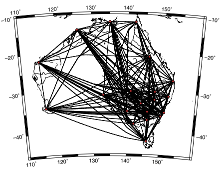


Ensemble variance





True

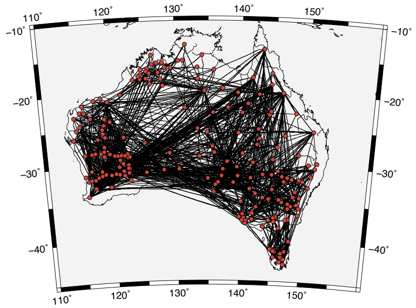


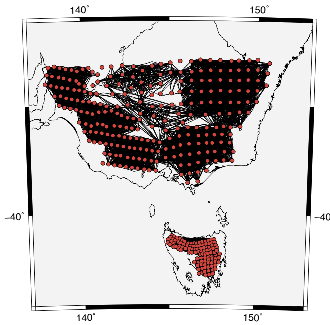
Data

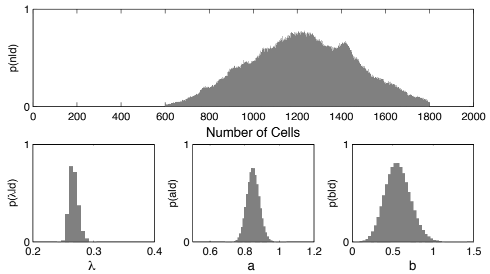
Ambient noise imaging

Transdimensional Hierarchical Bayes

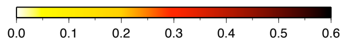
$$\mathbf{m}^T = (\mathbf{v}, \lambda_i, \mathbf{x}, k)^T$$



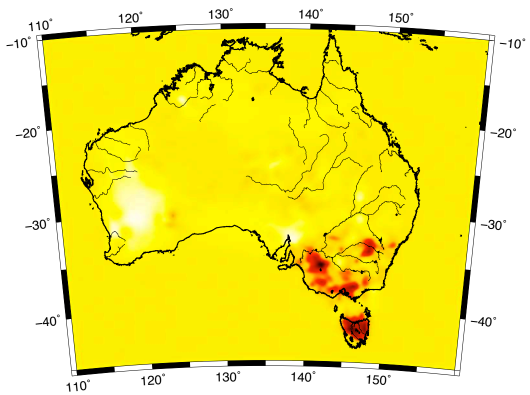


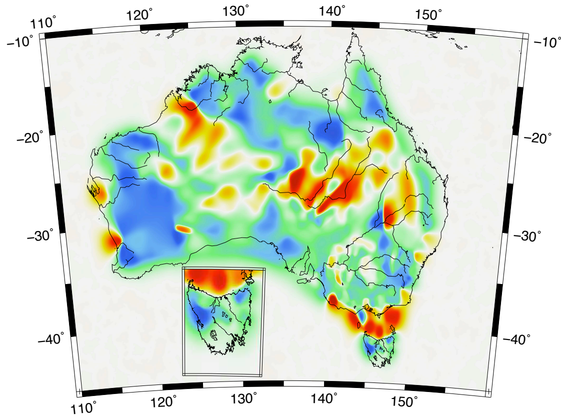


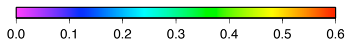
PDF Data uncertainty



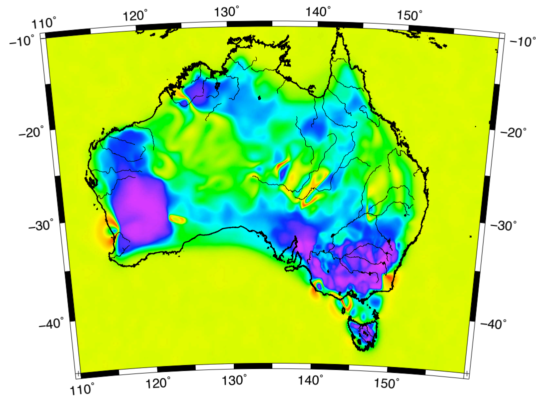
Velocity cell density







Output: Error in the ensemble mean



From Bodin et al. (2011)

Hierarchical Bayes

When data uncertainty is uncertain

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{(\sigma\sqrt{2\pi})^N} \exp\left\{-\frac{1}{2}\sum_i^N \frac{|\mathbf{d} - g(\mathbf{m})|^2}{\sigma^2}\right\}$$

$$\sigma \longrightarrow \lambda\sigma$$

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{(|C_d|(2\pi)^N)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{d} - g(\mathbf{m}))^T C_d^{-1}(\mathbf{d} - g(\mathbf{m}))\right\}$$

$$C_d(\lambda_1, \lambda_2)$$

We can solve for
data noise correlation too

Noise is what we don't fit with the data

$$p(\mathbf{m}, \lambda_i, k|\mathbf{d}) = C \times L(\mathbf{d}|\mathbf{m}, \lambda_i, k)\ p(\mathbf{m}|k)p(\lambda_i)p(k)$$

Markov chain, $(\mathbf{m}, \lambda_i, k) \rightarrow (\mathbf{m}', \lambda_i', k')$

$$\alpha = 1 \wedge \left[\frac{p(\mathbf{m}', \lambda_i', k')|\mathbf{d})q(\mathbf{m}, \lambda_i, k|\mathbf{m}', \lambda_i', k')}{p(\mathbf{m}, \lambda_i, k)|\mathbf{d})q(\mathbf{m}', \lambda_i', k'|\mathbf{m}, \lambda_i, k)}|J|\right]$$

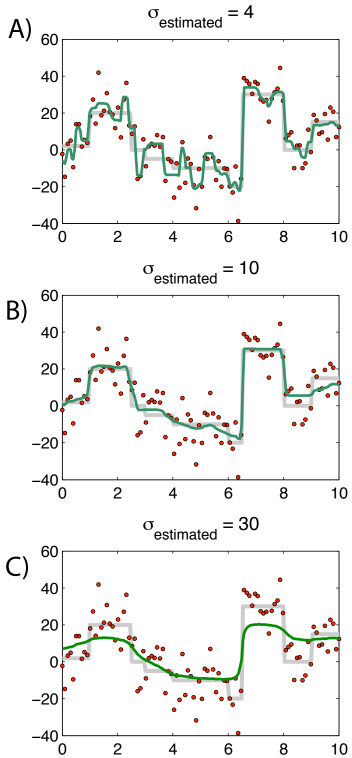
Hierarchical Bayes

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{(\lambda\sigma\sqrt{2\pi})^N} \exp\left\{-\frac{1}{2\lambda^2}\sum_i^N \frac{|\mathbf{d} - g(\mathbf{m})|^2}{\sigma^2}\right\}$$

Transdimensional with fixed values for data uncertainty

$$\mathbf{m}^T = (\mathbf{f}, \mathbf{x}, k)^T$$

λ and k are related



Model too complex
Data over fit

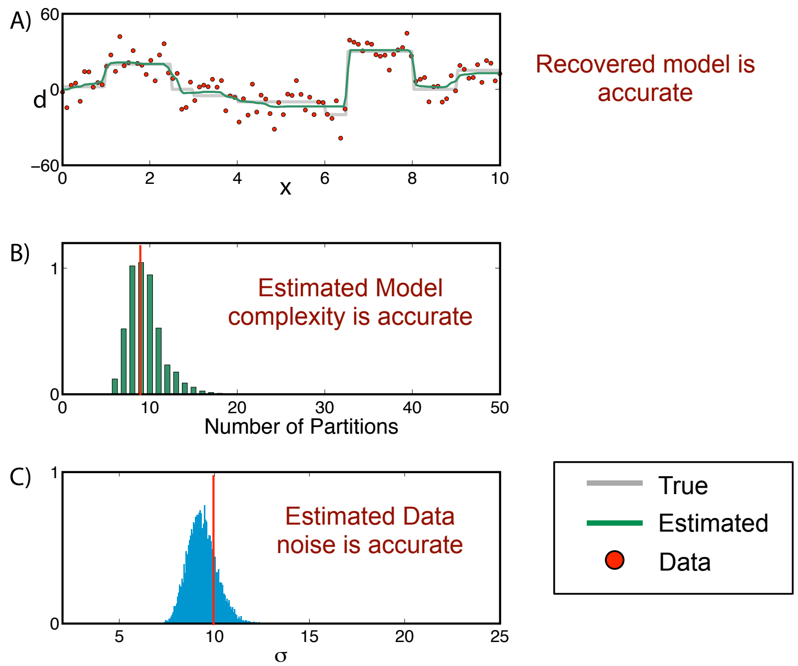
Model complexity OK
Data fit good

Model too smooth
Data under fit

Transdimensional with variable data uncertainty

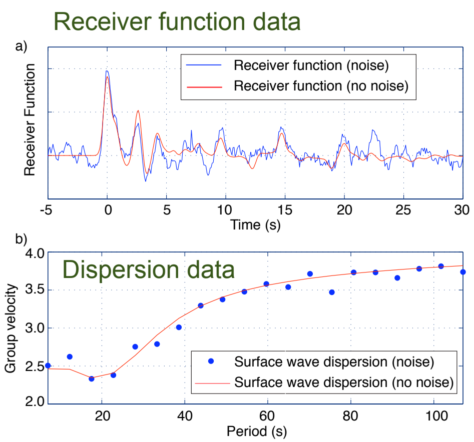
$$\mathbf{m}^T = (\mathbf{f}, \mathbf{x}, \lambda, k)^T$$

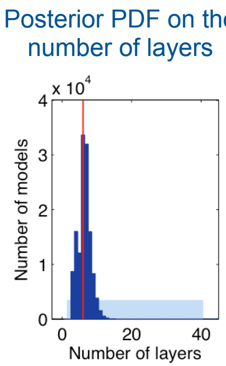
Hierarchical algorithm

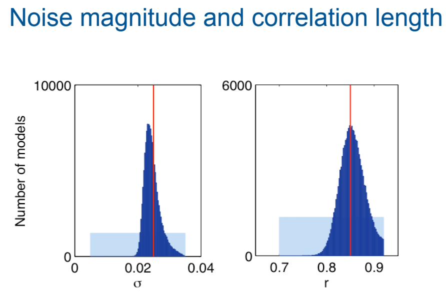


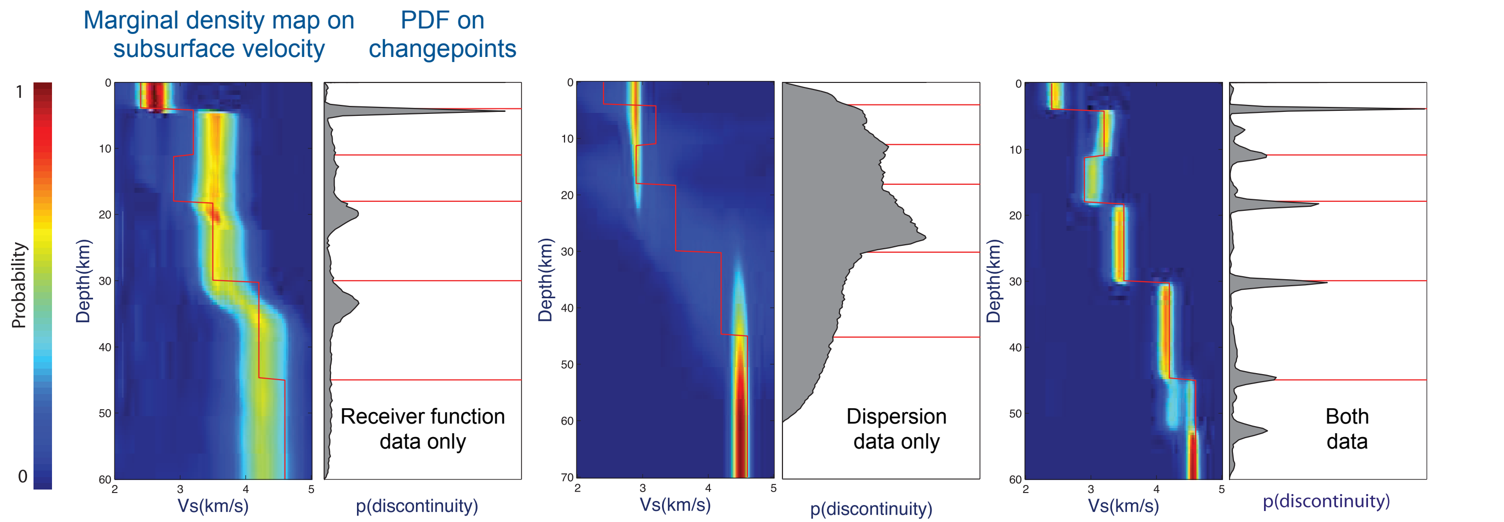
Joint inversion of multiple classes of data

$$\mathbf{m}^T = (\mathbf{v}, \lambda_i, \mathbf{x}, k)^T$$









After Bodin et al. (2012)

Summary

- It is possible to use the data to constrain both the parametrization and the noise with sampling based approaches.

- Trans-dimensional ensemble based approaches look promising for a range of applications in the geosciences. General methodology.

- Ensemble properties are reasonably robust and can be used as an estimator of the solution. Make use of the *wisdom of the crowd*.

The end

Some papers on Trans-dimensional inversion

Electrical resistivity inversion

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a 107 nonlinear geophysical problem, *Geophysical Journal International*, **151**(3), 675–688.

Malinverno, A. & Briggs, V., 2004. Expanded uncertainty quantification in inverse prob1029 lems: Hierarchical Bayes and empirical Bayes, *Geophysics*, **69**, 1005.

Climate histories

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**(3), 1430–1439.

Hopcroft, P., Gallagher, K., & Pain, C., 2009. A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion, *Geophysical Journal International*, **178**(2), 651–666.

Thermochronology

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M., & 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 Petroleum Geology*, **26**(4), 525–535.

Stratigraphic modelling

Charvin, K., Gallagher, K., Hampson, G., & Labourdette, R., 2009a. A Bayesian approach to inverse modelling of stratigraphy, part 1: method, *Basin Research*, **21**(1), 5–25.

Some papers on Trans-dimensional inversion

Geochronlogy

Jasra, A., Stephens, D., Gallagher, K., & Holmes, C., 2006. Bayesian mixture modelling in geochronology via Markov chain Monte Carlo, *Mathematical Geology*, **38**(3), 269–300.

General

Gallagher, K., Bodin, T., Sambridge, M., Weiss, D., Kylander, M., & Large, D., 2011. Inference of abrupt changes in noisy geochemical records using Bayesian Transdimensional changepoint models, *Earth and Planetary Science Letters*, (October 2011) doi:10.1016/j.epsl.2011.09.015

ANU papers on Seismology, Regression and other geophysical problems

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

Seismic tomography with the reversible jump algorithm,
Bodin, T. & Sambridge, M., 2009. *Geophysical Journal International*, **178**(3), 1411–1436.

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

Data inference in the 21st Century: Some ideas from outside the box,
Sambridge, M., Bodin, T., Reading, A., and Gallagher, K., 2010. *Australian Soc. of Exploration Geophysics 21st International Geophysics Conference and Exhibition,* Extended Abstracts, 22-26 August, Sydney, Australia.

Transdimensional Inversion of Receiver Functions 1**and Surface Wave Dispersion**
T. Bodin, M. Sambridge, H. Tkalcic, P. Arroucau, K. Gallagher,and N. Rawlinson,
Journal of Geophysical research submitted, 2011.

Some papers are at http://rses.anu.edu.au/~malcolm/