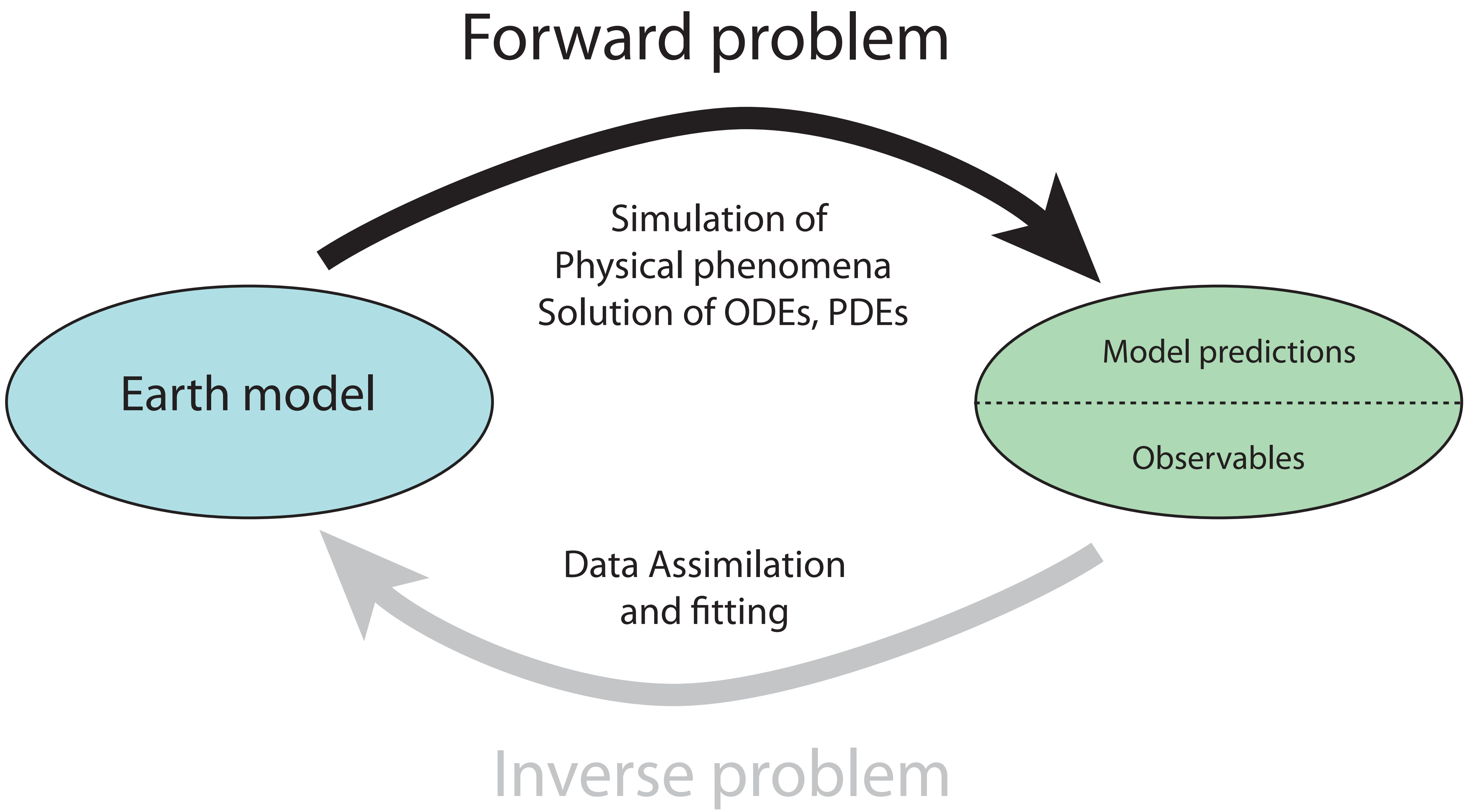
Non-deterministic approaches to inverse problems

Malcolm Sambridge & Thomas Bodin
Research School of Earth Sciences,
Australian National University
Geomath08, 15-17th Sept. 2008
Santa Fe, NM, USA.

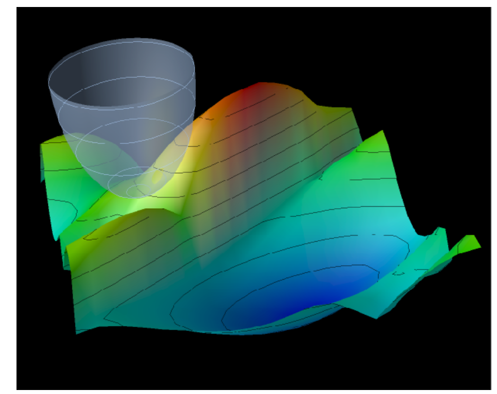
Forward and inverse problems

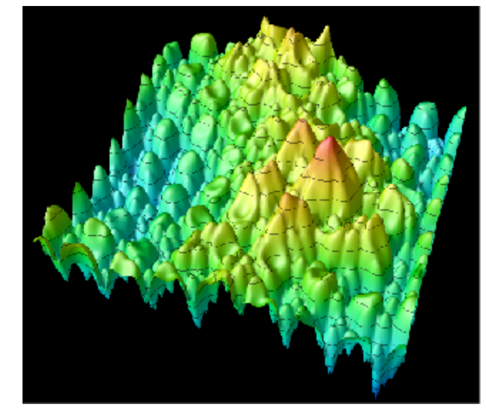


Irritating properties of inverse problems

Nonlinear

$$\mathbf{d} = g(\mathbf{m})$$





Ill-posed

$$d(s) = \int_a^b g(s,x)m(x)dx$$

Small changes in data lead to large changes in a solution

$G^T G$ has large condition number

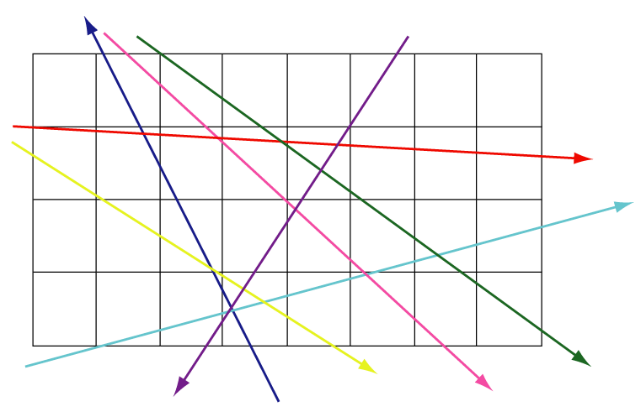
Fredholm integral equations
Deconvolution
Fourier transform inversion

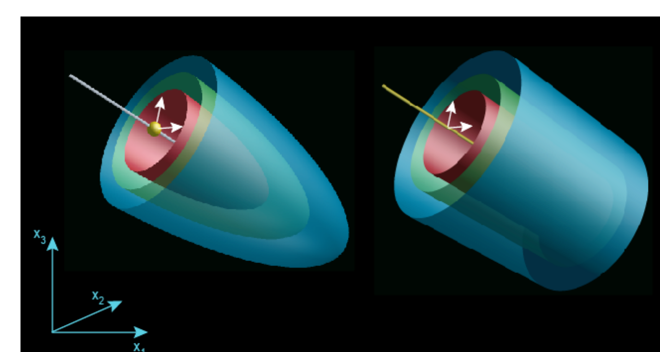
Non-unique

$$\mathbf{d} = G\mathbf{m}$$

$$\mathbf{m} = (G^T G)^{-1} G^T \mathbf{d}$$

$G^T G$ has zero eigenvalues

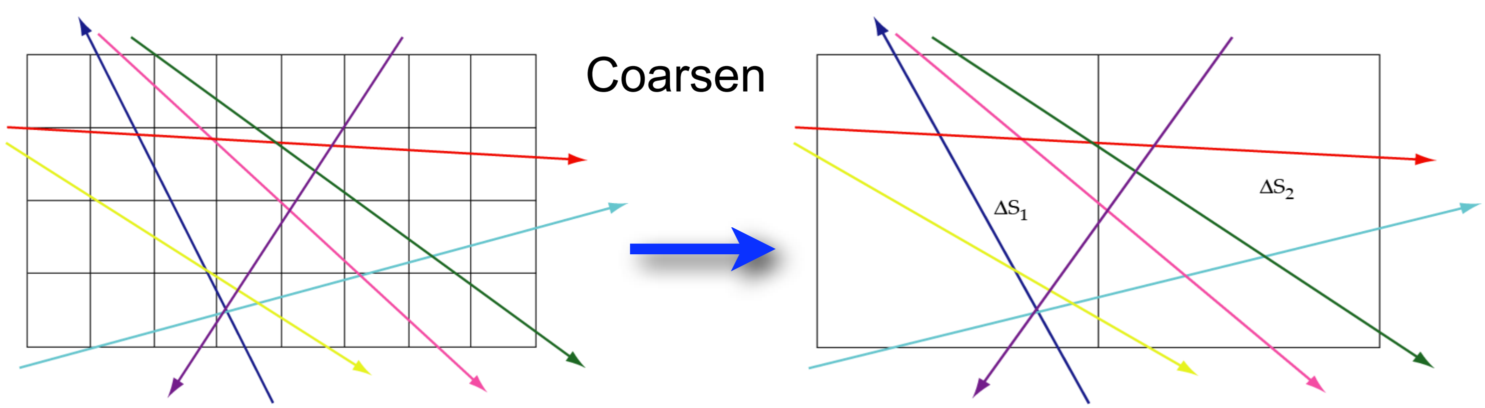




Non-uniqueness in Linear(ized) problems

$$\mathbf{d} = G\mathbf{m} \qquad \mathbf{m} = (G^T G)^{-1} G^T \mathbf{d}$$

A Parametrization solution

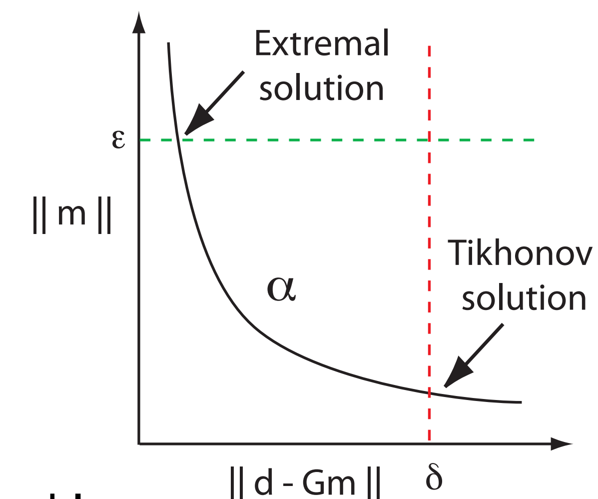


Global reduction in resolution -> reduces information retrieval

A Tikhonov solution

$$\min \quad ||G\mathbf{m} - \mathbf{d}||_2^2 + \alpha^2 ||\mathbf{m}||_2^2$$

$$\mathbf{m} = (G^T G + \alpha^2)^{-1} G^T \mathbf{d}$$



Global reduction in resolution. Formal errors are biased !

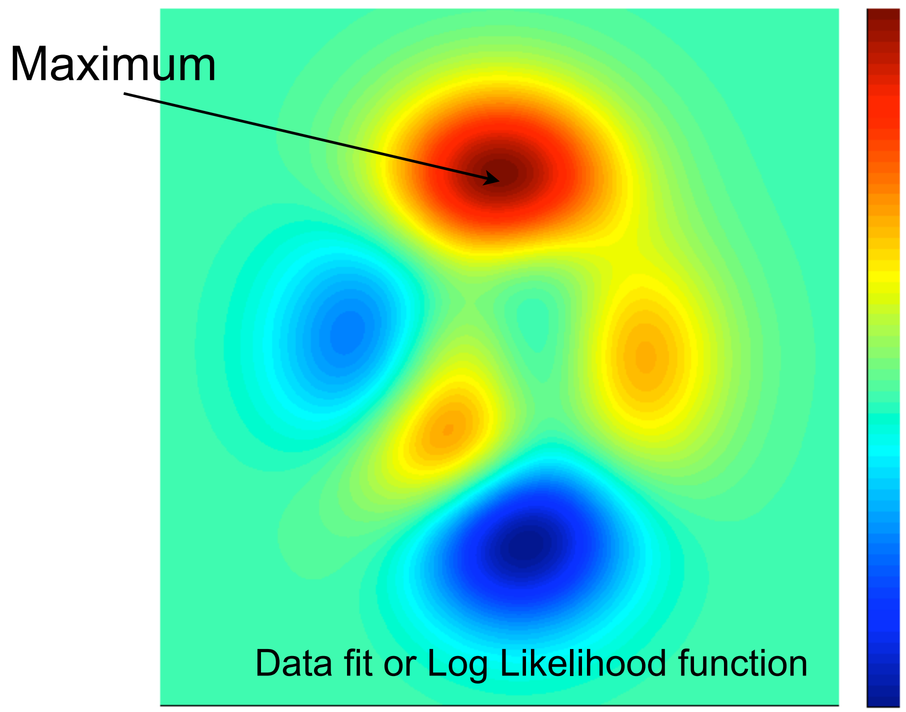
Looking for a single 'best' solution is not the only game in town

Take home messages

- Replace optimization framework with a randomized sampling based approach
- Single 'best' solutions are replaced by properties of many partial solutions
- Global (crude) regularization term replaced with data adaptive smoothing
- Obtain statistical measures of model error

An ensemble approach to inversion

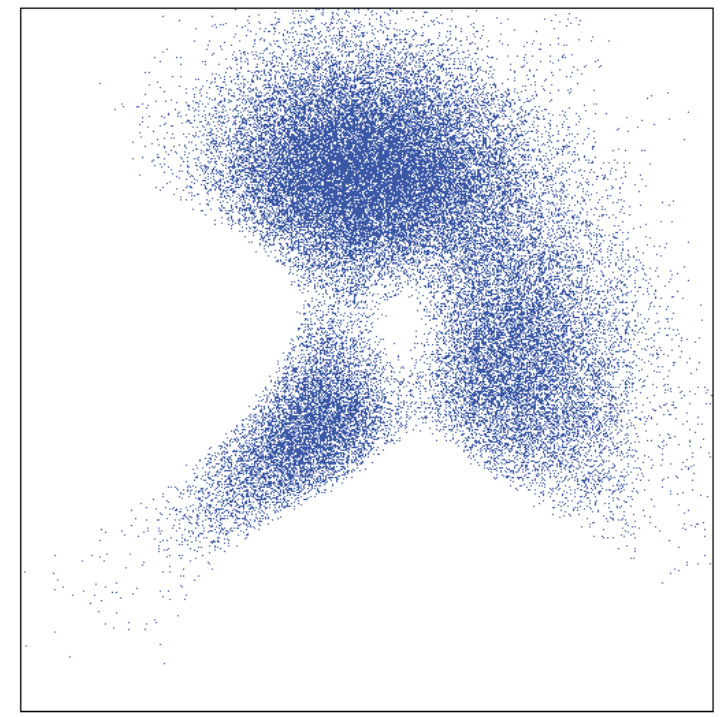
Optimization framework



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

Bayesian sampling framework

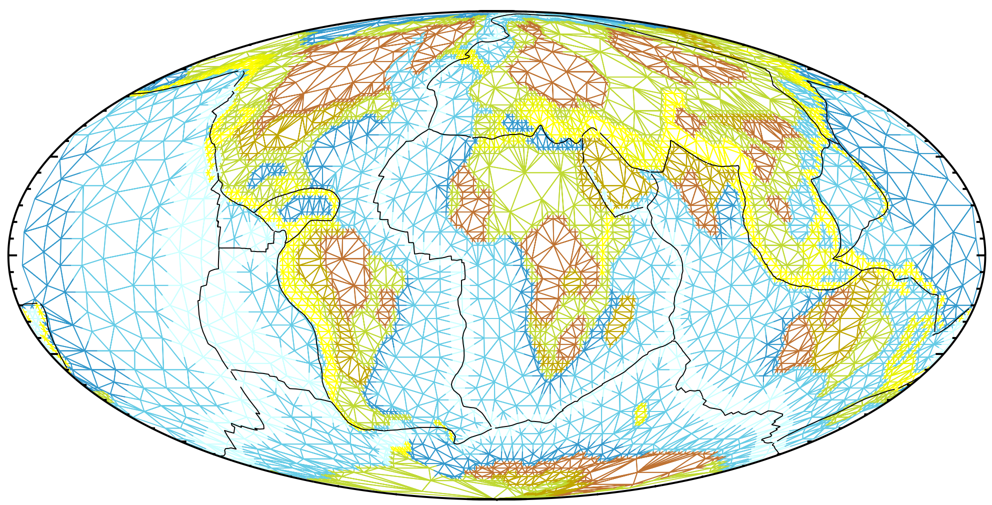
A nonlinear problem



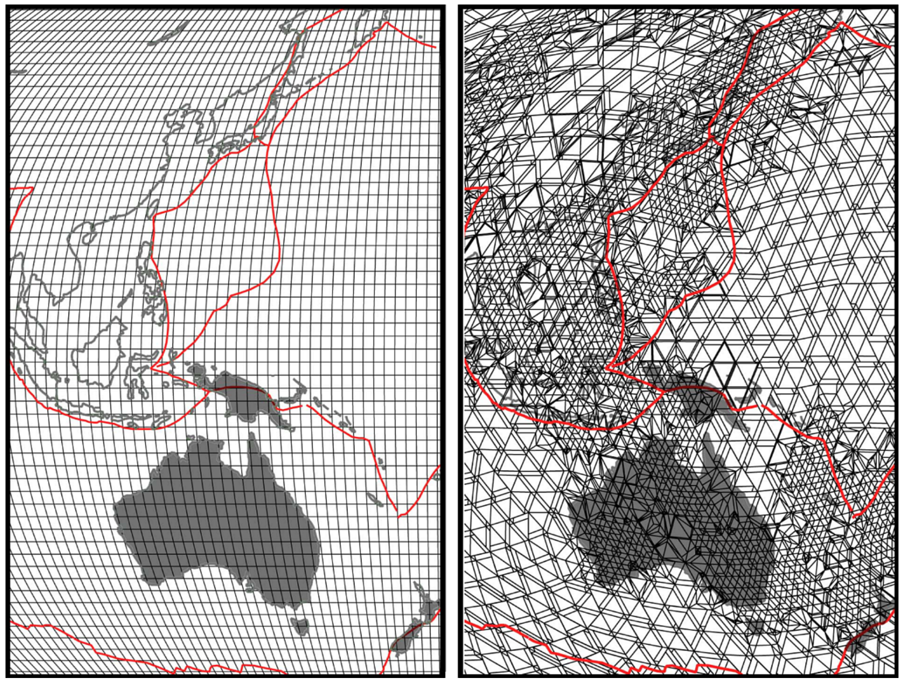
$$p(\mathbf{m}|\mathbf{d}) = k \times p(\mathbf{d}|\mathbf{m})\ p(\mathbf{m})$$

- Probabilistic framework -> inference described by probabilities
- Seek many potential solutions rather than just one
- Look at properties of solutions
- Computation high, but not infeasible.

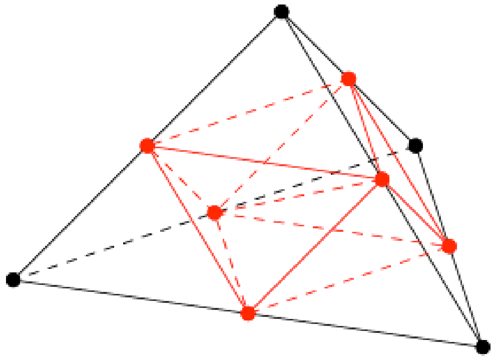
Adaptive parametrizations in seismic tomography

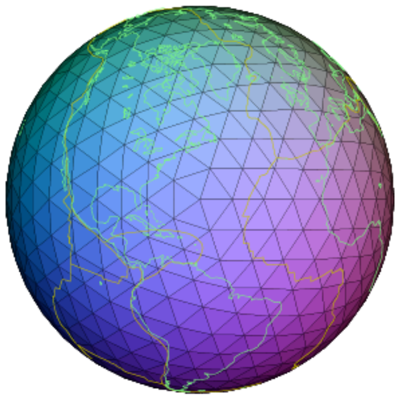


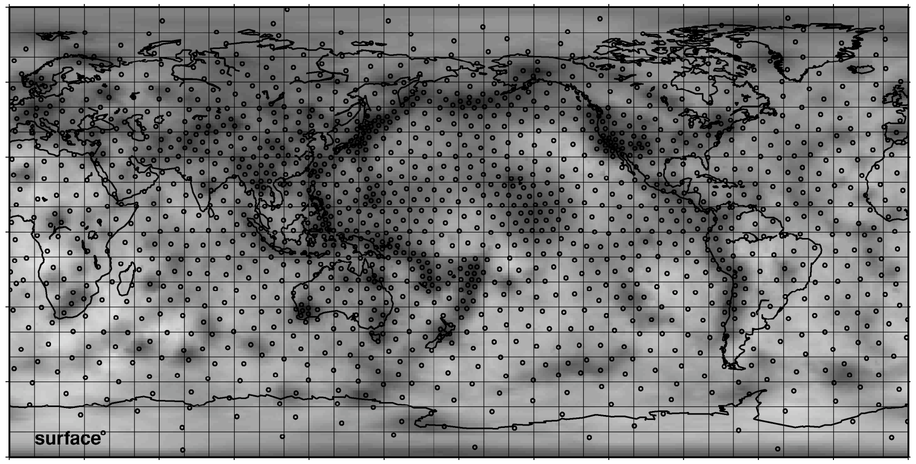
Gudmundsson & Sambridge (1998)



Sambridge & Rawlinson (2005)





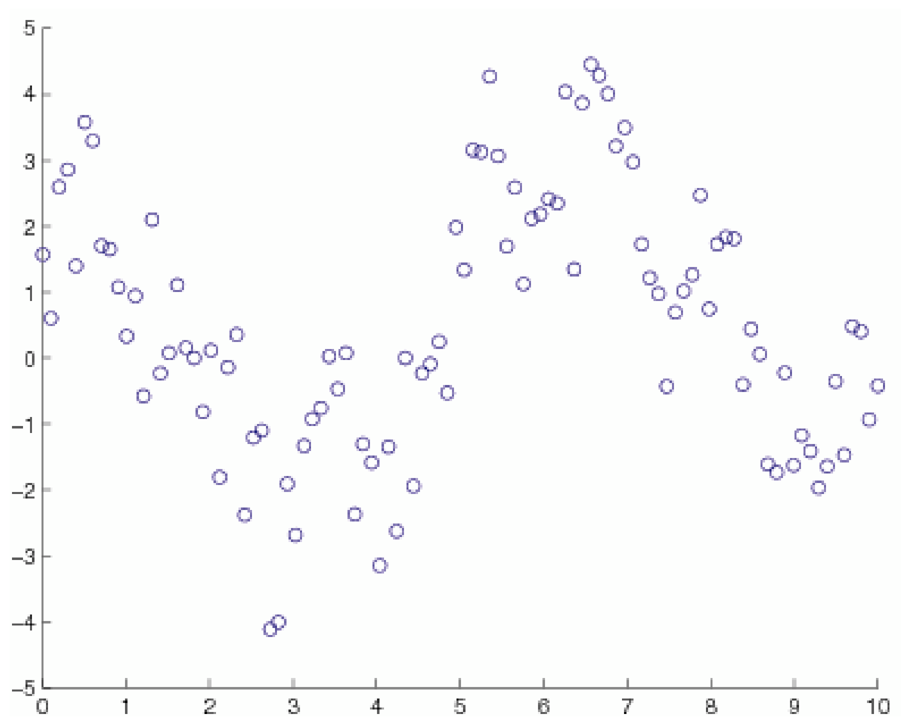


Nolet & Montelli (2005)

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).

Partition Modelling

A Bayesian MCMC based technique used for classification and Regression problems in Medical Statistics

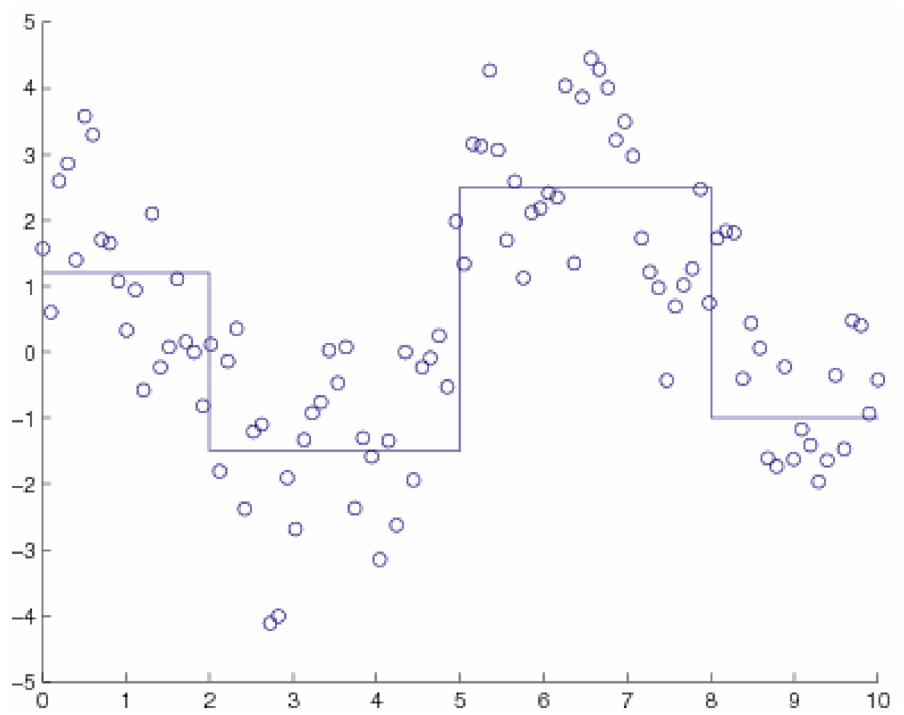


What is the underlying function ?

How many unknowns ?

Partition Modelling

A Bayesian MCMC based technique used for classification and Regression problems in Medical Statistics

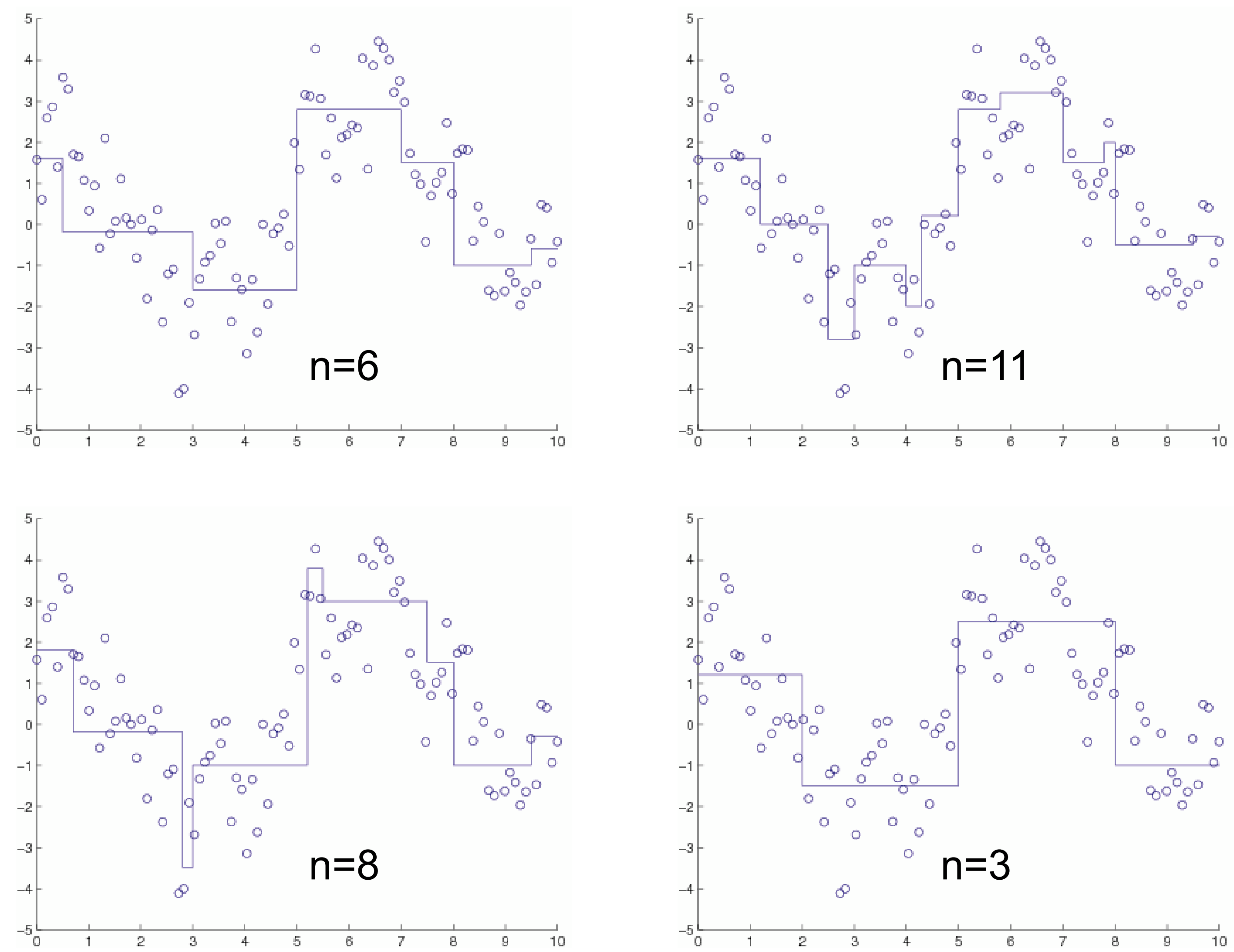


What is the underlying function ?

How many unknowns ?

Partition Modelling

Dynamically fit the height and the position of discontinuity



Randomized sampling of partition models within a Bayesian Framework

Mean solution

True solution

Data driven smoothing !

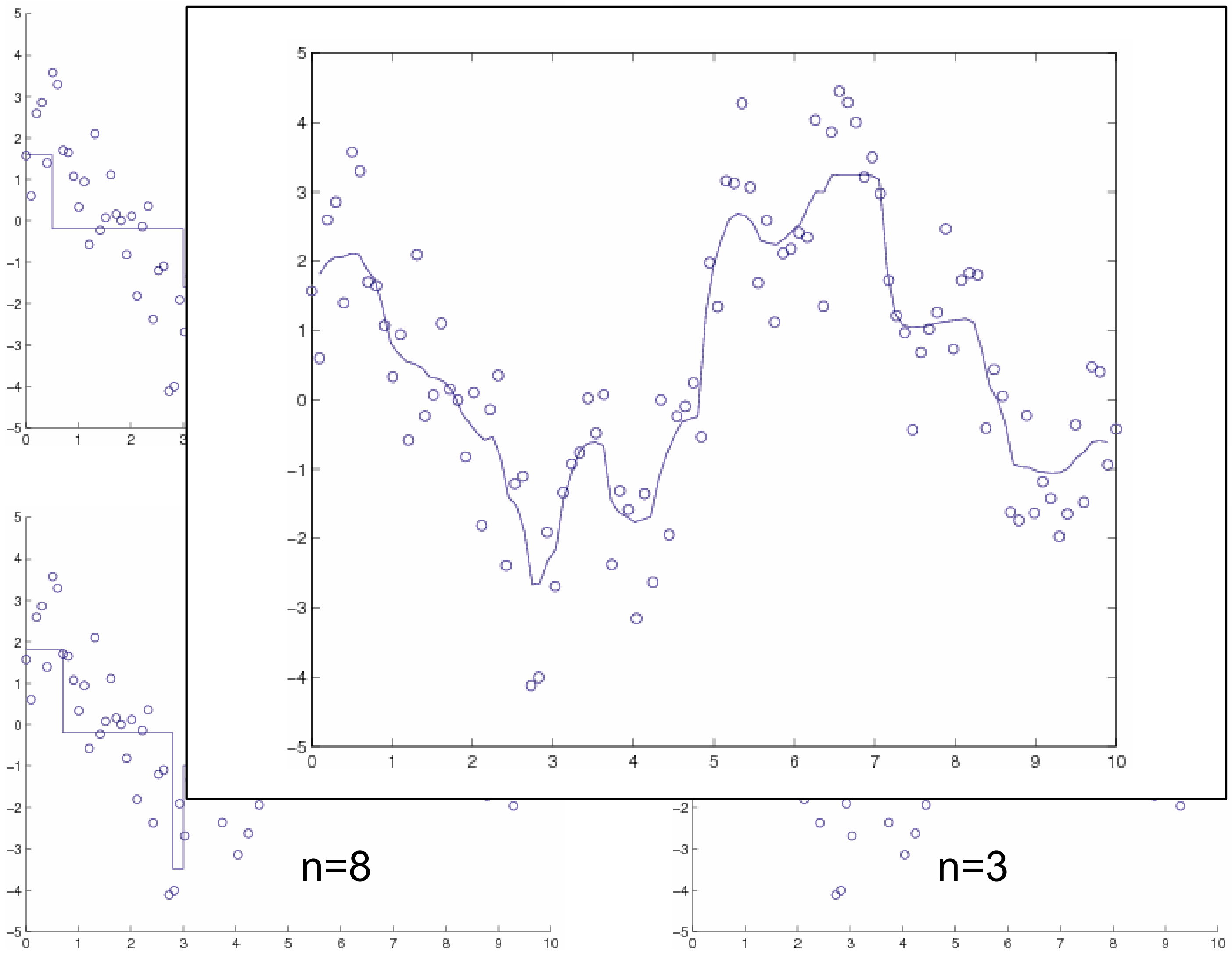
No explicit regularization

Self adaptive

Can we apply these concepts to tomography ?

Partition Modelling

Dynamically fit the height and the position of discontinuity



Randomized sampling of partition models within a Bayesian Framework

Mean solution

True solution

Data driven smoothing !

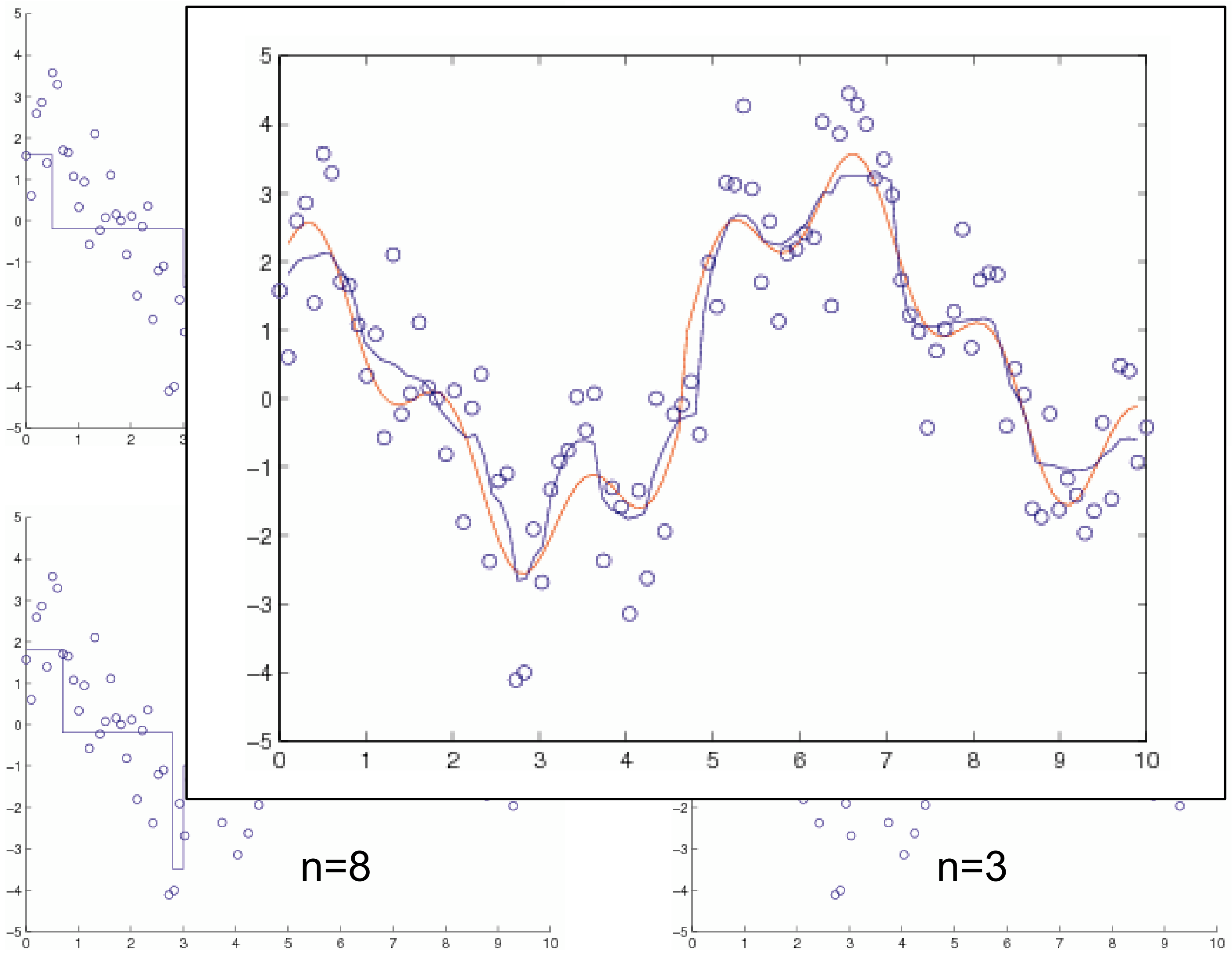
No explicit regularization

Self adaptive

Can we apply these concepts to tomography ?

Partition Modelling

Dynamically fit the height and the position of discontinuity



Randomized sampling of partition models within a Bayesian Framework

Mean solution

True solution

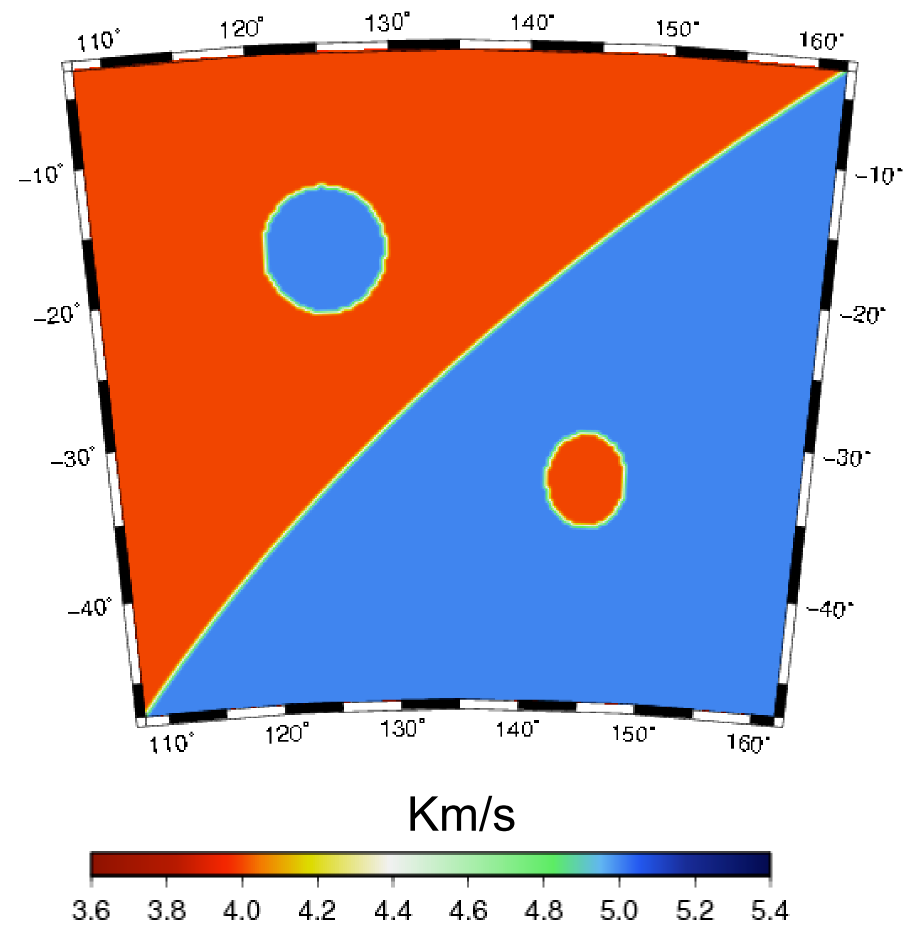
Data driven smoothing !

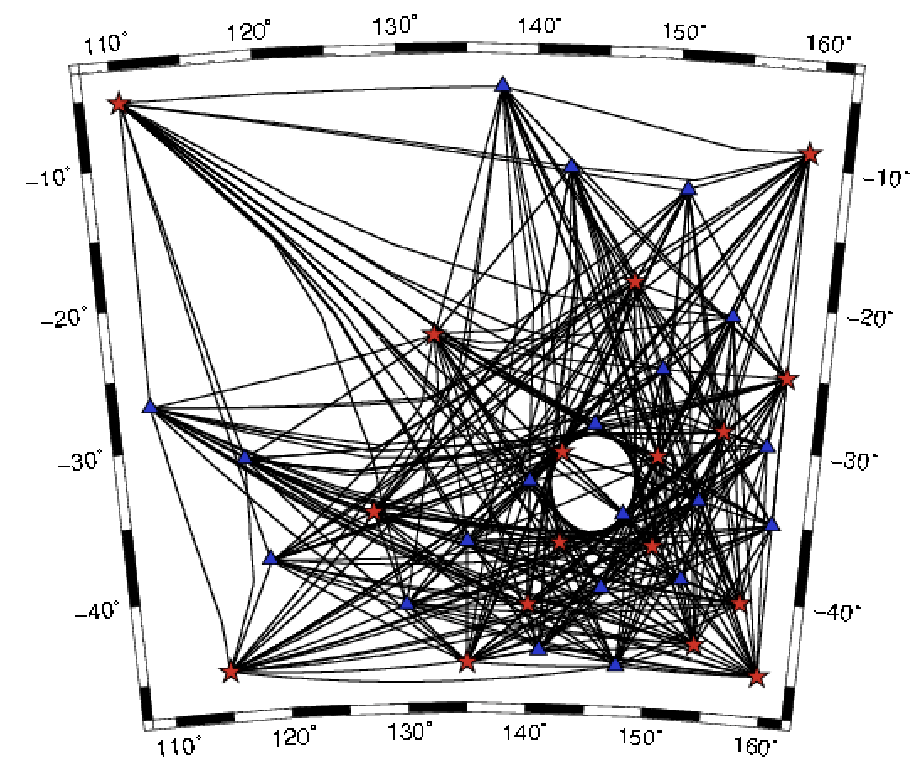
No explicit regularization

Self adaptive

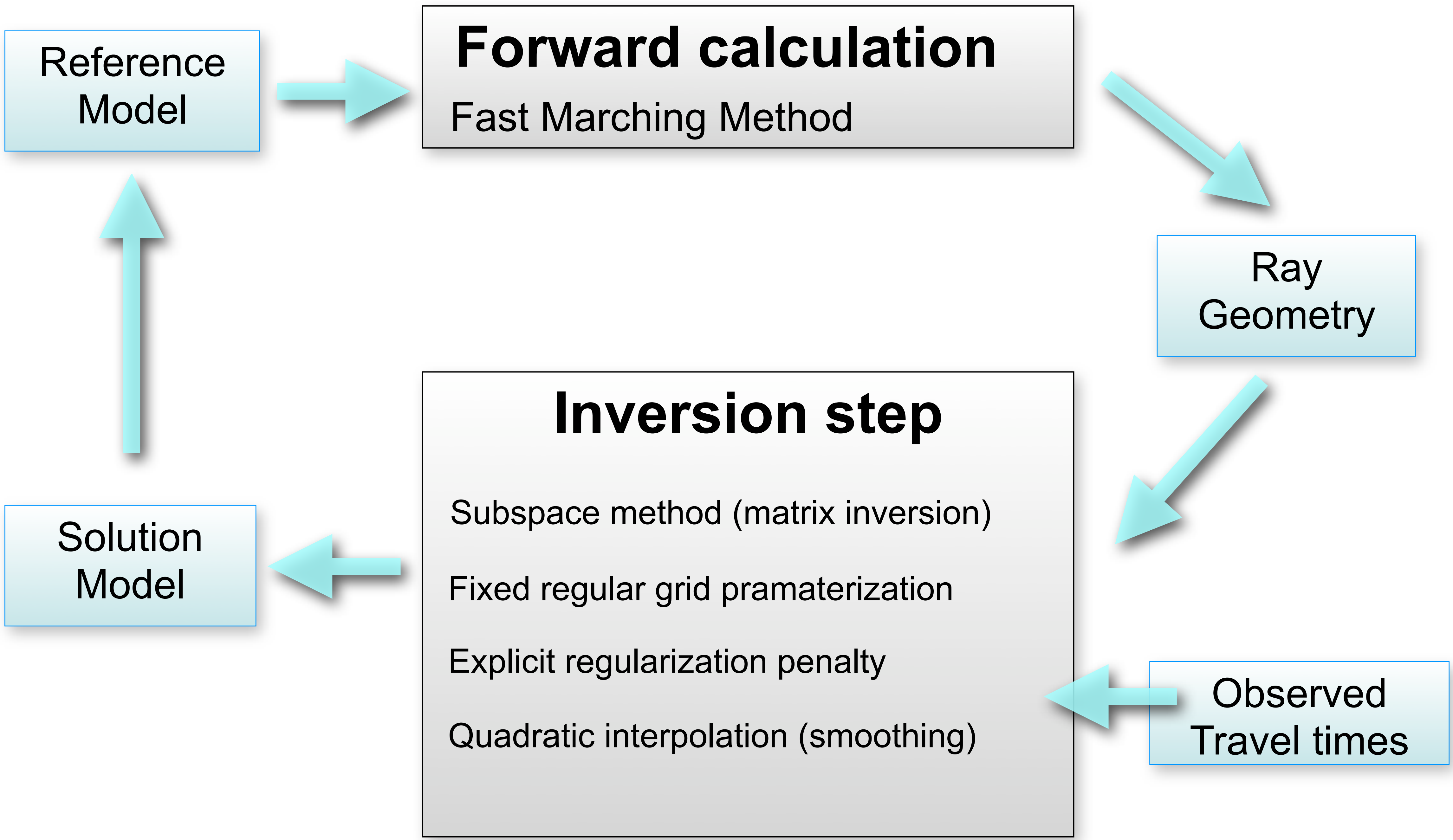
Can we apply these concepts to tomography ?

A simple tomographic test problem





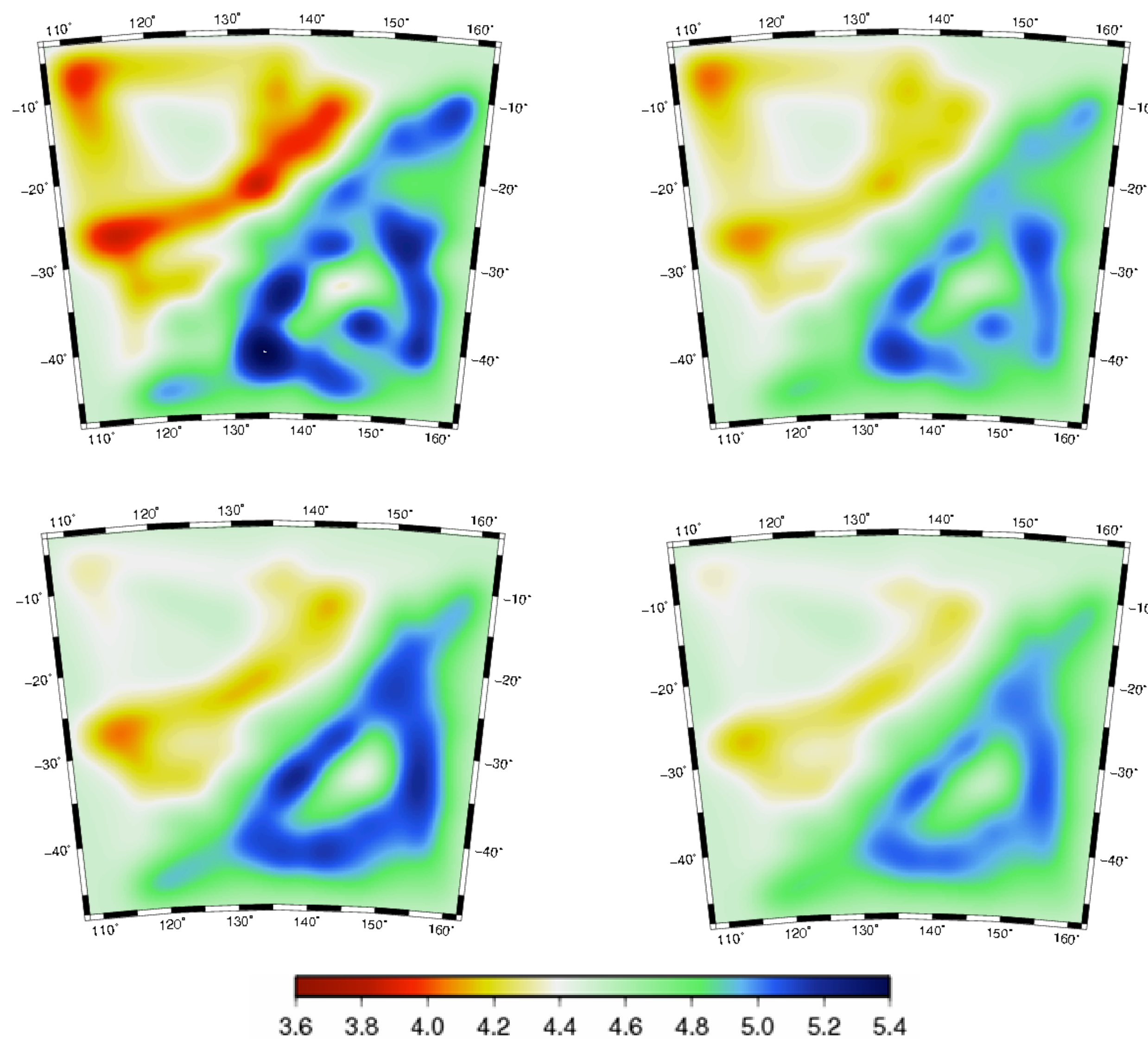
Iterative linearized inversion



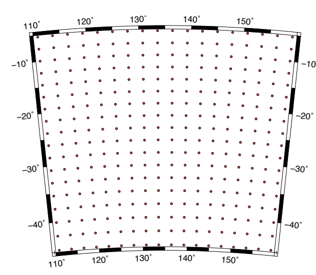
Tikhonov solutions

Damping

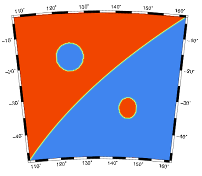
Smoothing

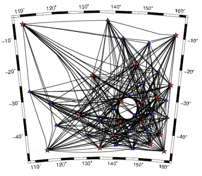


Standard
Fixed grid
tomography



20 x 20
B-splines nodes

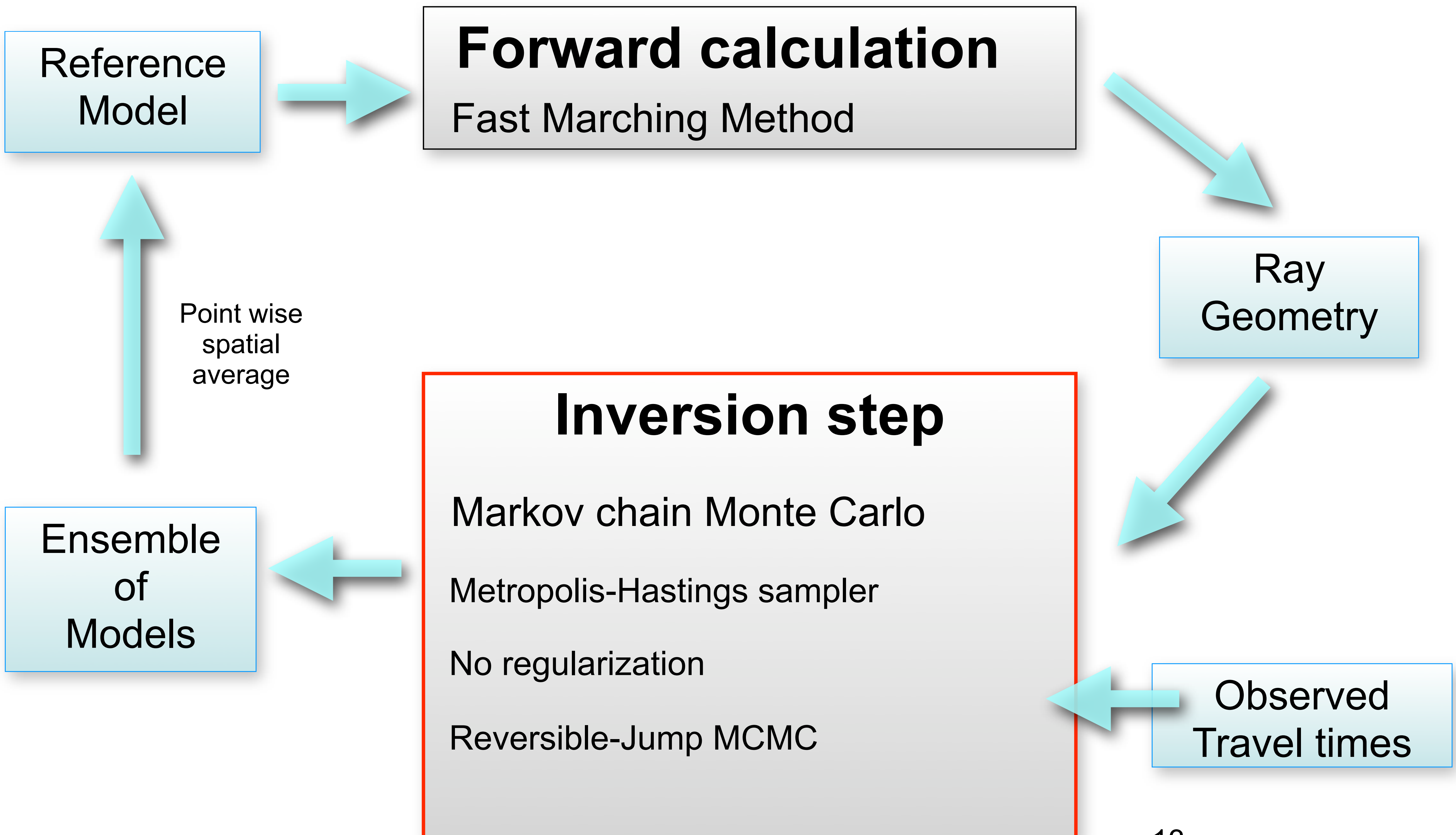




$$\min \quad ||G\mathbf{m} - \mathbf{d}||_2^2 + \alpha^2 ||L\mathbf{m}||_2^2$$

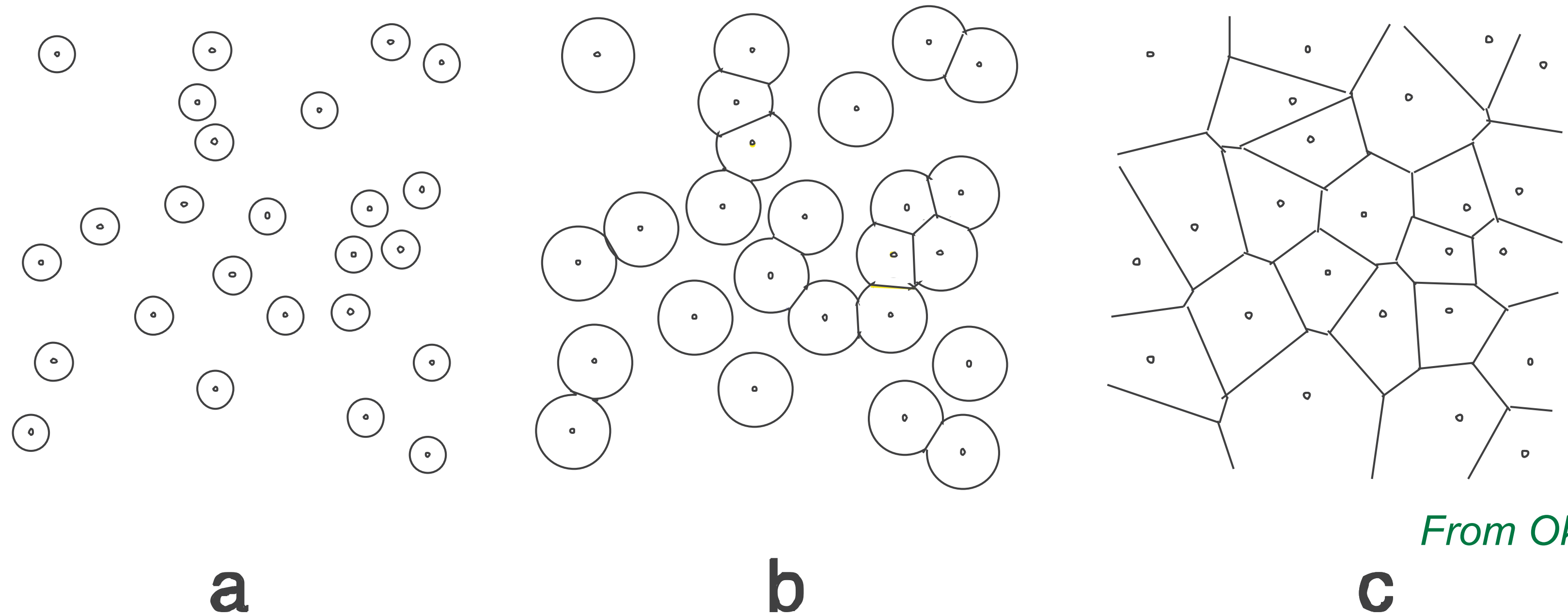
$$\mathbf{m} = (G^T G + \alpha^2)^{-1} G^T \mathbf{d}$$

Partition modelling applied to linearized tomography



Voronoi cells for unstructured meshes

2-D partitions = Voronoi cells



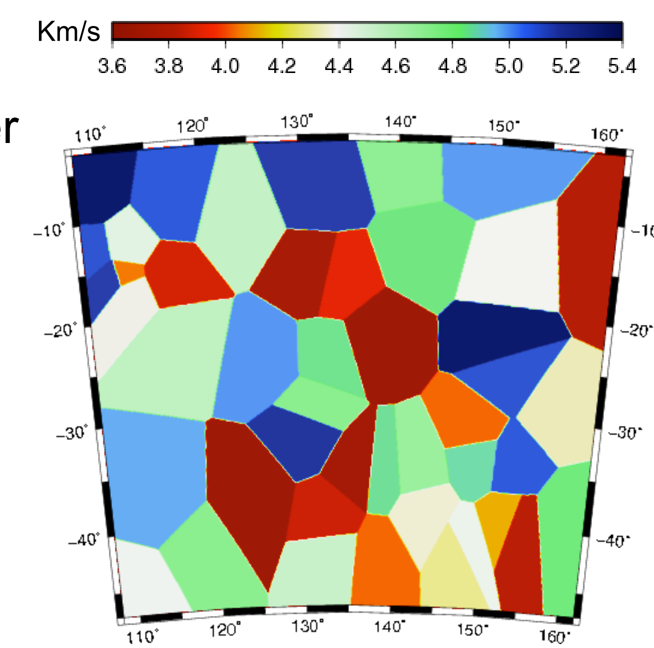
From Okabe et al. (1995)

Tomography parameterization

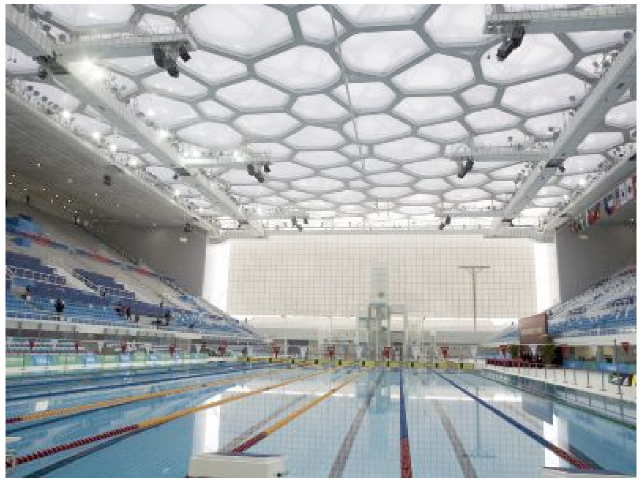
Each cell contains a slowness parameter

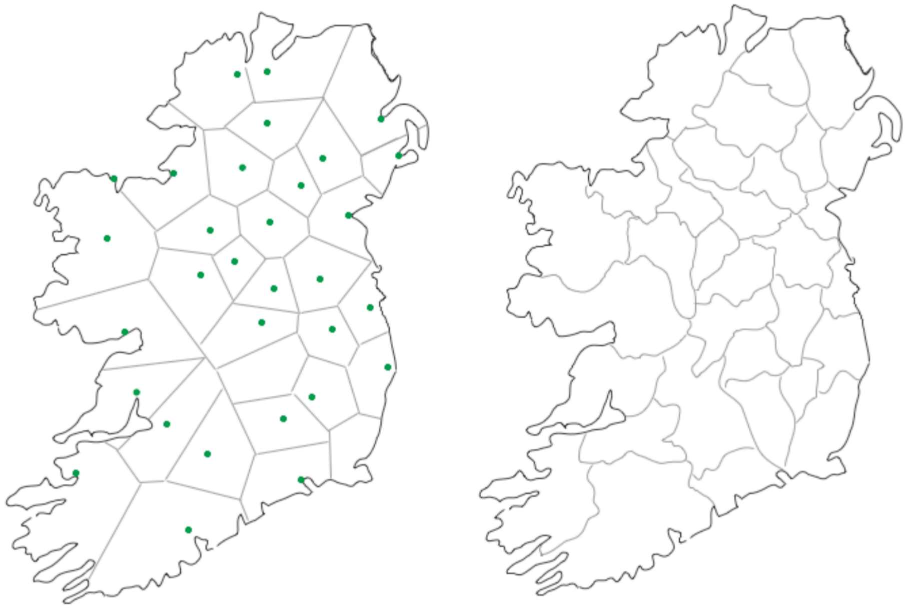
Cell nuclei can move position
(grow and shrink)

Cell nuclei can be added (birth)
or removed (death)



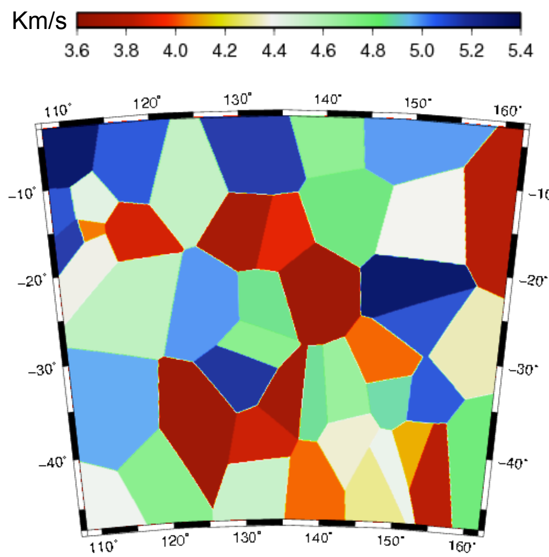
Voronoi cells are everywhere



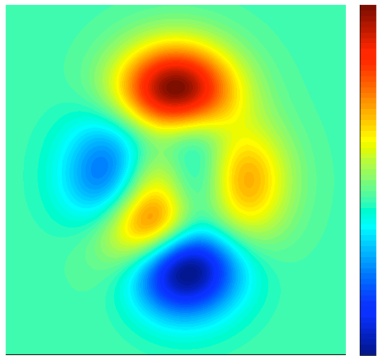


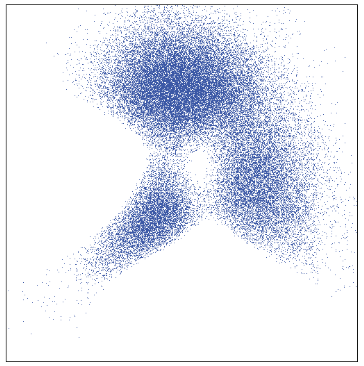
(From Okabe et. al. 1995, after Cox and Agnew, 1976)

Outline of general procedure



Will converge to sample from the Bayesian posterior Probability density function (PDF)





- With equal probability propose a new model using either:
  - cell birth,
  - cell death,
  - cell move,

    } $p = 1/2$
  - perturb slowness in randomly chosen cell $p = 1/2$

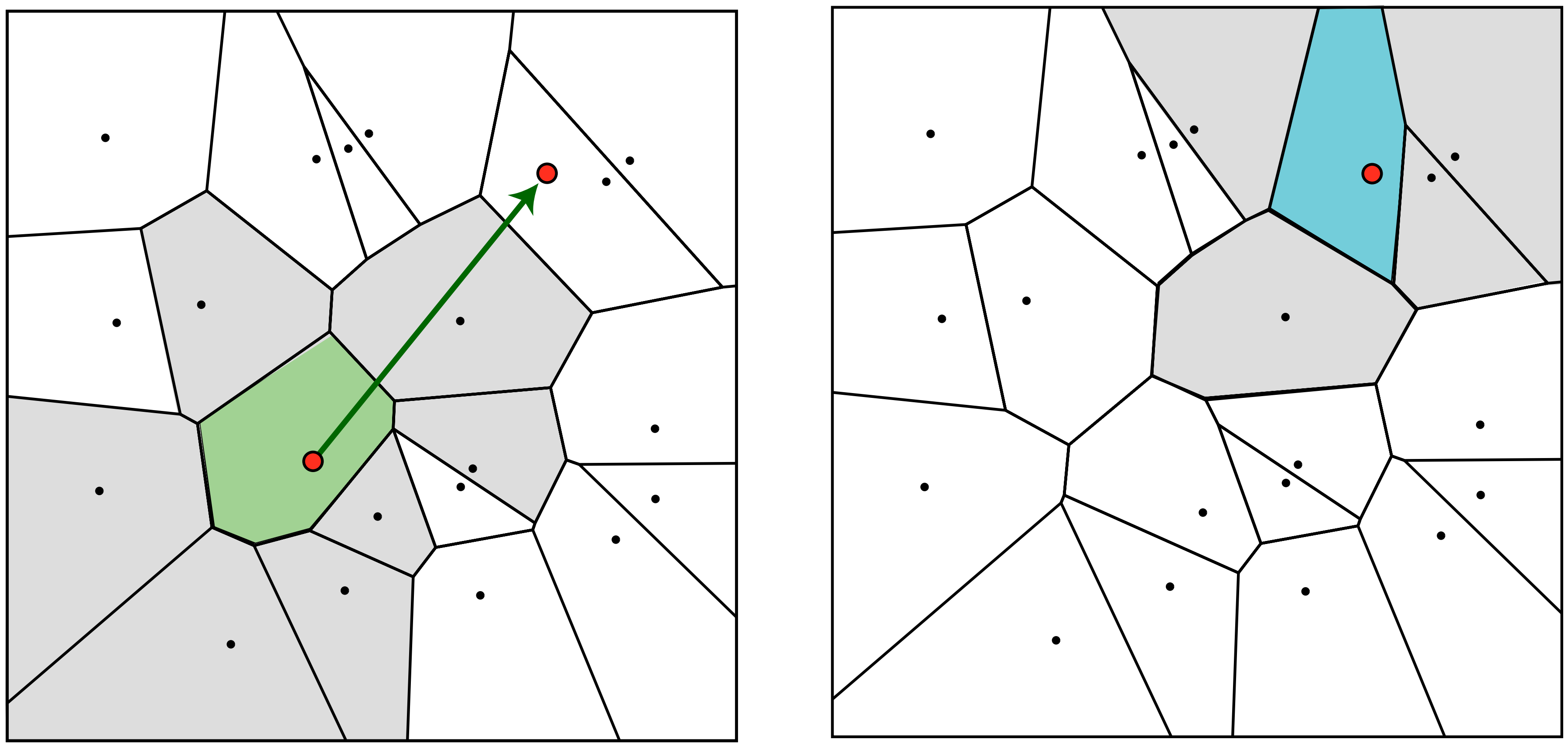
$$\mathbf{x} \rightarrow \mathbf{x}'$$

- Compute travel times
- Accept new model with probability, p:

$$p = \min\left\{1, \frac{p(\mathbf{d}|\mathbf{x}')\ p(\mathbf{x}')\ q(\mathbf{x}|\mathbf{x}')}{p(\mathbf{d}|\mathbf{x})\ p(\mathbf{x})\ q(\mathbf{x}'|\mathbf{x})}|J|\right\}$$

- Go to start

Moving Voronoi cells



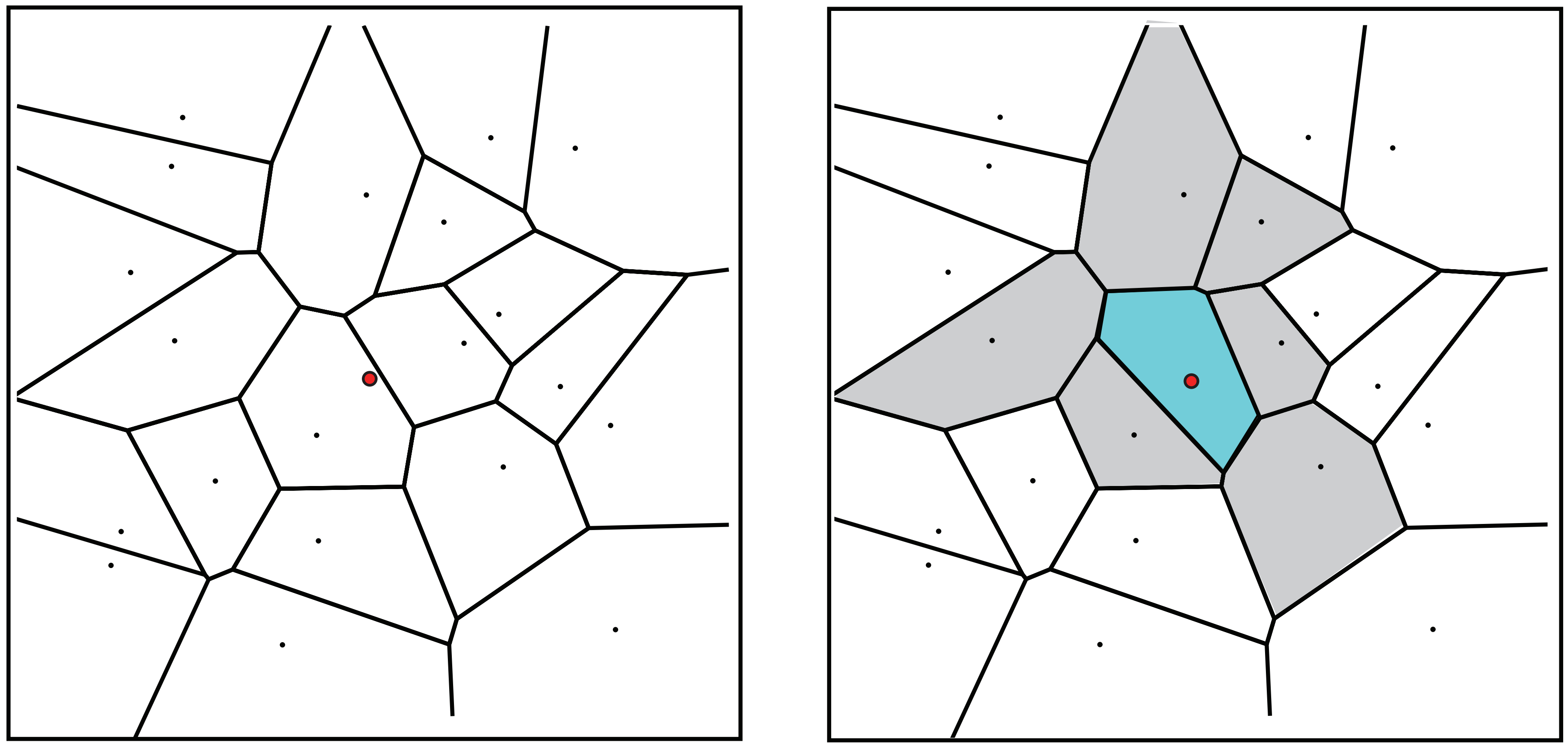
Move location of cell nucleus according to a chosen PDF
e.g. A Gaussian with chosen width

$$q(\mathbf{x}'|\mathbf{x}) \propto \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}')^T C_M^{-1}(\mathbf{x}-\mathbf{x}')\}$$

$$x_i' = x_i + \sigma_i \times N(0,1), \quad (i = 1, 2)$$

Changing the number of Voronoi cells

The birth step

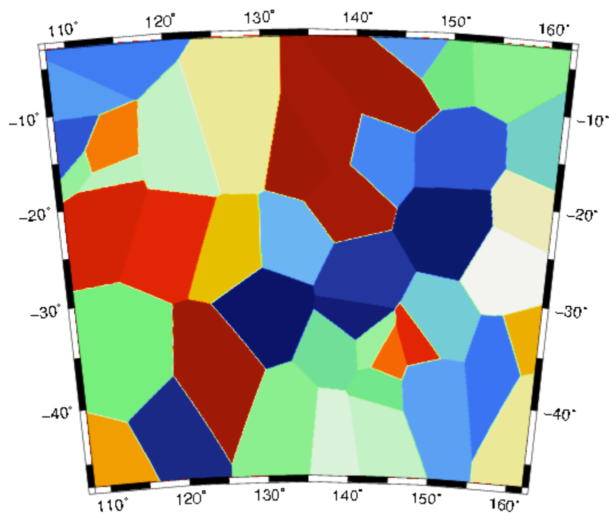


Generate location of cell nucleus according to a chosen PDF, e.g. the uniform prior

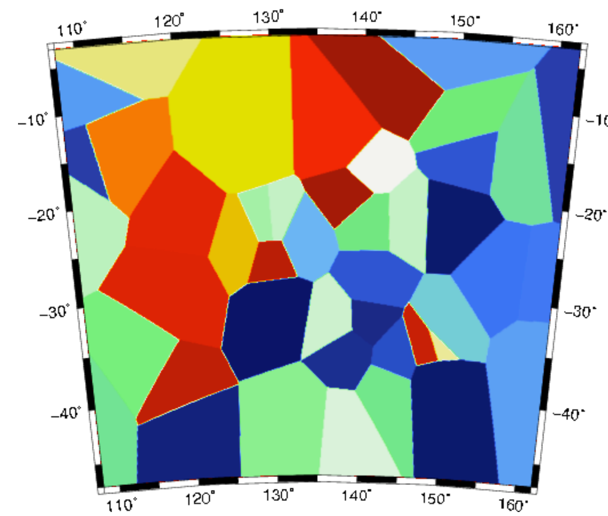
$$q(\mathbf{x}'|\mathbf{x}) \propto \frac{1}{\Delta x_1 \Delta x_2}$$

$$x_i' = x_{i,0} + \Delta x_i \times U(0,1), \quad (i = 1, 2)$$

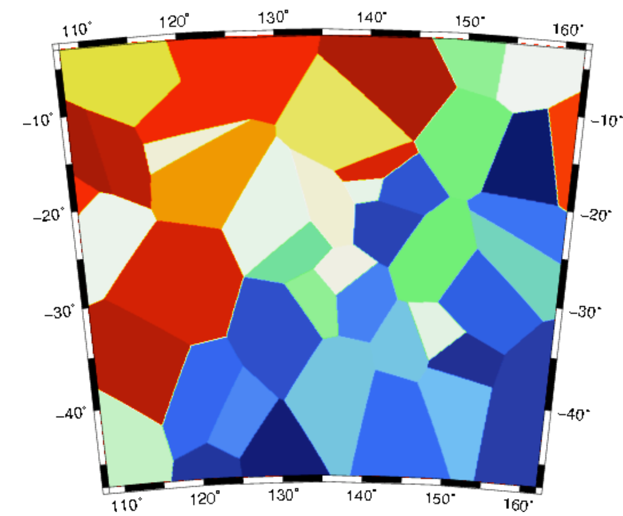
A convergent Markov chain



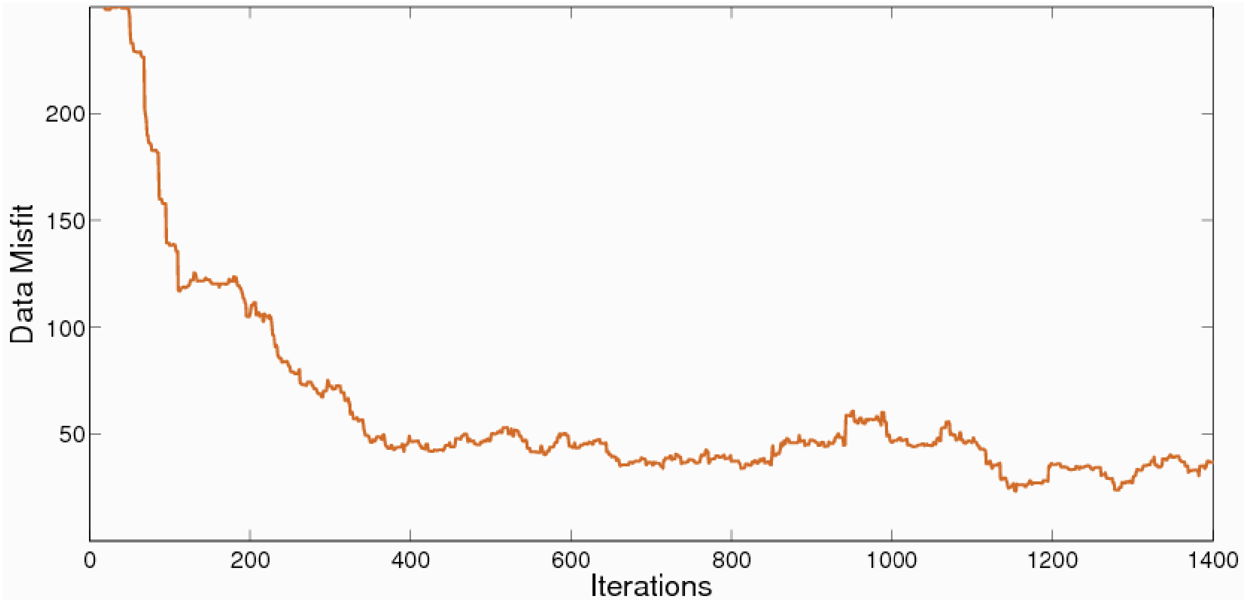
Step 150

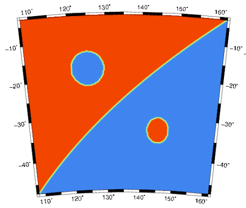


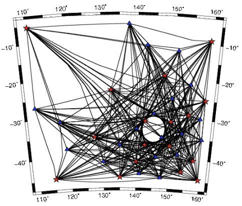
Step 300



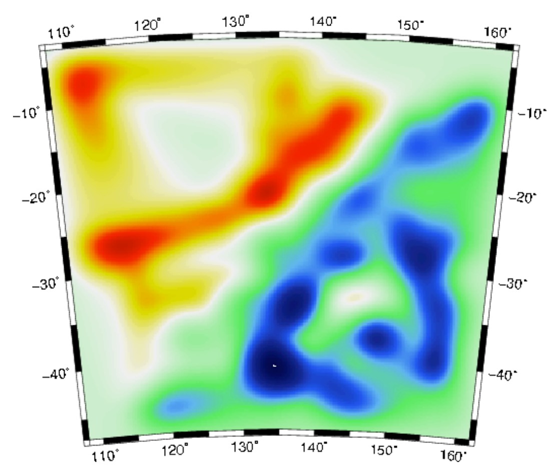
Step 1000

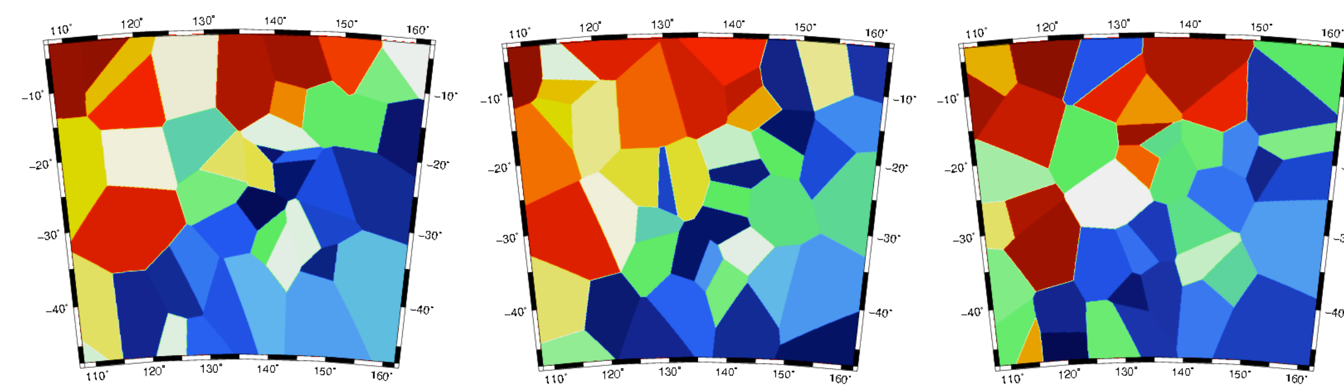




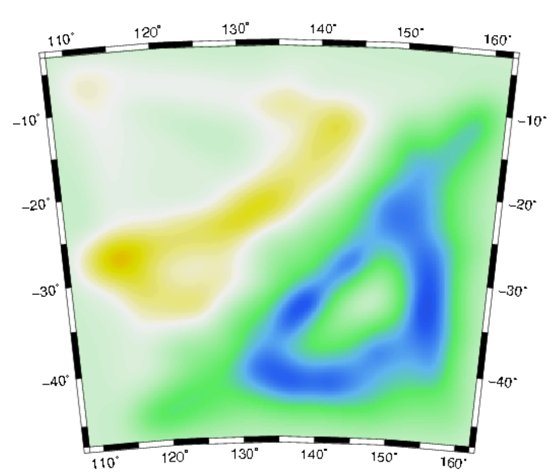


Fixed grid + Regularization vs Partition Modelling results



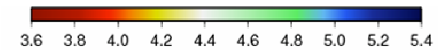


Best data fit model

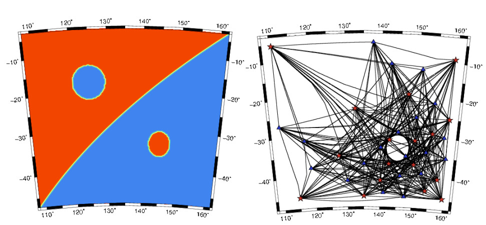


400 cells + regularization

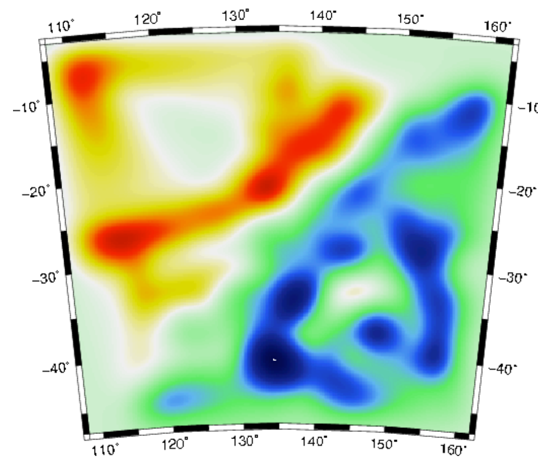
Information is contained in properties common to models in the ensemble

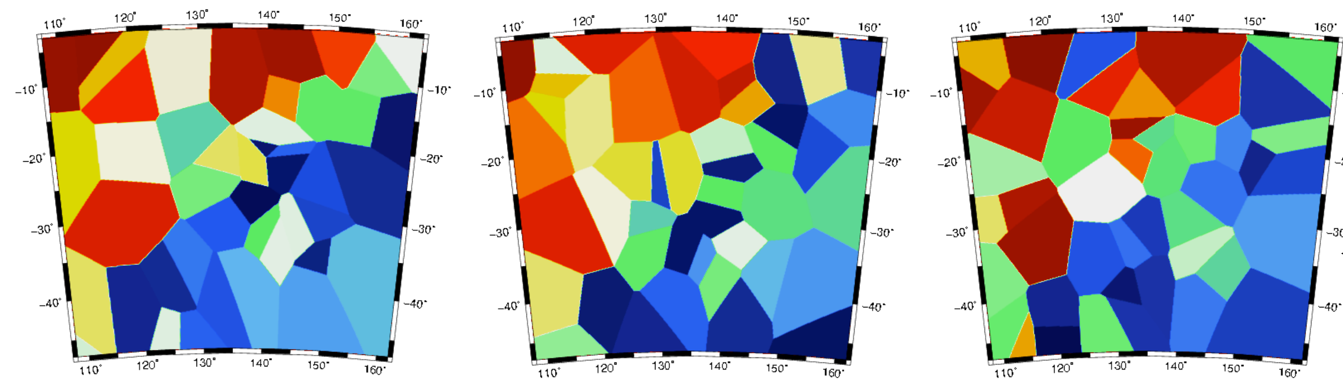


Details:
45 mobile cells
No birth,
No death
Linearized case

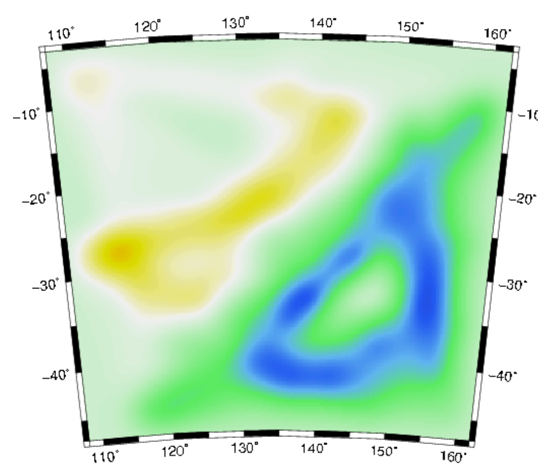


Fixed grid + Regularization vs Partition Modelling results

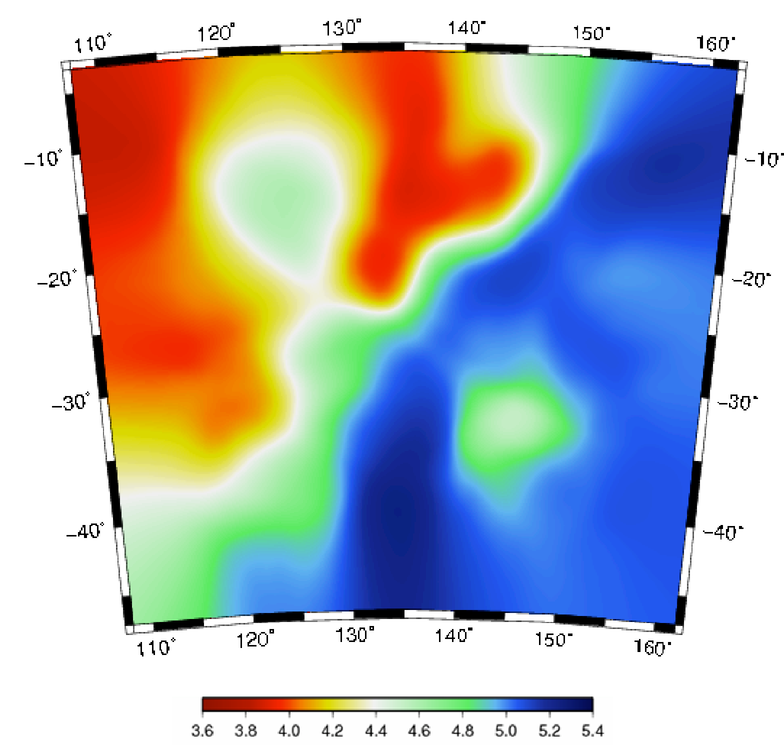




Best data fit model

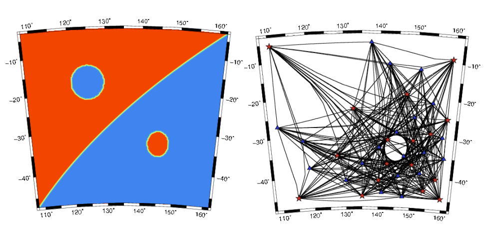


400 cells + regularization



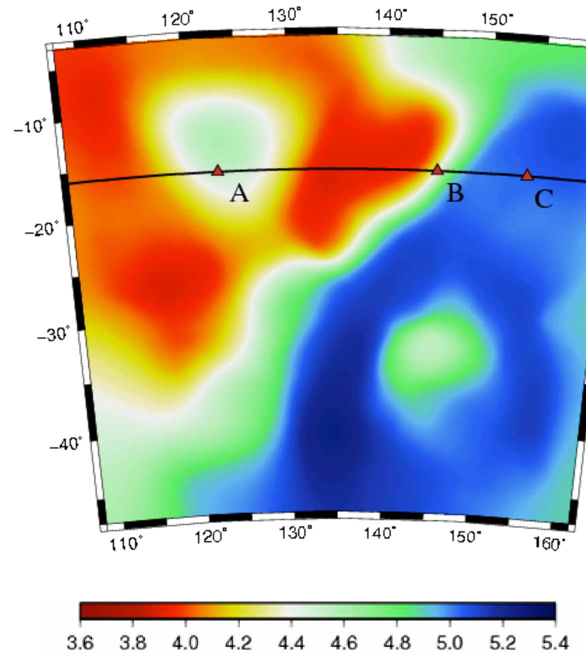
Mean model

Details:
45 mobile cells
No birth,
No death
Linearized case

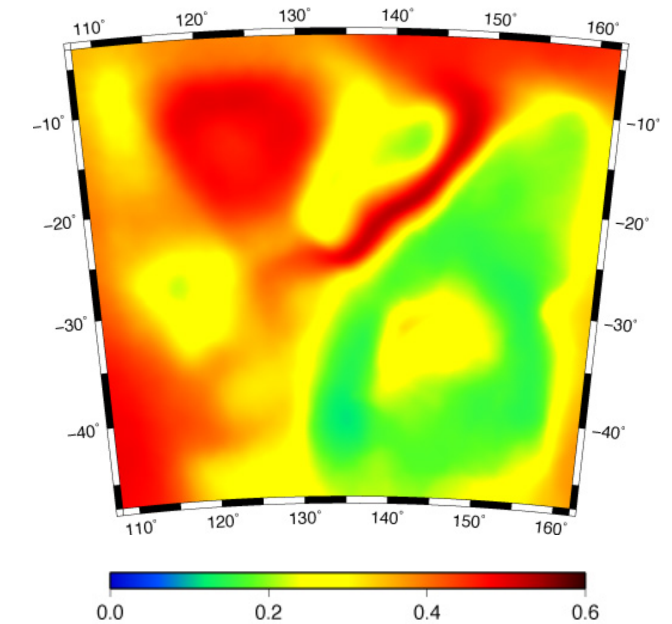


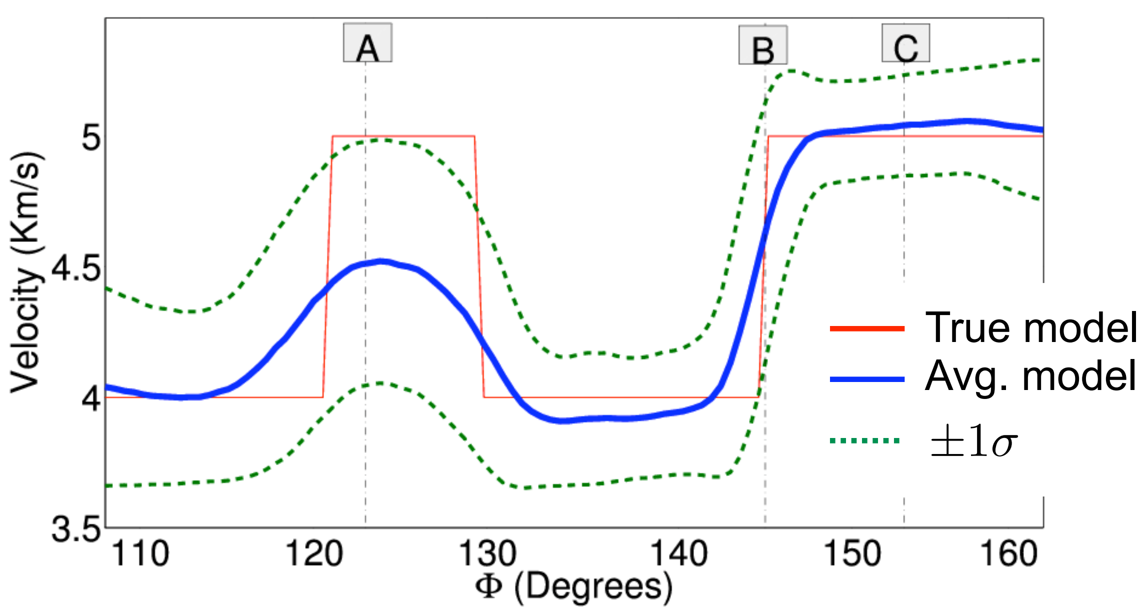
Model uncertainty

Ensemble average

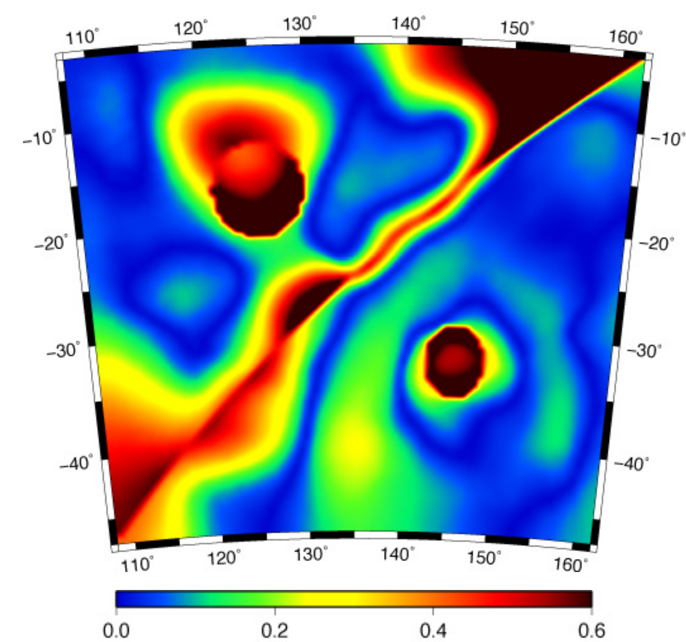


Ensemble standard deviation



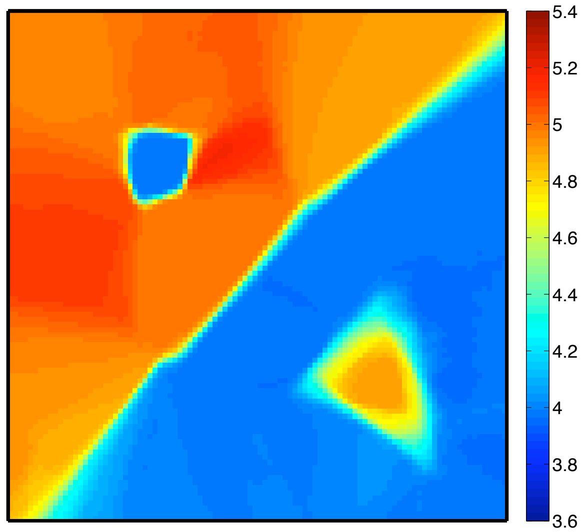


True absolute error

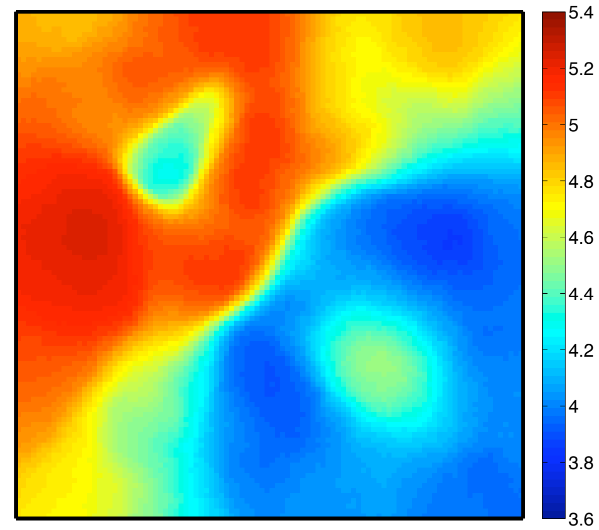


Including birth, death and noise

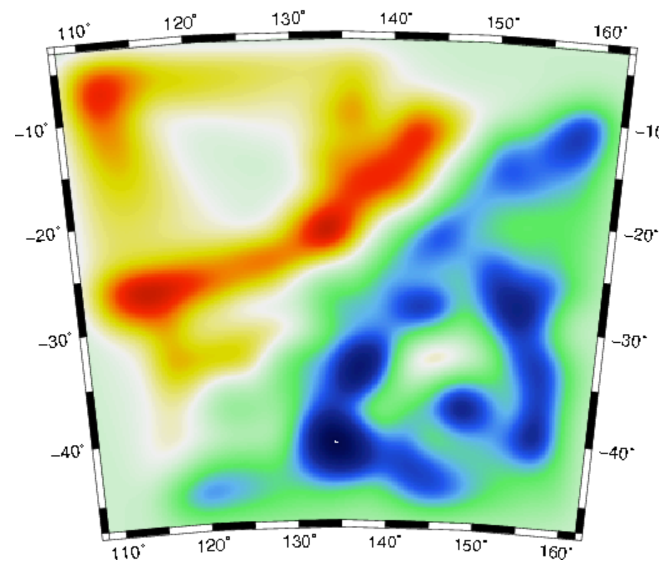
Spatially averaged models from ensemble



2 % data noise

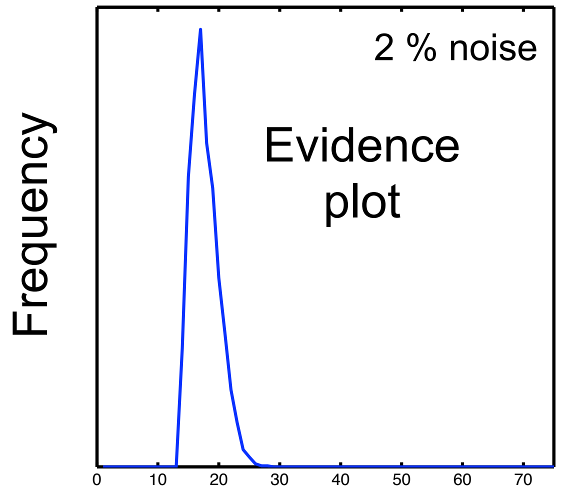


10 % data noise

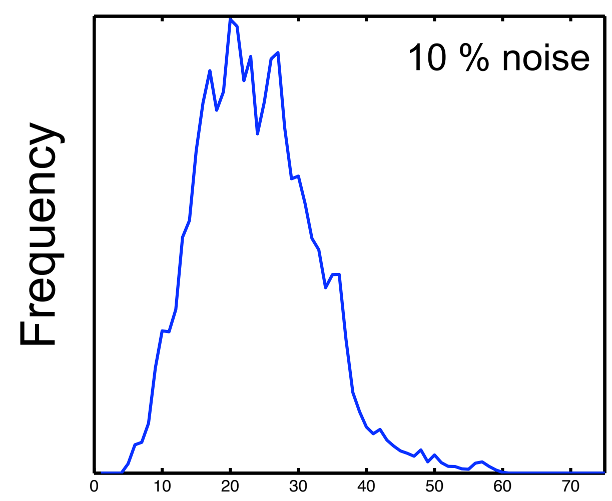


Optimal
Regularized solution

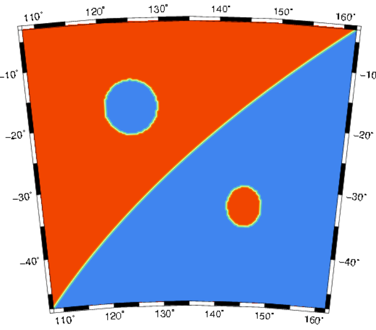
Histograms of number of cells from MCMC



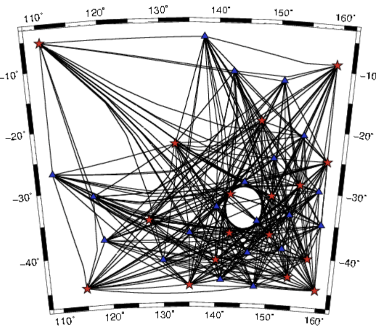
Number of cells



Number of cells



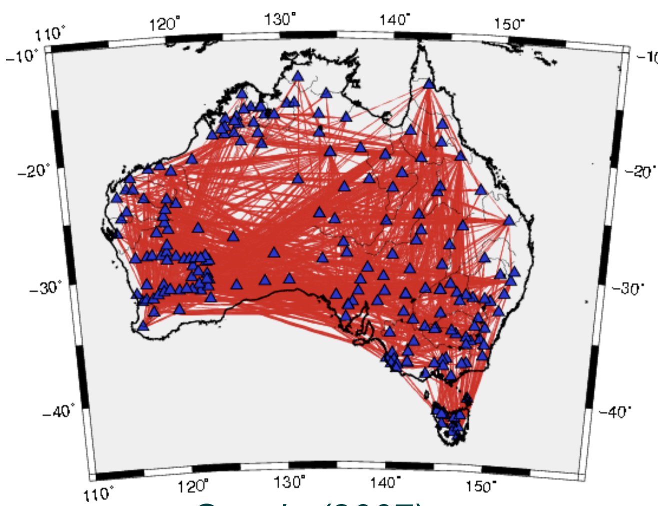
True model



Ray paths

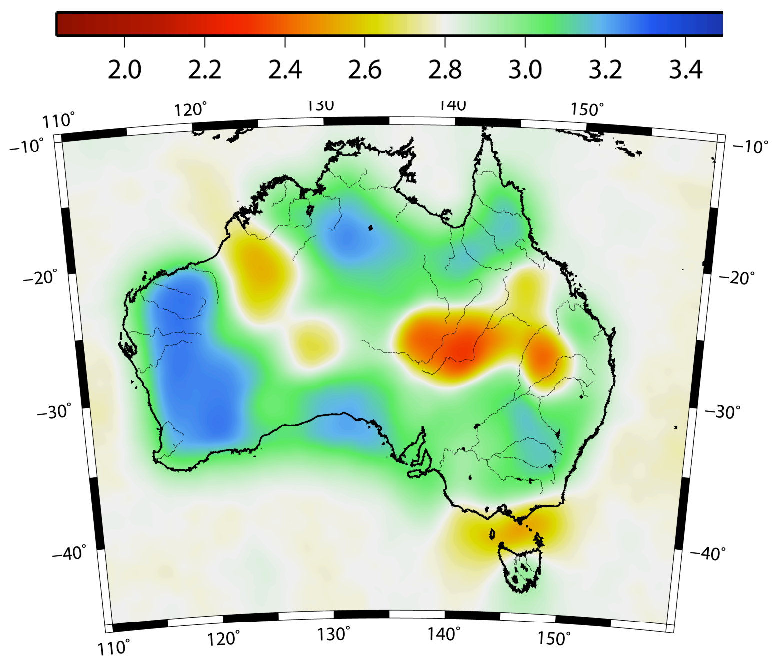
Application to ambient noise tomography is Australia

Cross correlation of
Ambient seismic noise
for Rayleigh wave
group velocity at 5s

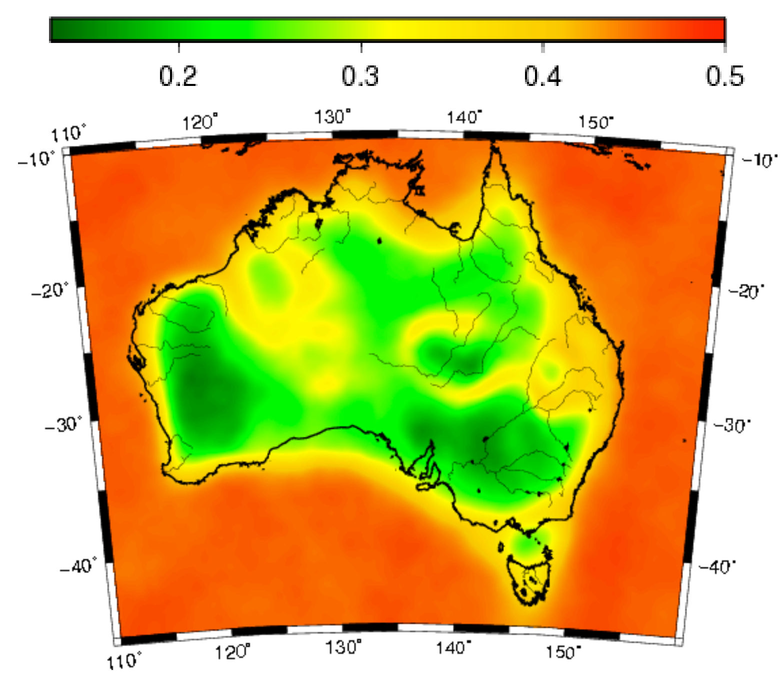


Saygin (2007)

Partition modelling tomography



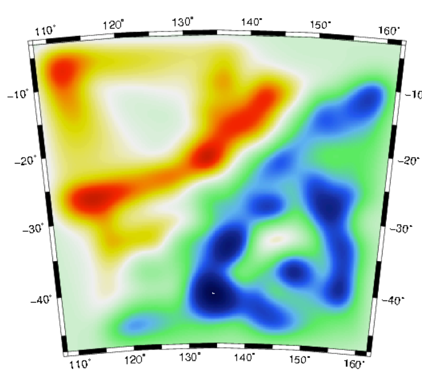
Average velocity field
of ensemble

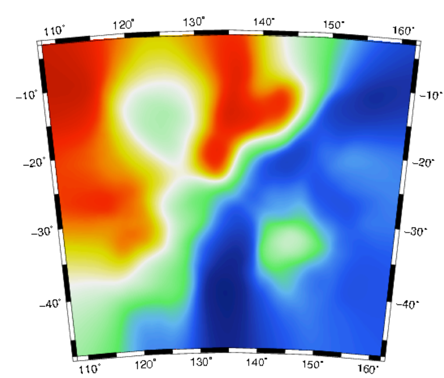


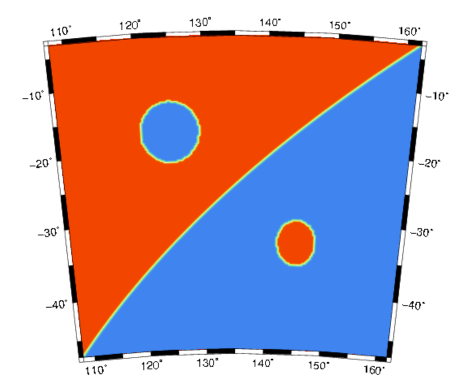
Standard deviation of ensemble

Summing up

- Inversion and Optimization are not the same thing
- Analysis of many candidate solutions can be better for interpretation than seeking single optimal (regularized) models.
- Computation not prohibitive (can employ same linearized approximations used in iterative schemes)







The End

But if you want more...