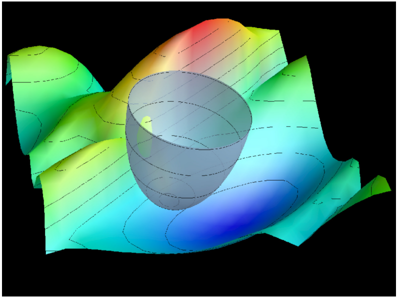
Seismic tomography: From damped least squares to transD

Nick Rawlinson
Bullard Labs
University of Cambridge

UNIVERSITY OF CAMBRIDGE



Acknowledgements: Malcolm Sambridge, Thomas Bodin, Simone Pilia, Mallory Young & Pierre Arroucau

In order to solve a linear or weakly non-linear inverse problem which is (a) under or mixed determined and (b) has data errors with a Gaussian distribution, one possible objective function is:

$$S(\mathbf{m}) = \Psi(\mathbf{m}) + \epsilon\Phi(\mathbf{m}) + \eta\Omega(\mathbf{m})$$

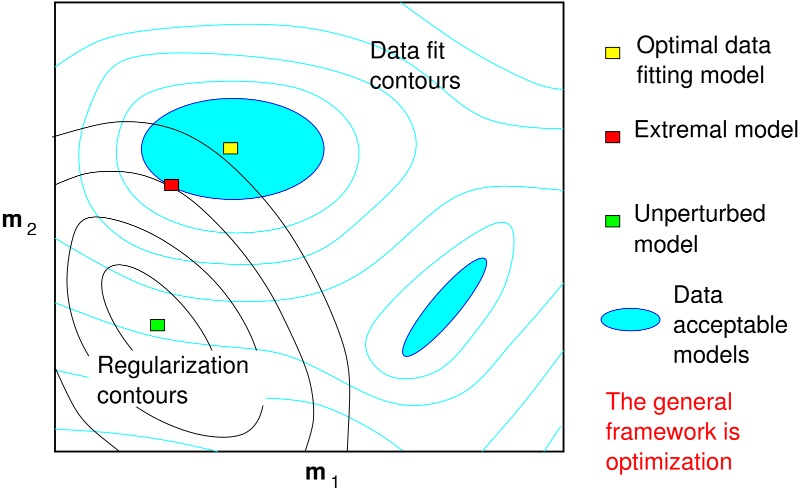
where

$$\Psi(\mathbf{m}) = (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})^T C_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})$$

$$\Phi(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_0)^T C_m^{-1} (\mathbf{m} - \mathbf{m}_0)$$

$$\Omega(\mathbf{m}) = \mathbf{m}^T D^T D\mathbf{m}$$

In general, we are faced with an ill-posed problem that has data noise and may be non-linear.



The general framework is optimization

Gauss-Newton:

$$\begin{aligned}\delta\mathbf{m}_n &= -[G_n^T C_d^{-1} G_n + \nabla_{\mathbf{m}} G_n^T C_d^{-1}(\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}) \\ &\quad + \epsilon C_m^{-1} + \eta D^T D]^{-1}[G_n^T C_d^{-1}[\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}] \\ &\quad + \epsilon C_m^{-1}(\mathbf{m}_n - \mathbf{m}_0) + \eta D^T D\mathbf{m}_n]\end{aligned}$$

Quasi-Newton:

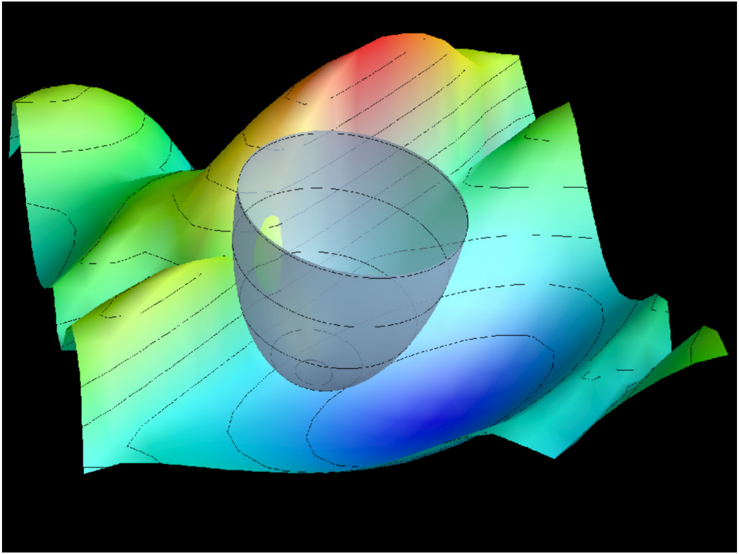
$$\begin{aligned}\delta\mathbf{m}_n &= -[G_n^T C_d^{-1} G_n + \epsilon C_m^{-1} + \eta D^T D]^{-1}[G_n^T C_d^{-1}[\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}] \\ &\quad + \epsilon C_m^{-1}(\mathbf{m}_n - \mathbf{m}_0) + \eta D^T D\mathbf{m}_n]\end{aligned}$$

Generalised subspace:

$$\delta\mathbf{m} = -A[A^T(G^T C_d^{-1} G + \epsilon C_m^{-1} + \eta D^T D)A]^{-1} A^T \hat{\gamma}$$

Damped and smoothed least squares:

$$\delta\mathbf{m} = [G^T C_d^{-1} G + \epsilon C_m^{-1} + \eta D^T D]^{-1} G^T C_d^{-1} \delta\mathbf{d}$$



Maximum likelihood or Stochastic inverse

$$\delta \mathbf{m} = [G^T C_d^{-1} G + C_m^{-1}]^{-1} G^T C_d^{-1} \delta \mathbf{d}$$

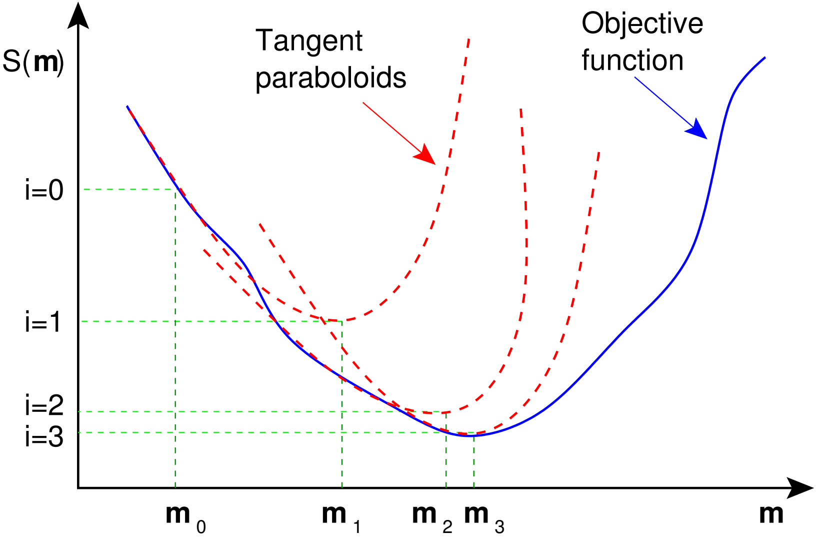
Damped least squares (DLS)

$$\delta \mathbf{m} = [G^T C_d^{-1} G + \epsilon C_m^{-1}]^{-1} G^T C_d^{-1} \delta \mathbf{d}$$

An equivalent approach is to find the least-squares solution of:

$$\begin{bmatrix} C_d^{-\frac{1}{2}} G \\ \sqrt{\epsilon} C_m^{-\frac{1}{2}} \\ \sqrt{\eta} D \end{bmatrix} \delta \mathbf{m} = \begin{bmatrix} C_d^{-\frac{1}{2}} \delta \mathbf{d} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

Application of SVD or iterative solvers like LSQR could be used to solve the equation as they can equally well be applied to non-square systems and will solve the equations in a least-squares sense.


S(**m**)
Tangent paraboloids
Objective function
i=0
i=1
i=2
i=3
$\mathbf{m}_0$
$\mathbf{m}_1$
$\mathbf{m}_2$
$\mathbf{m}_3$
m

Resolution and posterior covariance

$$R = G^{-g} G$$

$$R = [G^T C_d^{-1} G + \epsilon C_m^{-1} + \eta D^T D]^{-1} G^T C_d^{-1} G$$

$$C_M = G^{-g} C_d (G^{-g})^T$$

Synthetic reconstruction tests

Jackknife and Bootstrap

Linear and iterative non-linear sampling

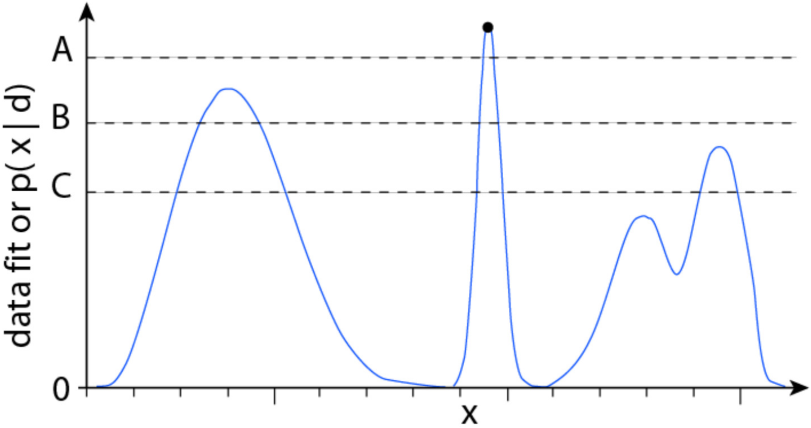
Strengths

- Can tackle very large problems (10s of millions of data measurements, millions of unknowns).
- Quantitative and qualitative estimates of posterior covariance and resolution relatively simple to generate.

Weaknesses

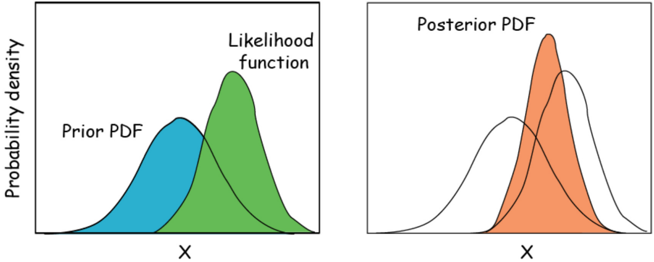
- Only applies to linear or weakly non-linear inverse problems
- Solution non-uniqueness a major issue
- Regularisation tends to be ad hoc and methods for assessing solution uncertainty often of limited value.

Multi-modal data misfit/objective function.



What value is there in an optimal model?

All information is expressed in terms of a probability density function.

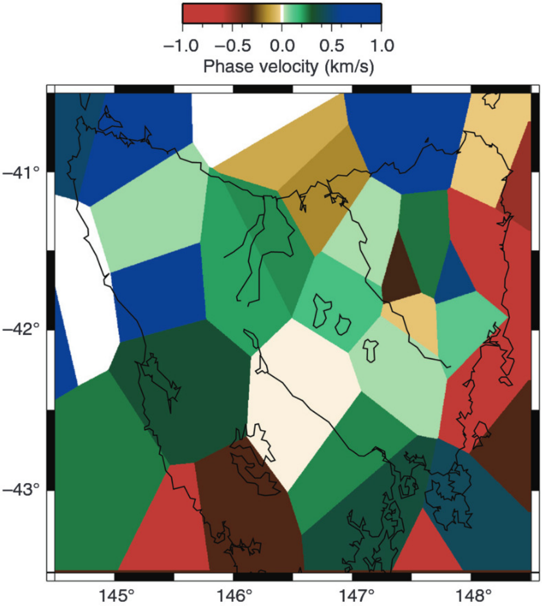


Bayes' rule (1763)

$$p(\mathbf{m}|\mathbf{d}, I) \propto p(\mathbf{d}|\mathbf{m}, I) \times p(\mathbf{m}|I)$$

Posterior probability density $\propto$ Likelihood $\times$ Prior probability density

Transdimensional tomography
Typical parameterization



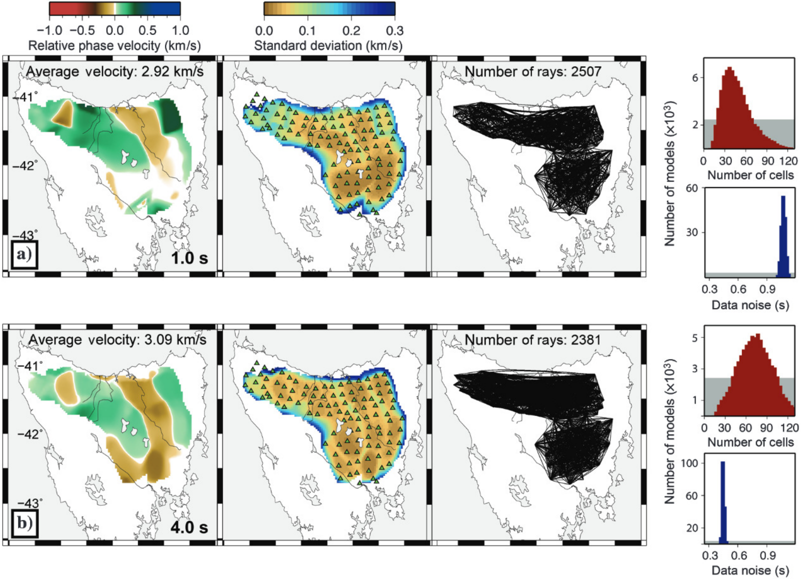
Unknowns to be inverted for include:

- Velocity: Constant velocity value in each cell
- Location: Coordinates of all Voronoi nodes
- Number of parameters: Number of Voronoi cells
- Data errors: Hyper-parameters defined by $C_d = f(h_1, h_2, ...)$. The simplest case is $h_1 = \sigma$, the standard deviation of the data errors.

In order to sample the posterior PDF, we use a variant of the Metropolis-Hastings algorithm commonly referred to as reversible-jump Markov chain Monte Carlo (rj-McMC).

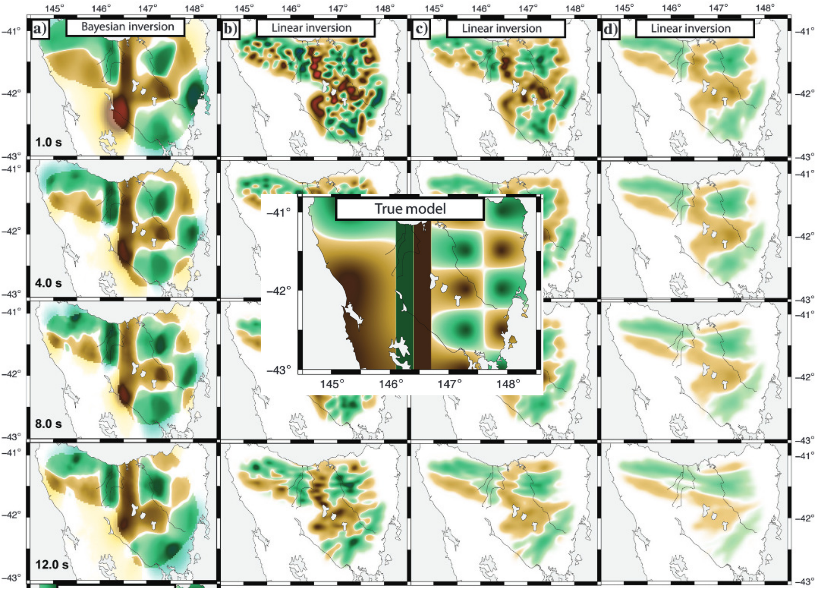
In the case of traveltime tomography, ray geometries can be updated infrequently for weakly non-linear problems, or frequently for fully non-linear problems.

Transdimensional tomography
Example - Tasmania

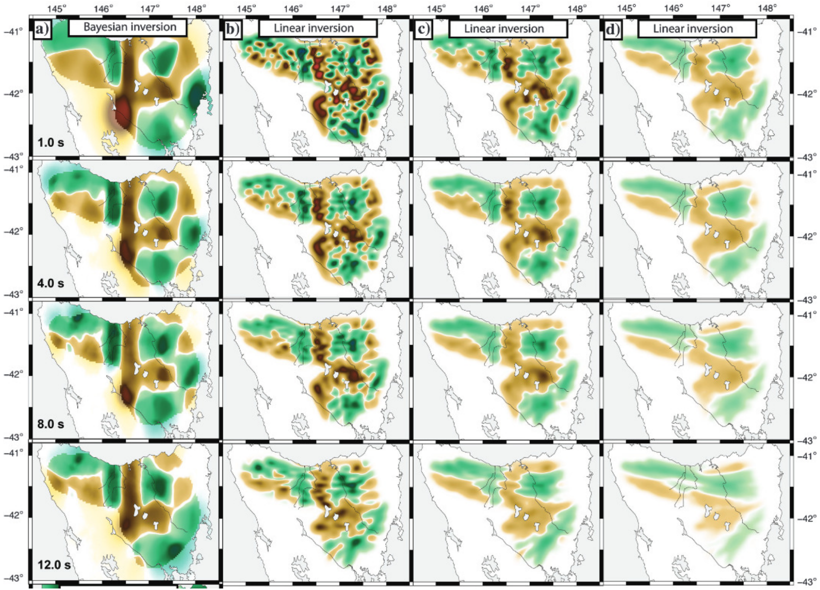


Transdimensional tomography

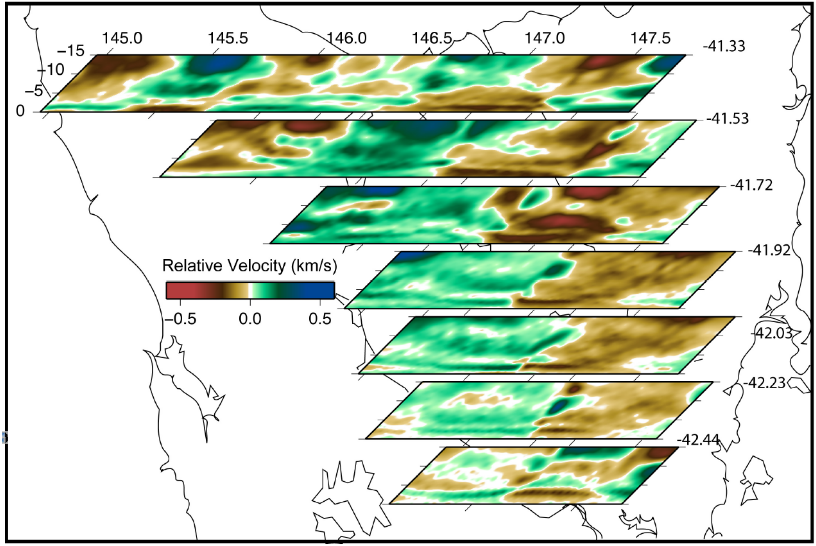
Example - Tasmania

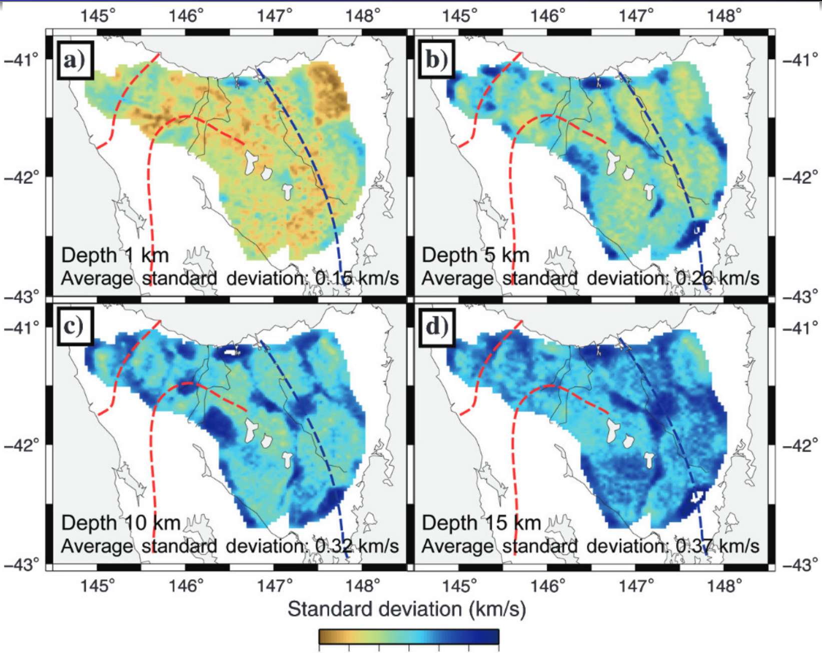


Transdimensional tomography
Example - Tasmania



Transdimensional tomography
Example - Tasmania

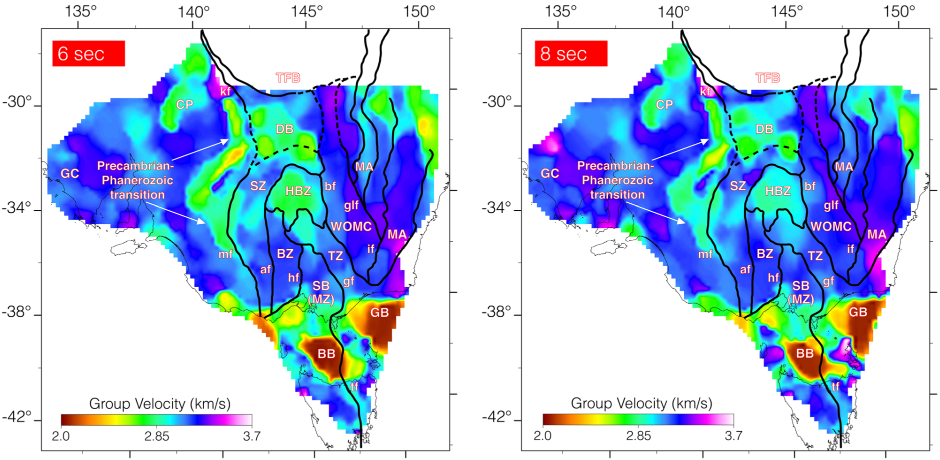




a)
Depth 1 km
Average standard deviation: 0.15 km/s
b)
Depth 5 km
Average standard deviation: 0.26 km/s
c)
Depth 10 km
Average standard deviation: 0.32 km/s
d)
Depth 15 km
Average standard deviation: 0.37 km/s
145°
146°
147°
148°
−41°
−42°
−43°
Standard deviation (km/s)

Transdimensional tomography

Example - SE Australia



References

- http://www.iearth.org.au/codes/ for transD software
- Bodin, T., & Sambridge, M. 2009. Seismic tomography with the reversible jump algorithm. Geophysical Journal International, 178, 1411-1436.
- Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P. 2012. Transdimensional tomography with unknown data noise. Geophysical Journal International, 189, 1536-1556.
- Young, M.K., Cayley, R.A., McLean, M.A., Rawlinson, N., Arroucau, P. & Salmon, M. 2013. Crustal Structure of the east Gondwana margin in southeast Australia revealed by transdimensional ambient seismic noise tomography. Geophysical Research Letters, 40, 4266-4271.
- Pilia, S., Rawlinson, N., Cayley, R. A., Bodin, T., Musgrave, R., Reading, A. M., Direen, N. G. & Young, M. K. 2015. Evidence of micro-continent entrainment during crustal accretion. Scientific Reports, 5, 8218.