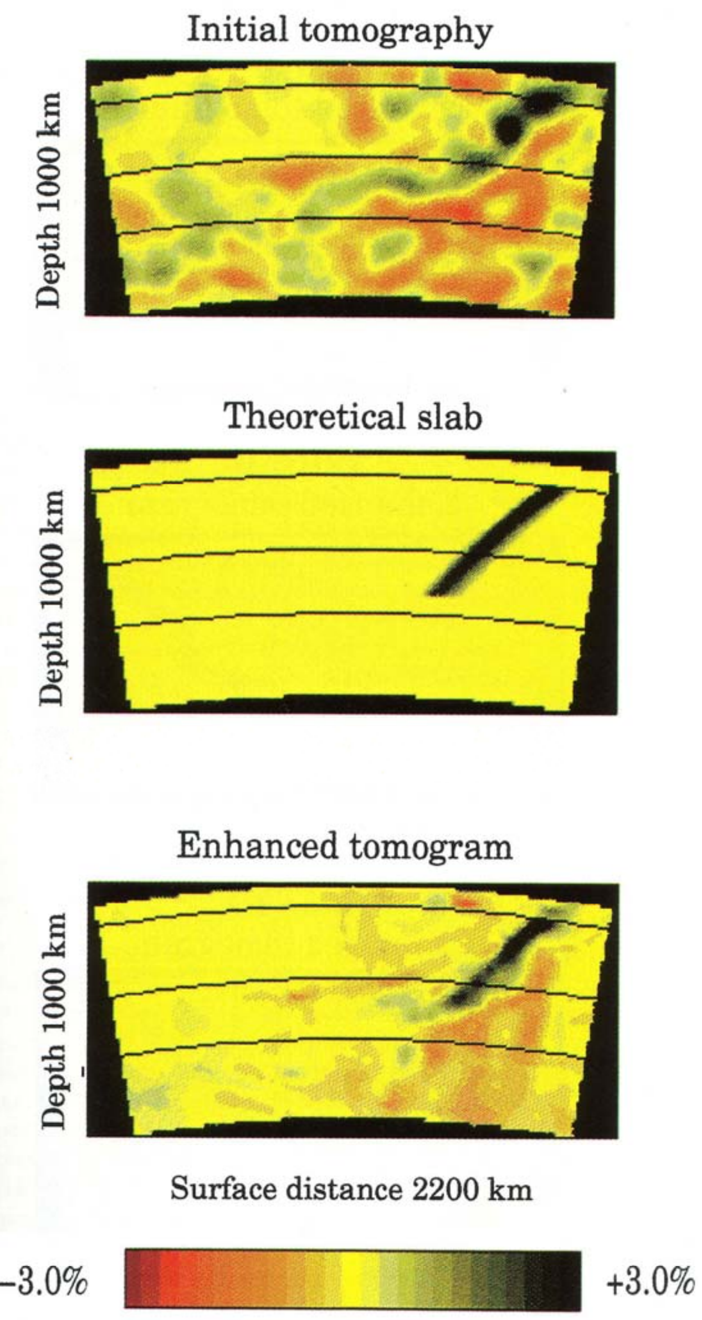
Playing dice with the Earth's interior

Jeannot Trampert
Utrecht University

Thanks to many excellent PhD students and postdocs

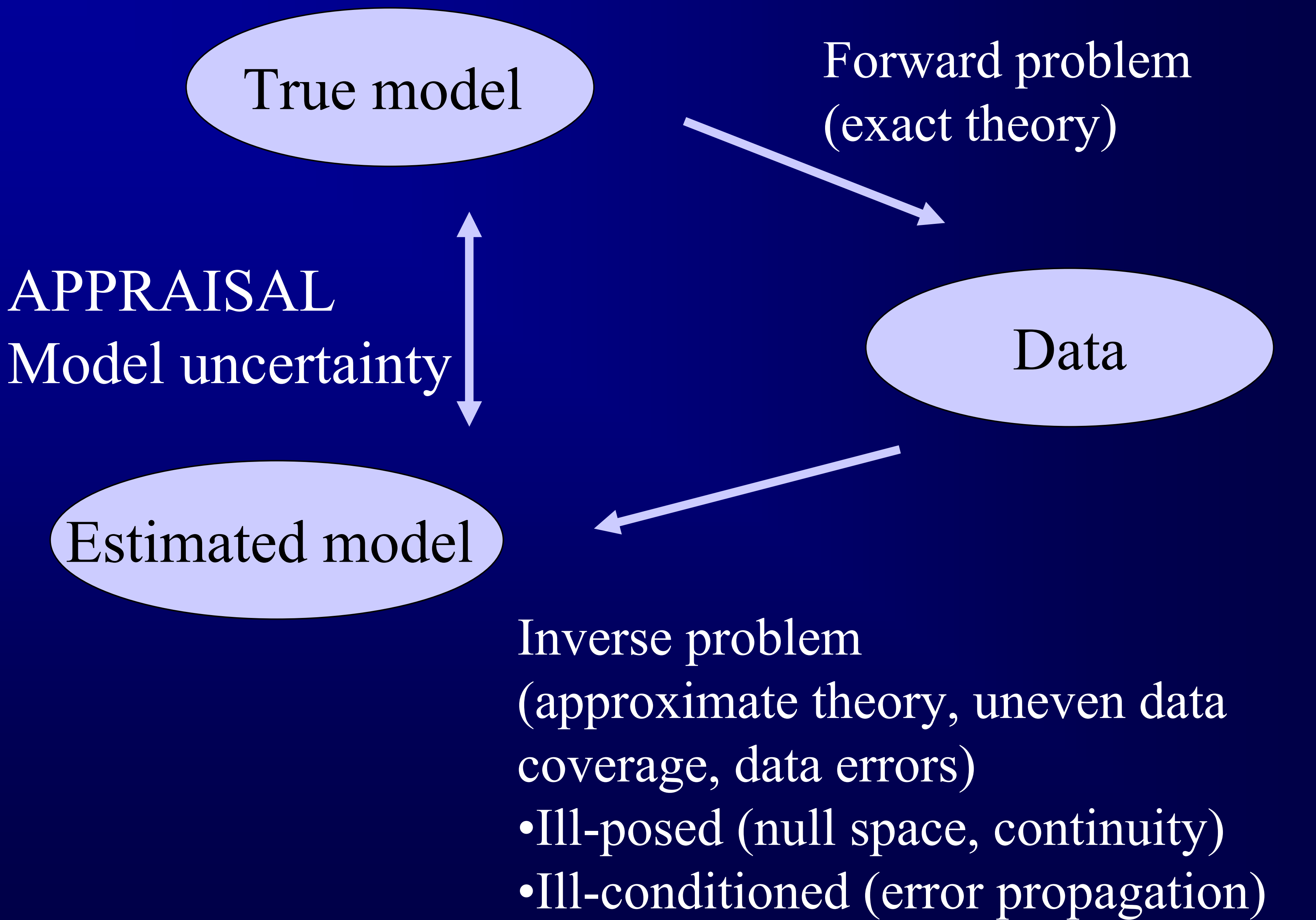


Is this the only model compatible with the data?

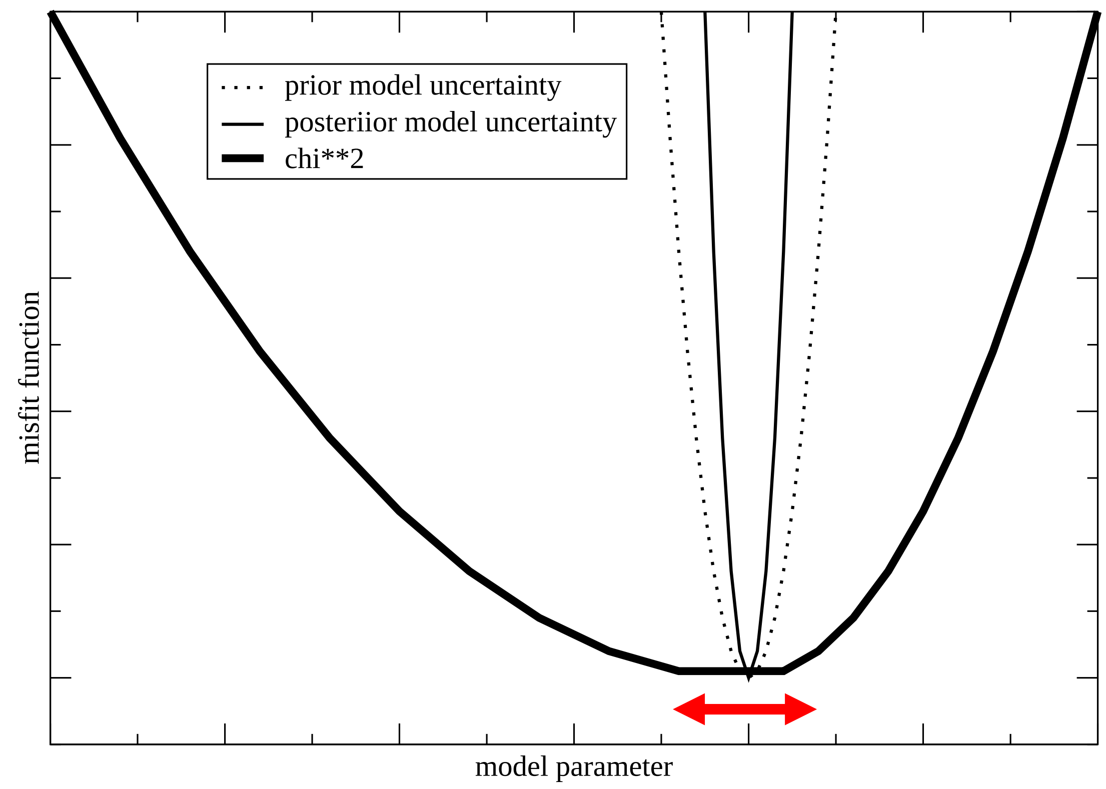
Project some chosen model onto the model null space of the forward operator

Then add to original model.
The same data fit is guaranteed!

Deal et al. JGR 1999

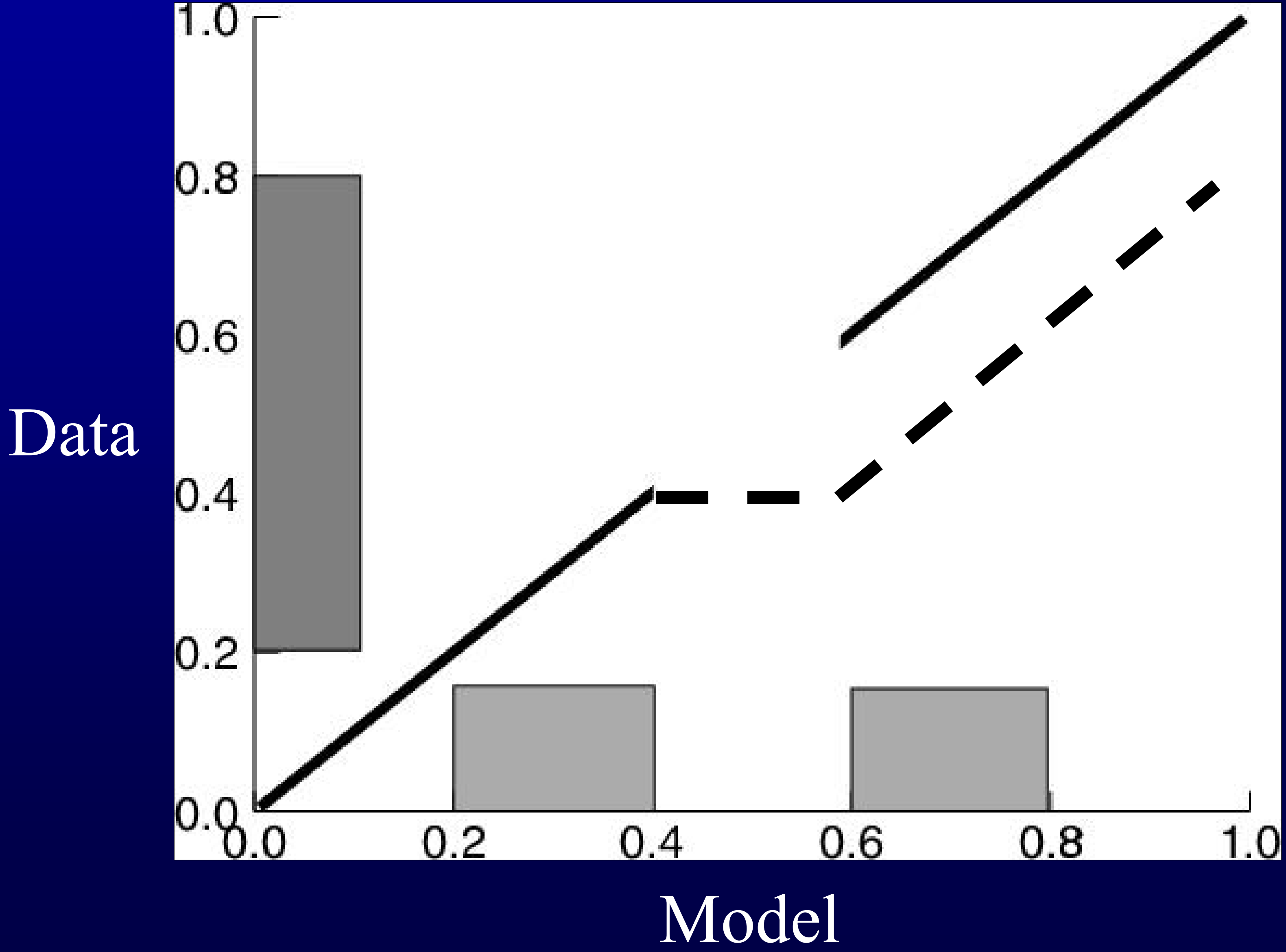

True model
Forward problem
(exact theory)
APPRAISAL
Model uncertainty
Data
Estimated model
Inverse problem
(approximate theory, uneven data coverage, data errors)
•Ill-posed (null space, continuity)
•Ill-conditioned (error propagation)

In the presence of a model null space, the cost function has a flat valley and the standard least-squares analysis underestimates uncertainty.



Realistic uncertainty for the given data

Multiple solutions are possible if the model does not continuously depend on the data, even for a linear problem.

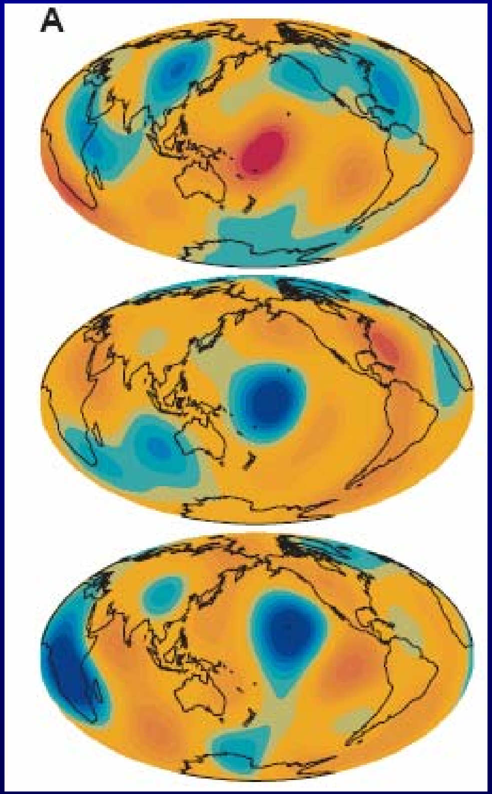


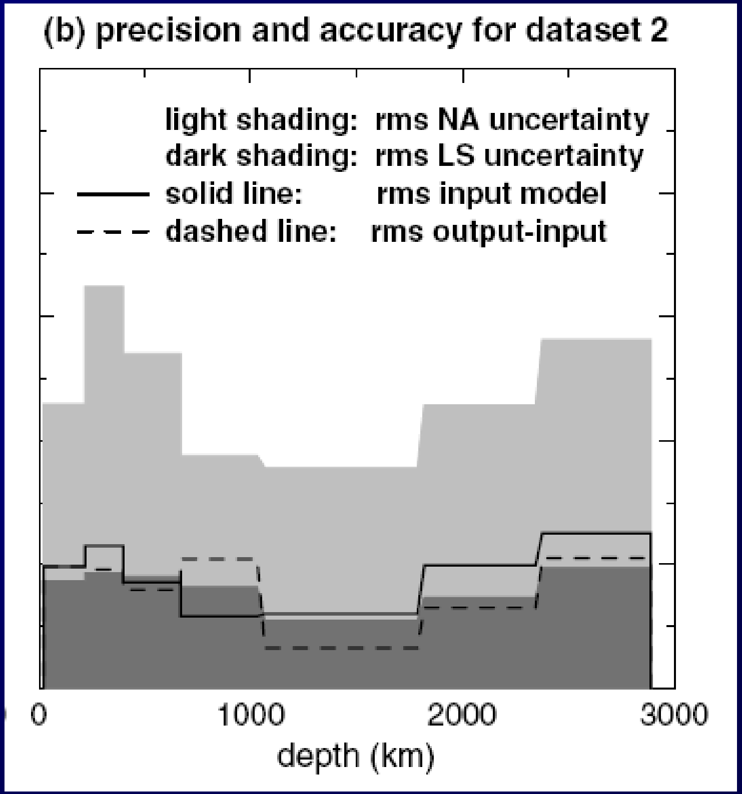
Does it matter? Yes!

dlnVs

dlnVphi

dlnrho





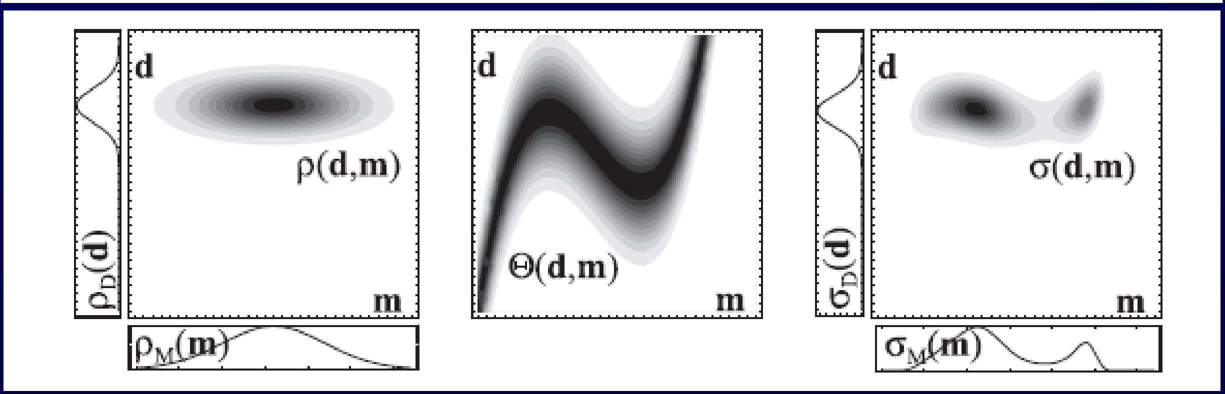
SPRD6 (Ishii and Tromp, 1999)

Input model is density from SPRD6 (Resovsky and Trampert, 2002)

The most general solution of an inverse problem (Bayes, Tarantola)

$$\sigma(d,m) = k\,\frac{\rho(d,m)\Theta(d,m)}{\mu(d,m)}$$

$$posterior = \frac{prior \text{ x } likelihood}{\text{evidence}}$$



A full model space search should estimate $\sigma(m) = k\rho_m(m)L(m)$

1. Exhaustive search
2. Brute force Monte Carlo (Shapiro and Ritzwoller, 2002)
3. Simulated Annealing (global optimisation with convergence proof)
4. Genetic algorithms (global optimisation with no covergence proof)
5. Sample $\rho(m)$ and apply Metropolis rule on $L(m)$. This will result in importance sampling of $\sigma(m)$ (Mosegaard and Tarantola, 1995)
6. Neighbourhood algorithm (Sambridge, 1999)
7. Neural networks (Meier et al., 2007)

The model space is HUGE!

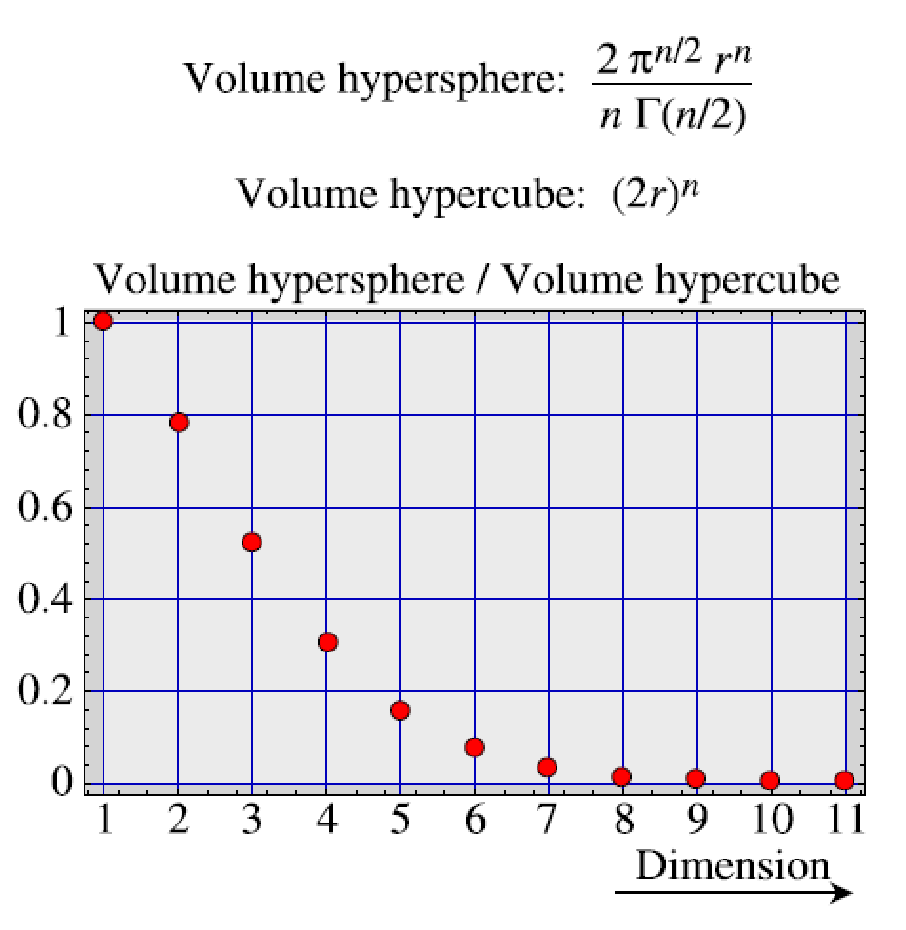
Draw a 1000 models per second where $m=\{0,1\}$

M=30 → 13 days

M=50 → 35702 years

Seismic tomography M = O(1000-100000)

The model space is EMPTY!



Tarantola, 2005

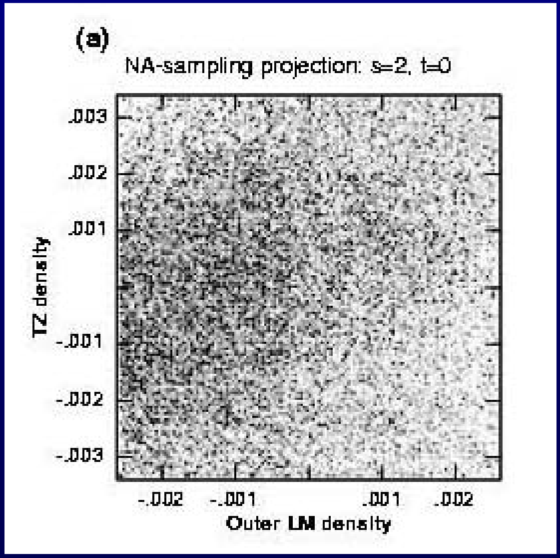
The curse of dimensionality → small problems M~30

$$\sigma(m) = k\rho_m(m)L(m)$$

1. Exhaustive search, IMPOSSIBLE
2. Brute force Monte Carlo, DIFFICULT
3. Simulated Annealing (global optimisation with convergence proof), DIFFICULT
4. Genetic algorithms (global optimisation with no covergence proof), ???
5. Sample $\rho(m)$ and apply Metropolis rule on L(m). This will result in importance sampling of $\sigma(m)$ (Mosegaard and Tarantola, 1995), BETTER, BUT HOW TO GET A MARGINAL
6. Neighbourhood algorithm (Sambridge, 1999), THIS WORKS
7. Neural networks, PROMISING

The neighbourhood algorithm:

Sambridge 1999



Stage 1:
Guided sampling of the model space.
Samples concentrate in areas (neighbourhoods) of better fit.

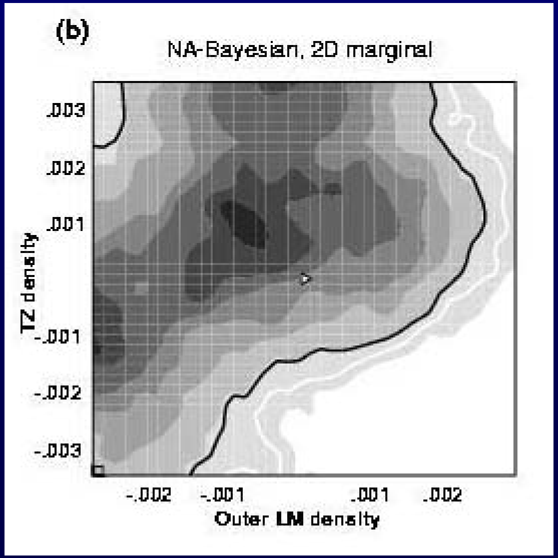
The neighbourhood algorithm (NA):

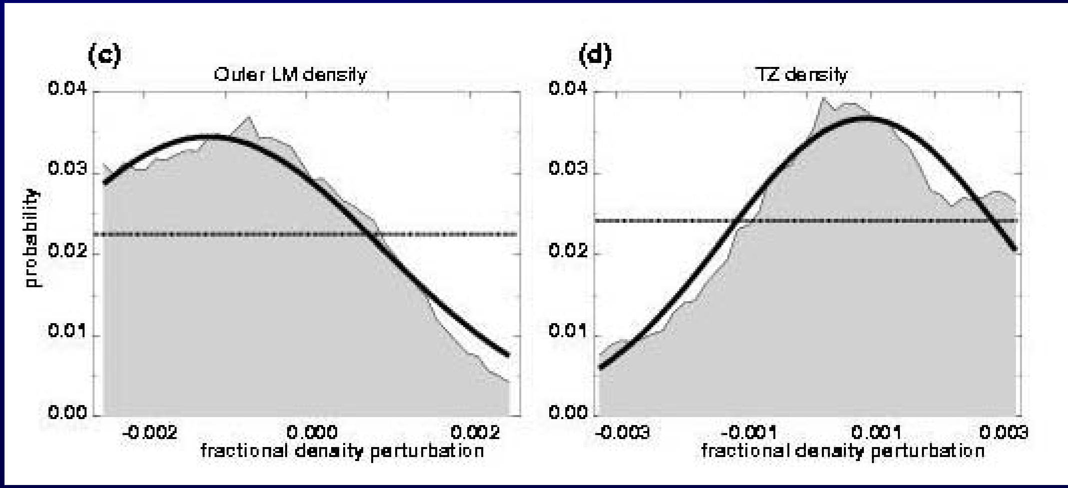
Stage 2: importance sampling

Resampling so that sampling density reflects posterior

2D marginal

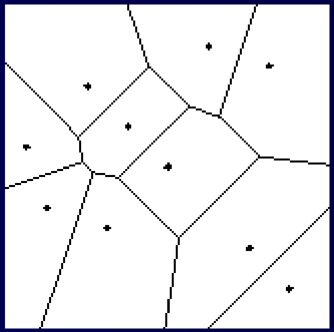
1D marginal



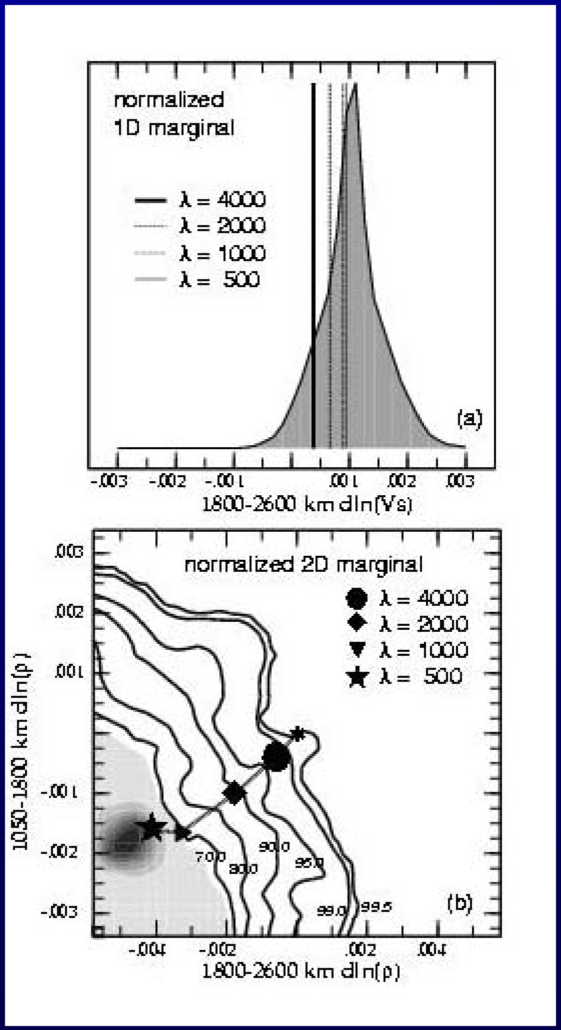


Advantages of NA

1. Interpolation in model space with Voronoi cells
2. Relative ranking in both stages (less dependent on data uncertainty)
3. Marginals calculated by Monte Carlo integration → convergence check



NA and Least Squares consistency



As damping is reduced, LS solution converges towards most likely NA solution.

In the presence of a null space, LS solution will diverge, but NA solution remains unaffected.

Finally some tomography!

We sampled all models compatible with the data using the Neighbourhood Algorithm (Sambridge, GJI 1999)

Data: 649 modes (NM, SW fundamentals and overtones)

He and Tromp, Laske and Masters, Masters and Laske, Resovsky and Ritzwoller, Resovsky and Pestena, Trampert and Woodhouse, Tromp and Zanzerkia, Woodhouse and Wong, van Heijst and Woodhouse, Widmer-Schnidrig

Parameterization: 5 layers

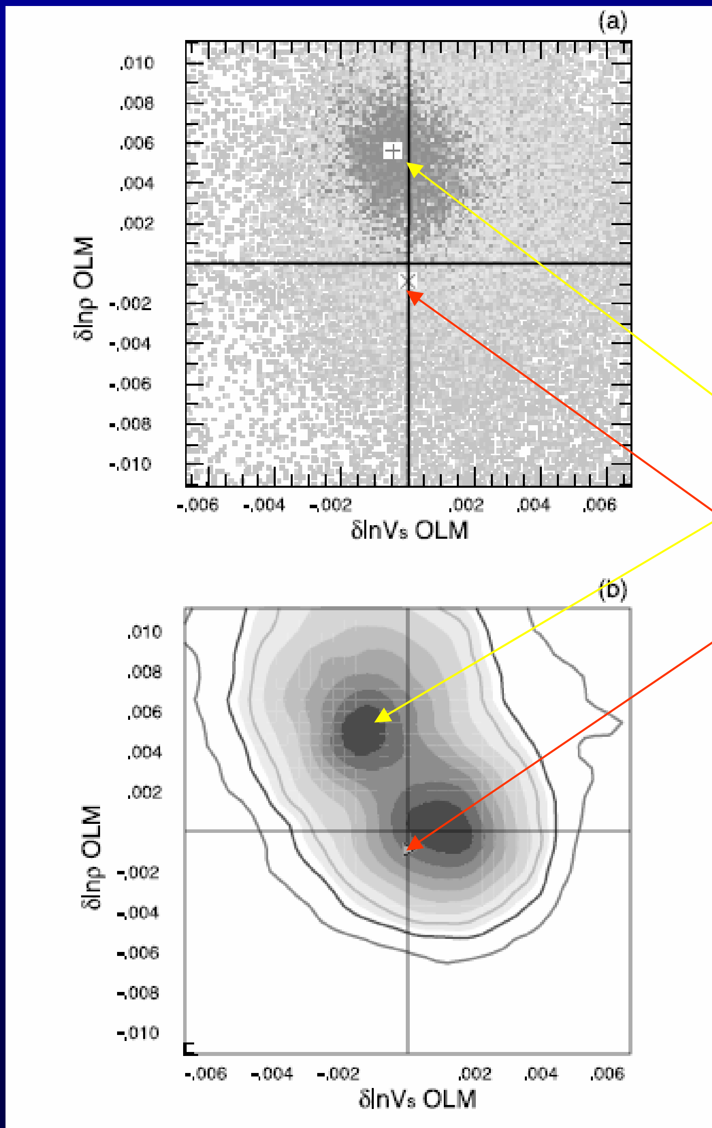
[2891-2000 km] [2000-1200 km] [1200-660 km]

[660-400 km]   [400-24 km]

Seismic parameters

dlnVs, dlnVΦ, dlnρ, relative topography on CMB and 660

Back to density

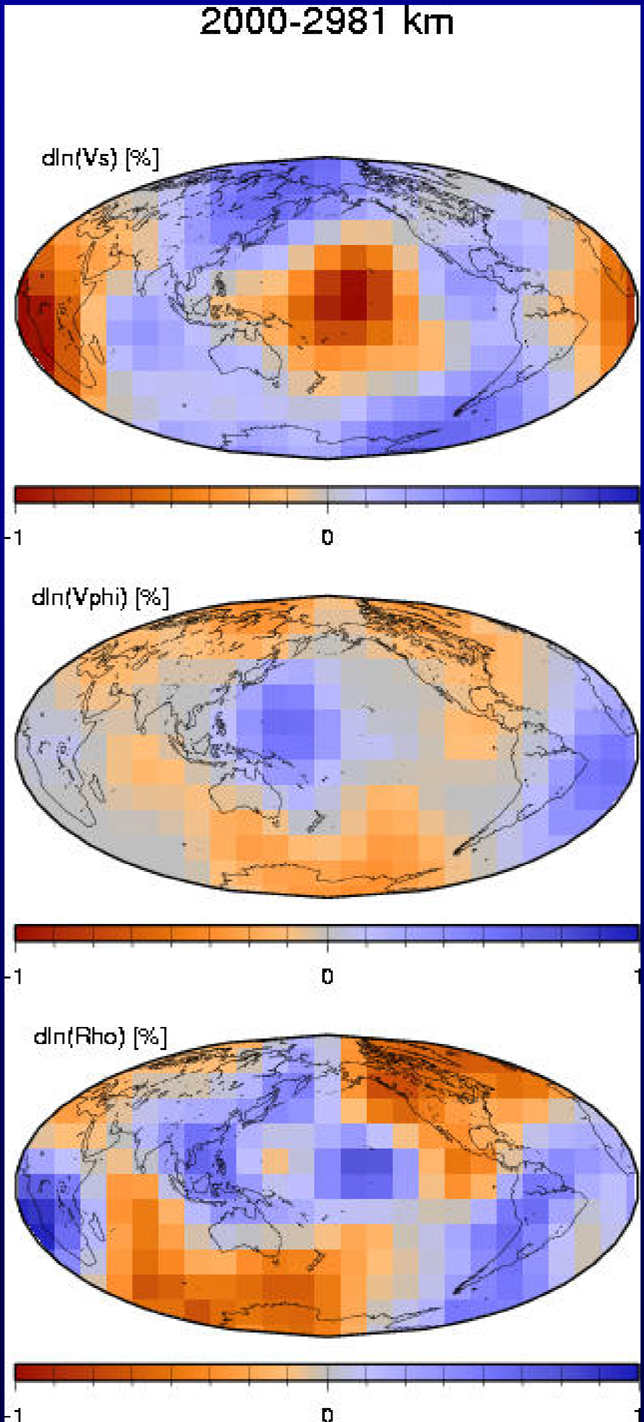


Stage 1

Most likely model

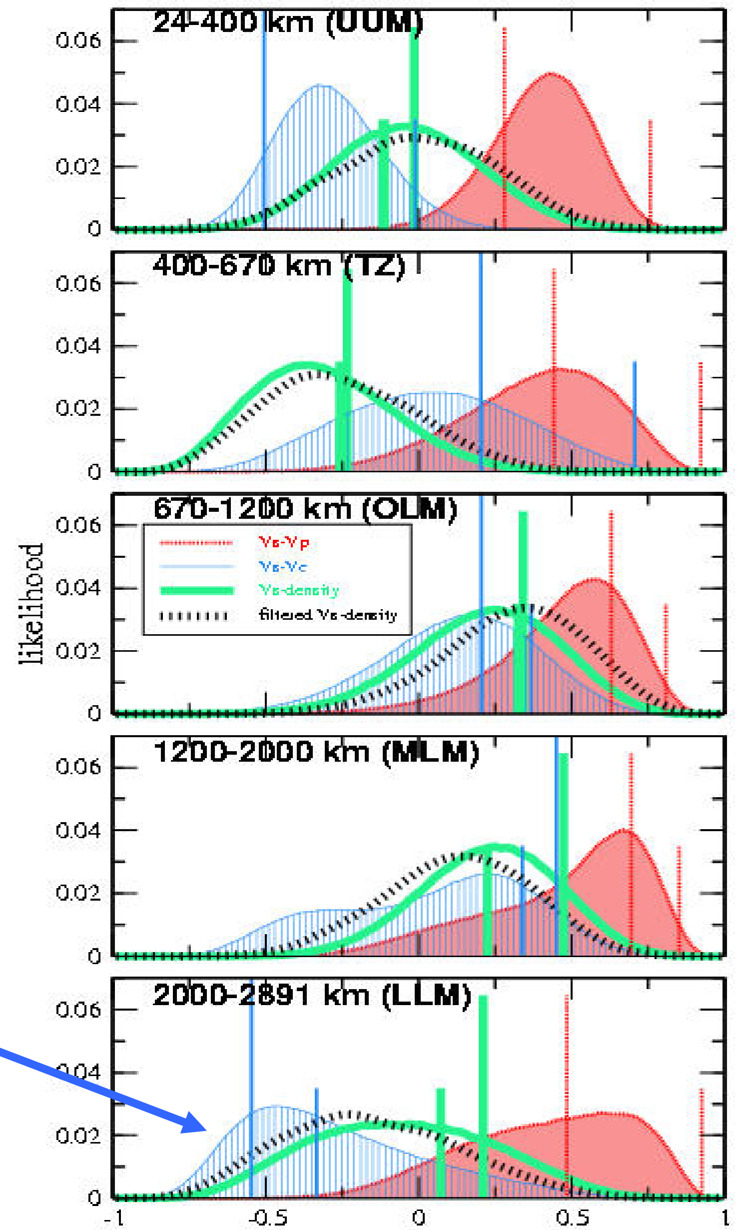
SPRD6

Stage 2



- Gravity filtered models over 15x15 degree equal area caps
- Likelihoods in each cap are nearly Gaussian
- Most likely model (above one standard deviation)
- Uniform uncertainties
  - dlnVs=0.12 %
  - dlnVΦ=0.26 %
  - dlnρ=0.48 %

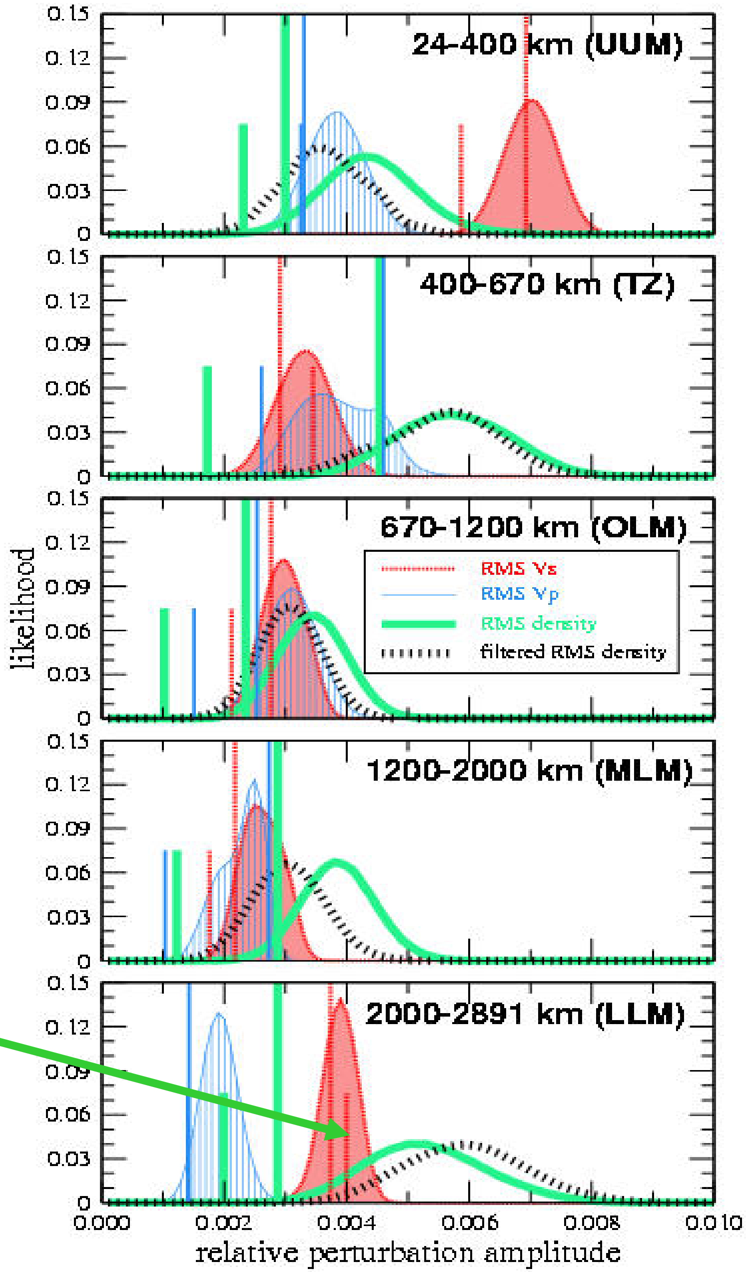
Likelihood of correlation between seismic parameters



Evidence for chemical heterogeneity Vs-VΦ

Half-height vertical bar corresponds to SPRD6

Likelihood of rms-amplitude of seismic parameters



Evidence for chemical heterogeneity RMS density

The neural network (NN) approach:

Bishop 1995, MacKay 2003

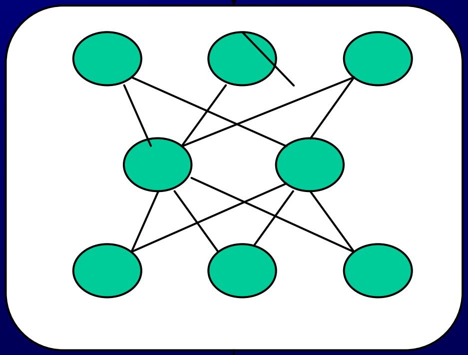
- A neural network can be seen as a non-linear filter between some input and output
- The NN is an approximation to any function g in the non-linear relation $d=g(m)$
- A training set is used to calculate the coefficients of the NN by non-linear optimisation

3) Forward propagating a new datum through the trained network (i.e. solving the inverse problem)

Network Input:
Observed or synthetic data for training

$\mathbf{d}^{obs}$

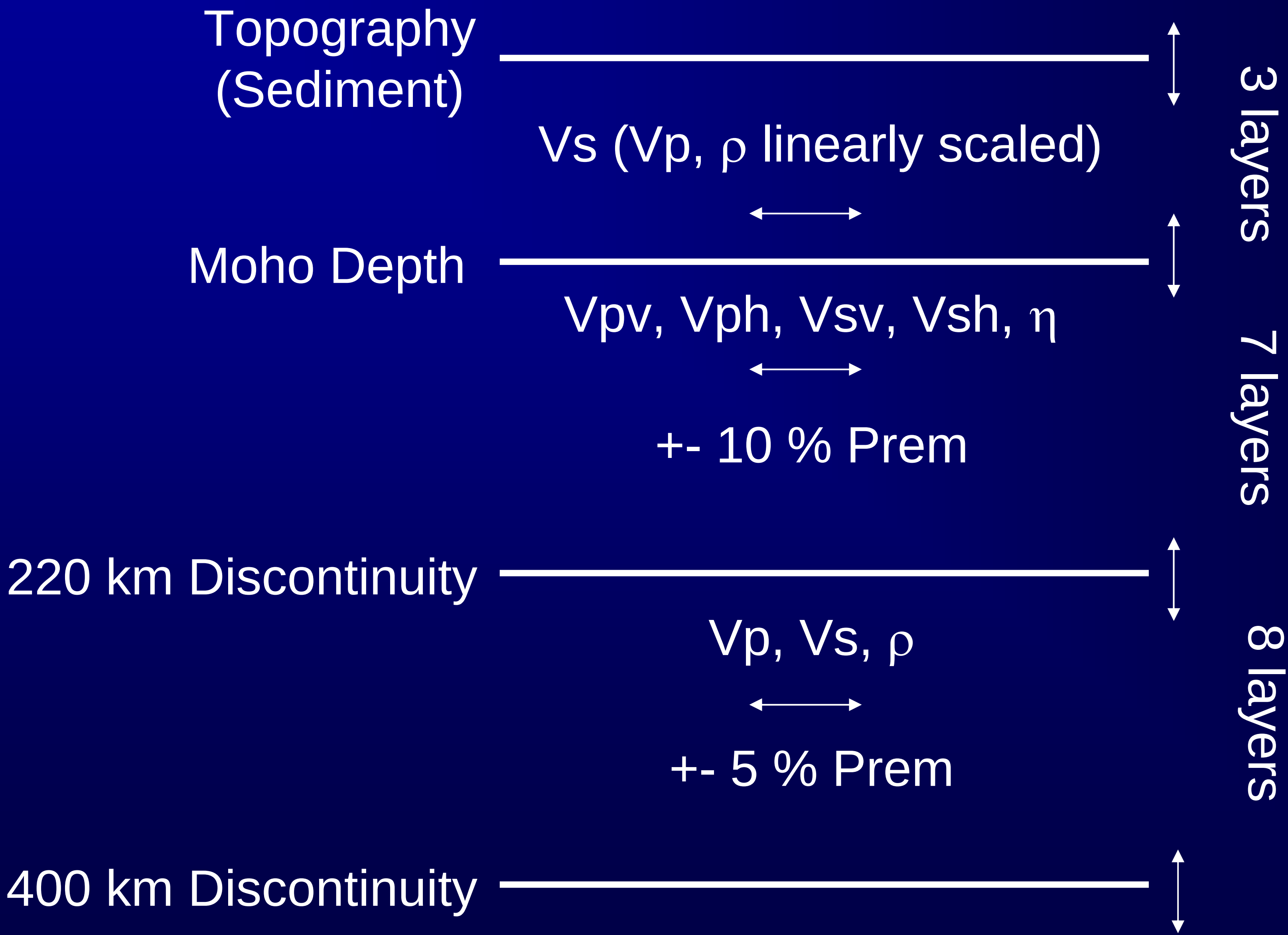
Neural Network



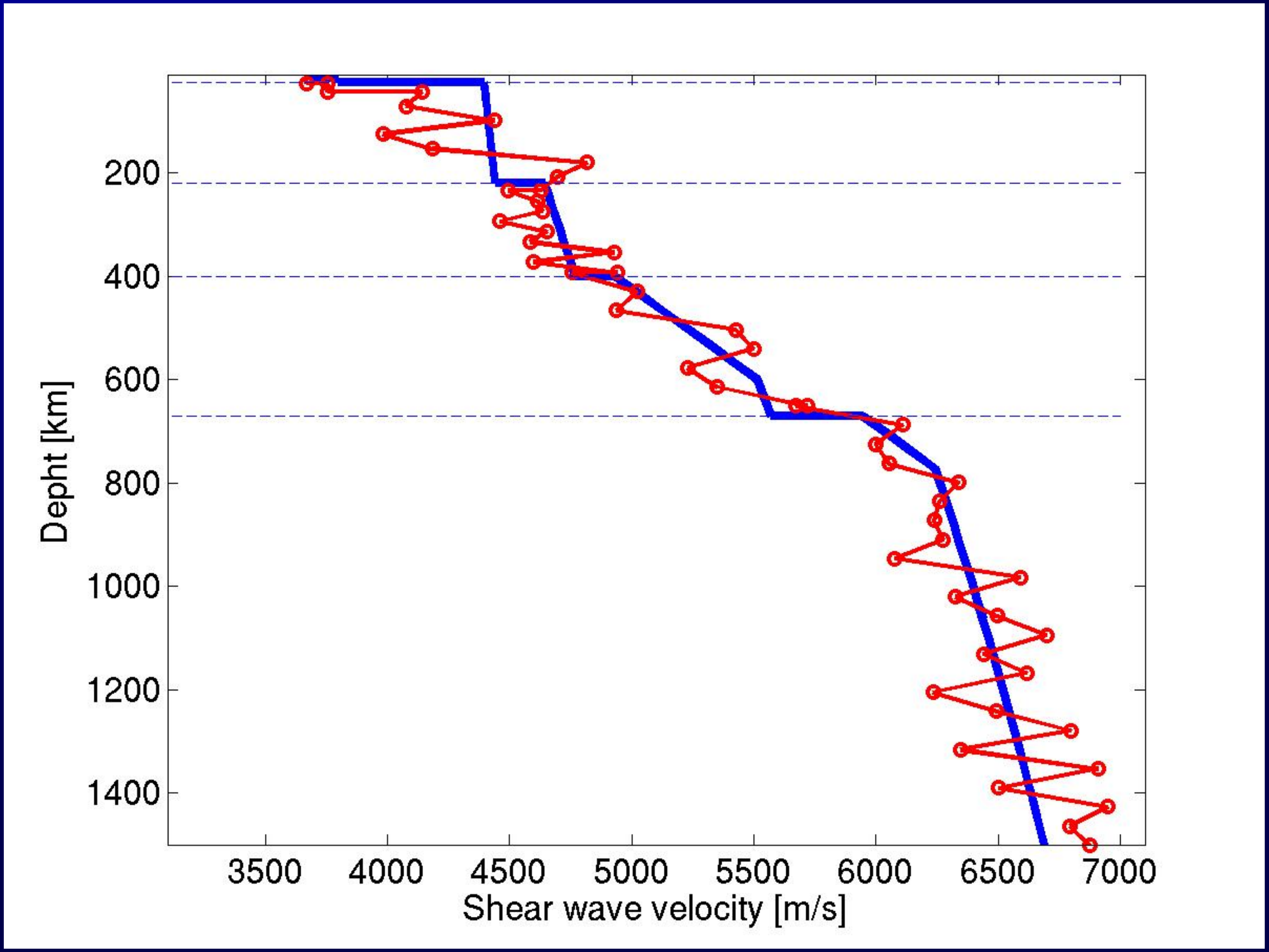
Network Output:
Conditional probability density

$\mathbf{p}(m|\mathbf{d}^{obs},\mathbf{w}^*)$

Model Parameterization

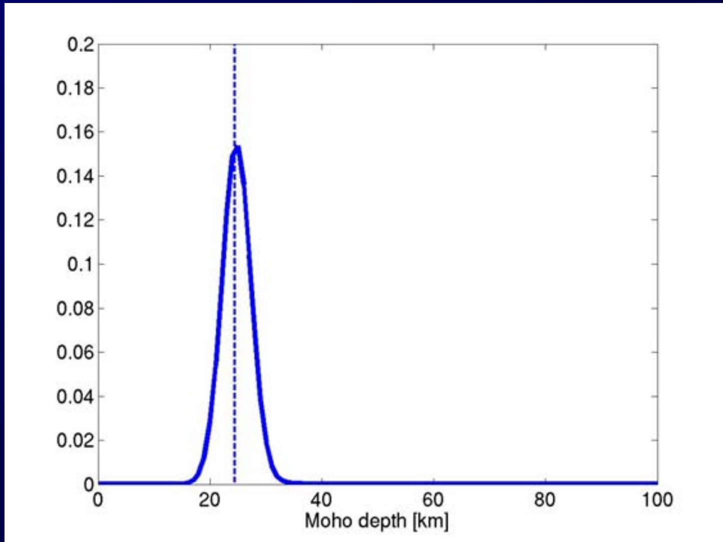


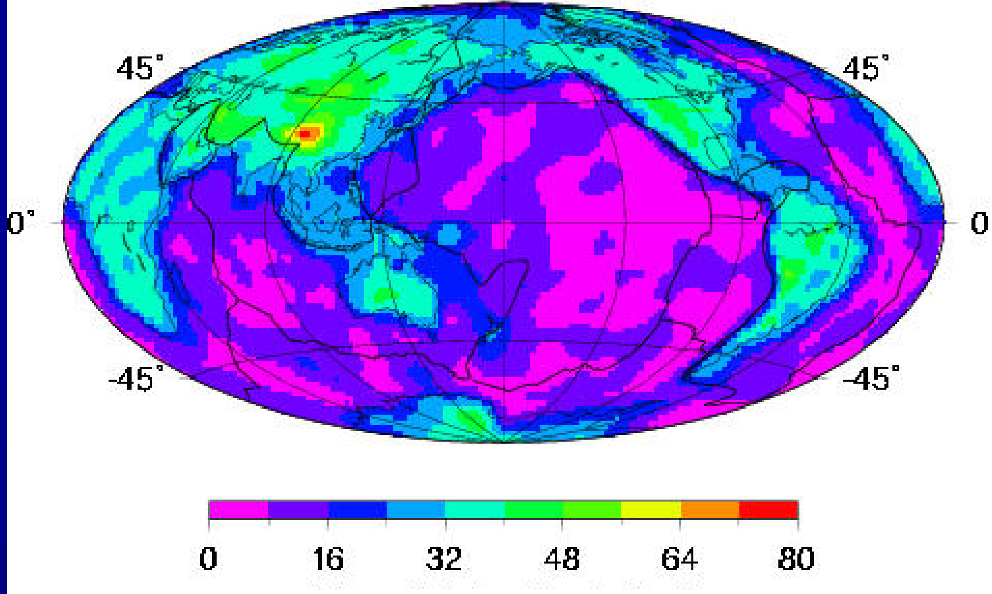
Example of a Vs realisation



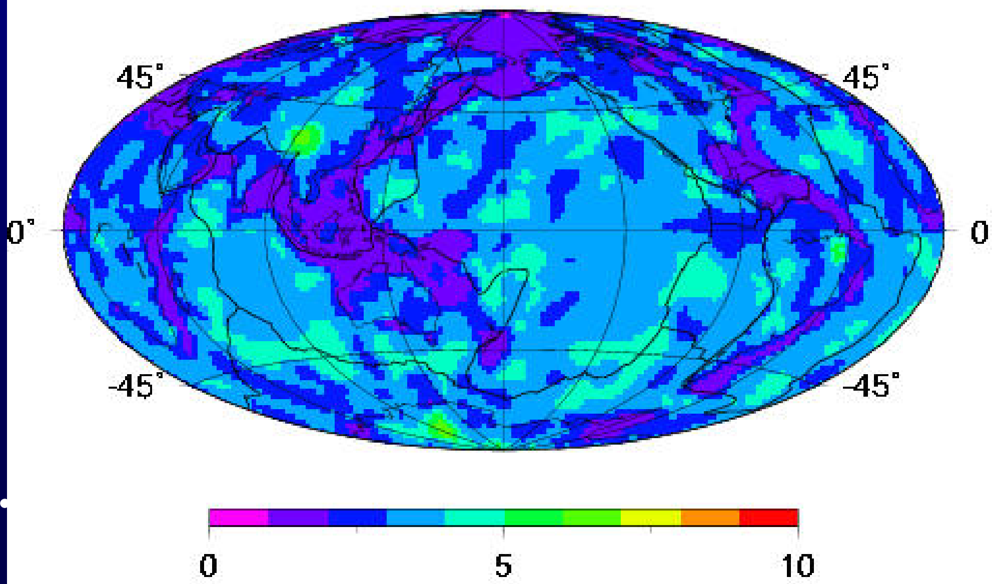
Advantages of NN

1. The curse of dimensionality is not a problem because NN approximates a function not a data prediction!
2. Flexible: invert for any combination of parameters
3. (Single marginal output)



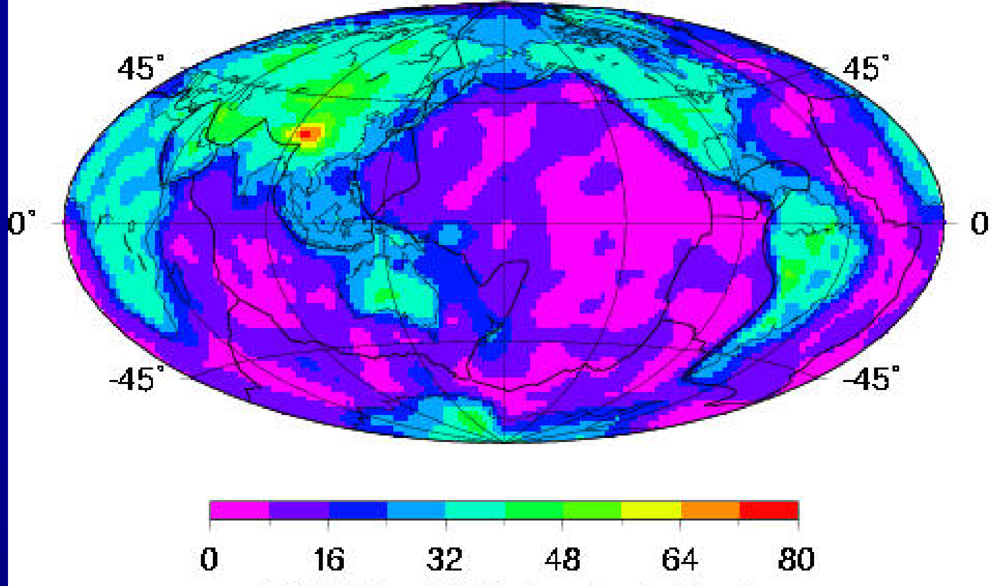


Mean Moho depth [km]

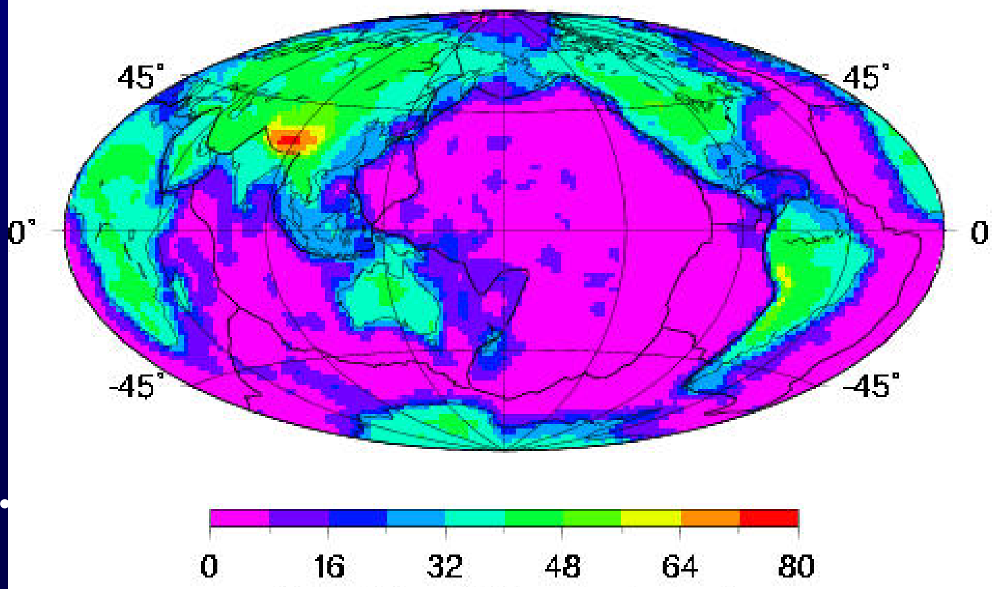


Standard deviation σ [km]

Meier et al.
2007



Mean Moho depth [km]

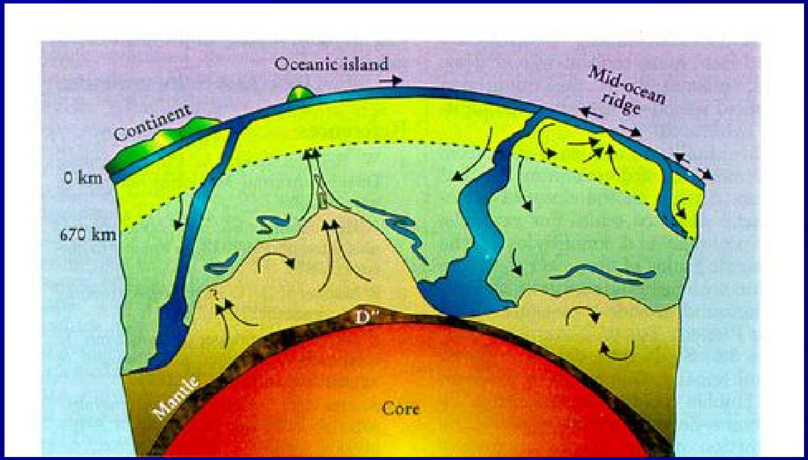


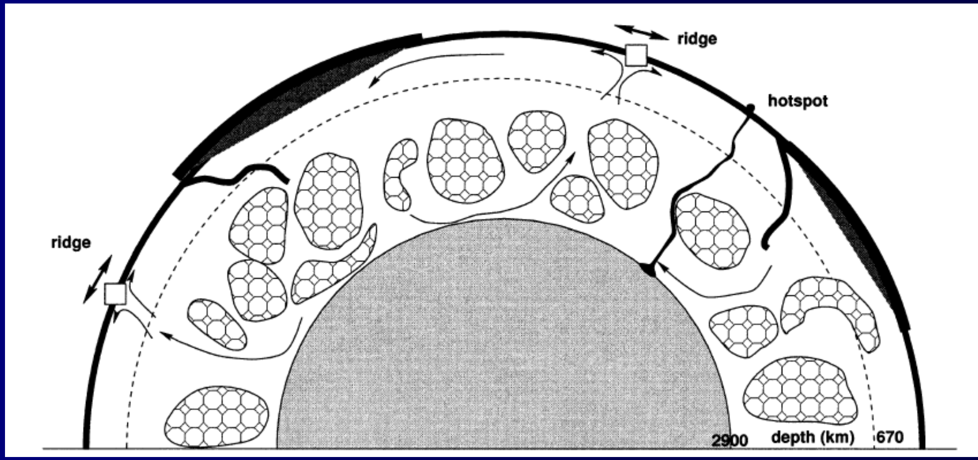
CRUST2.0 Moho depth [km]

Meier et al. 2007

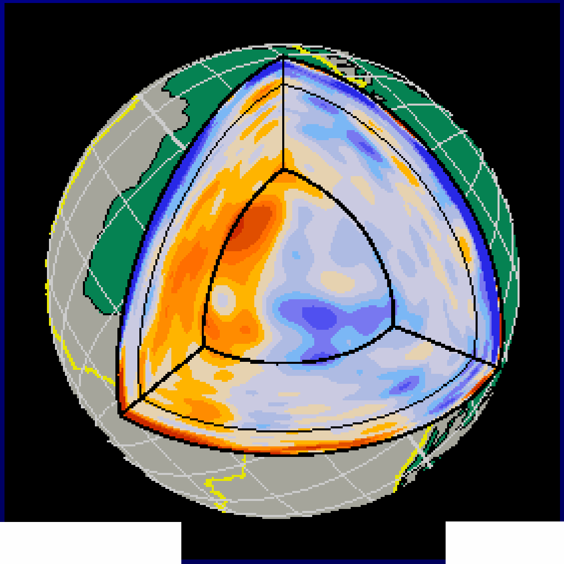
NA versus NN

- NA: joint pdf → marginals by integration → small problem
- NN: 1D marginals only (one network per parameter) → large problem

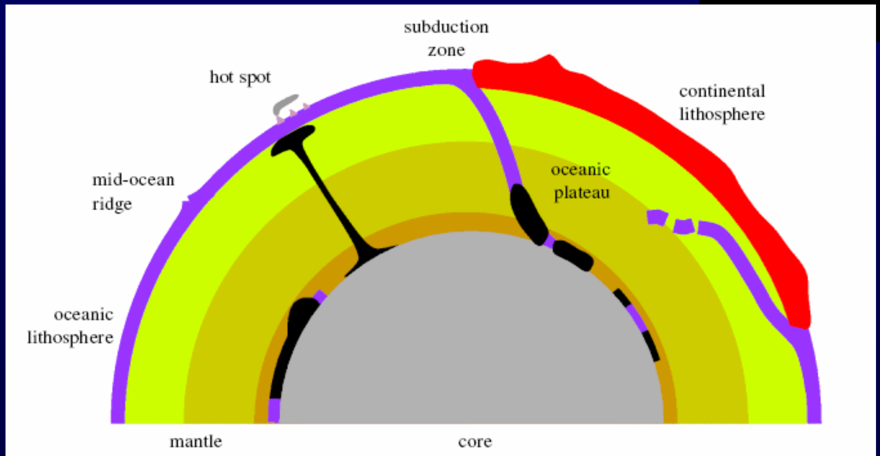


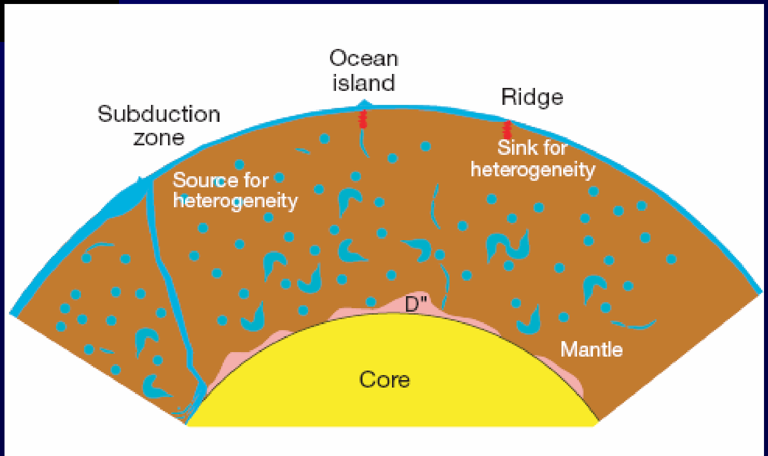


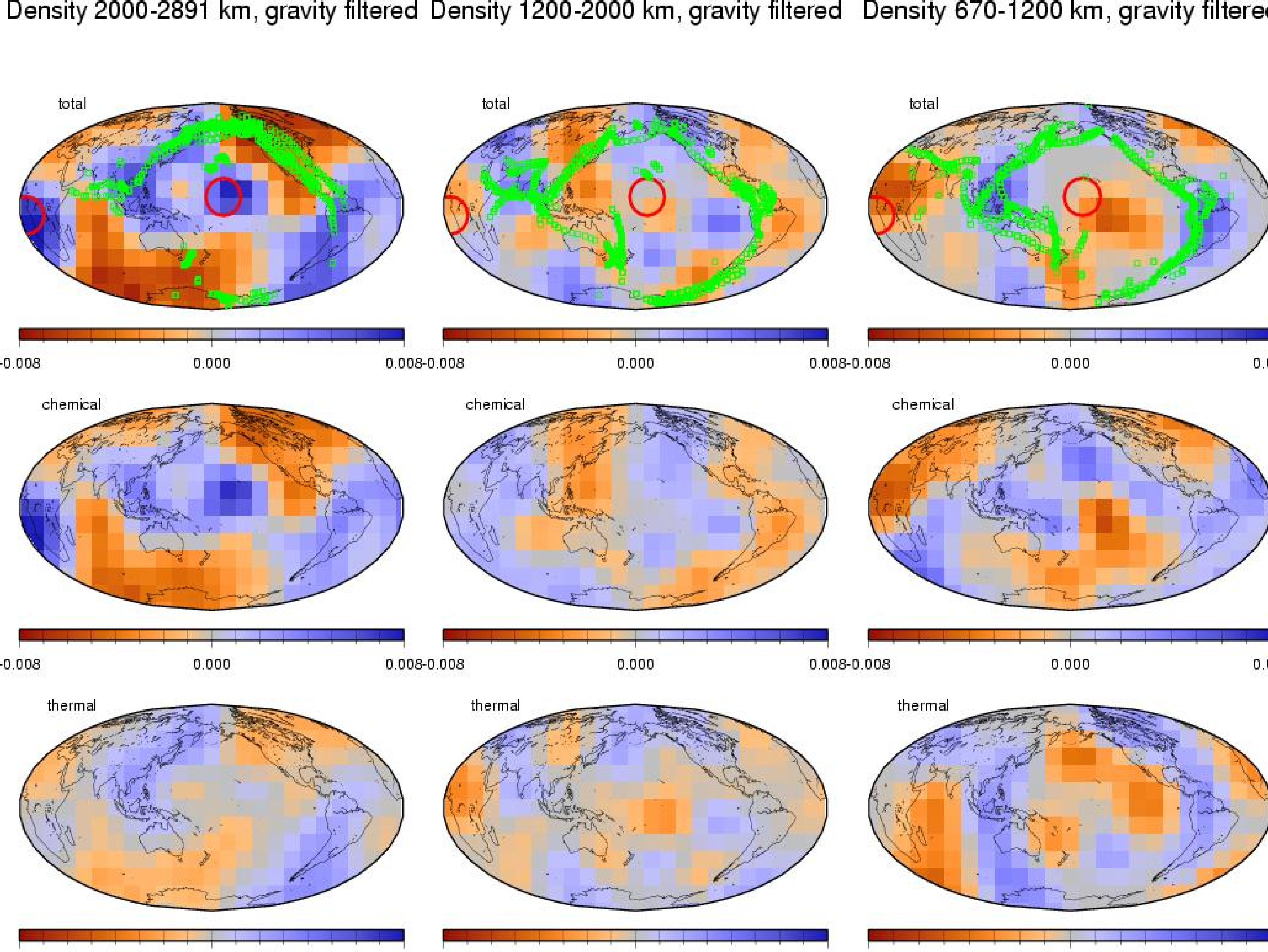
Advantage of having a pdf

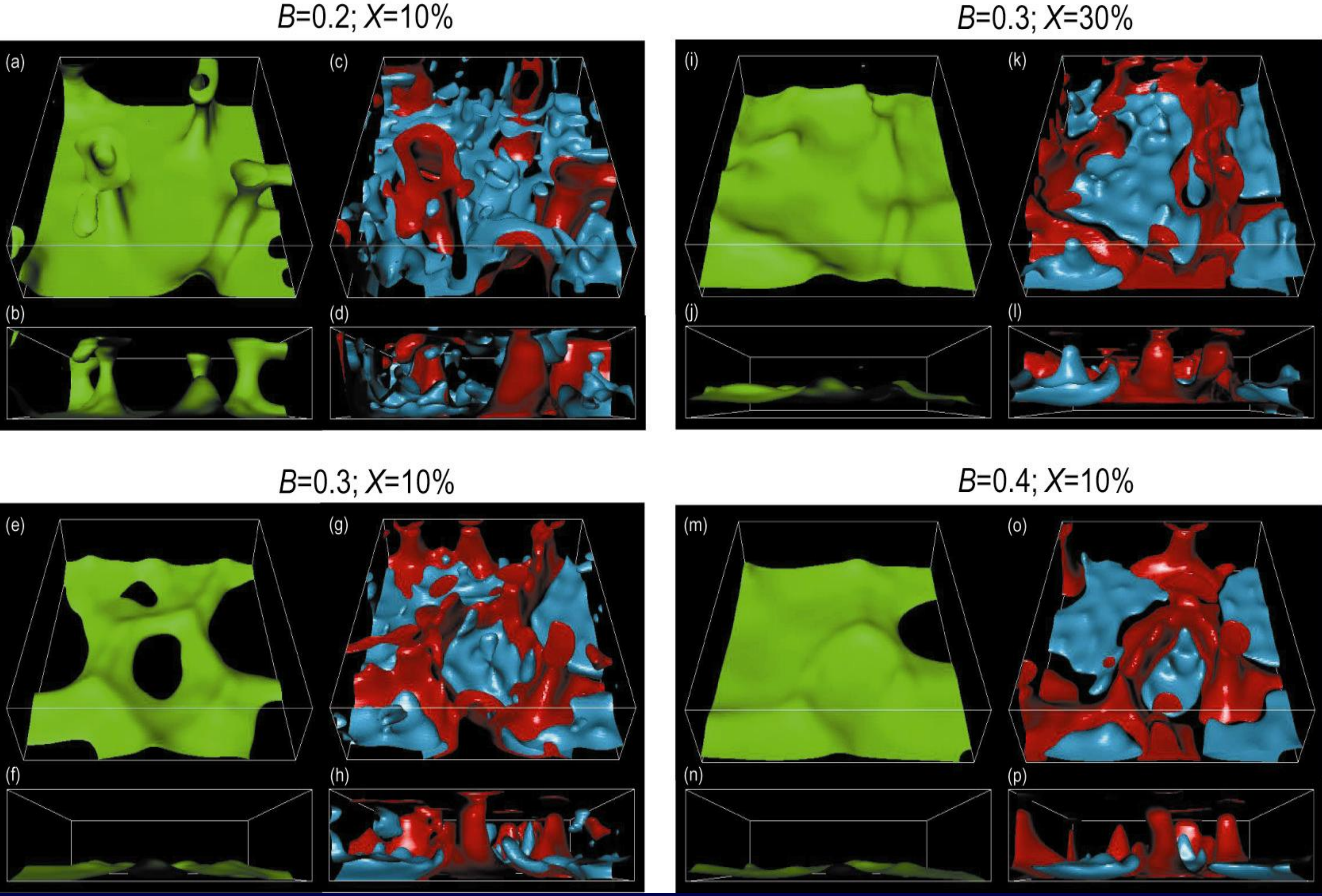


Quantitative comparison

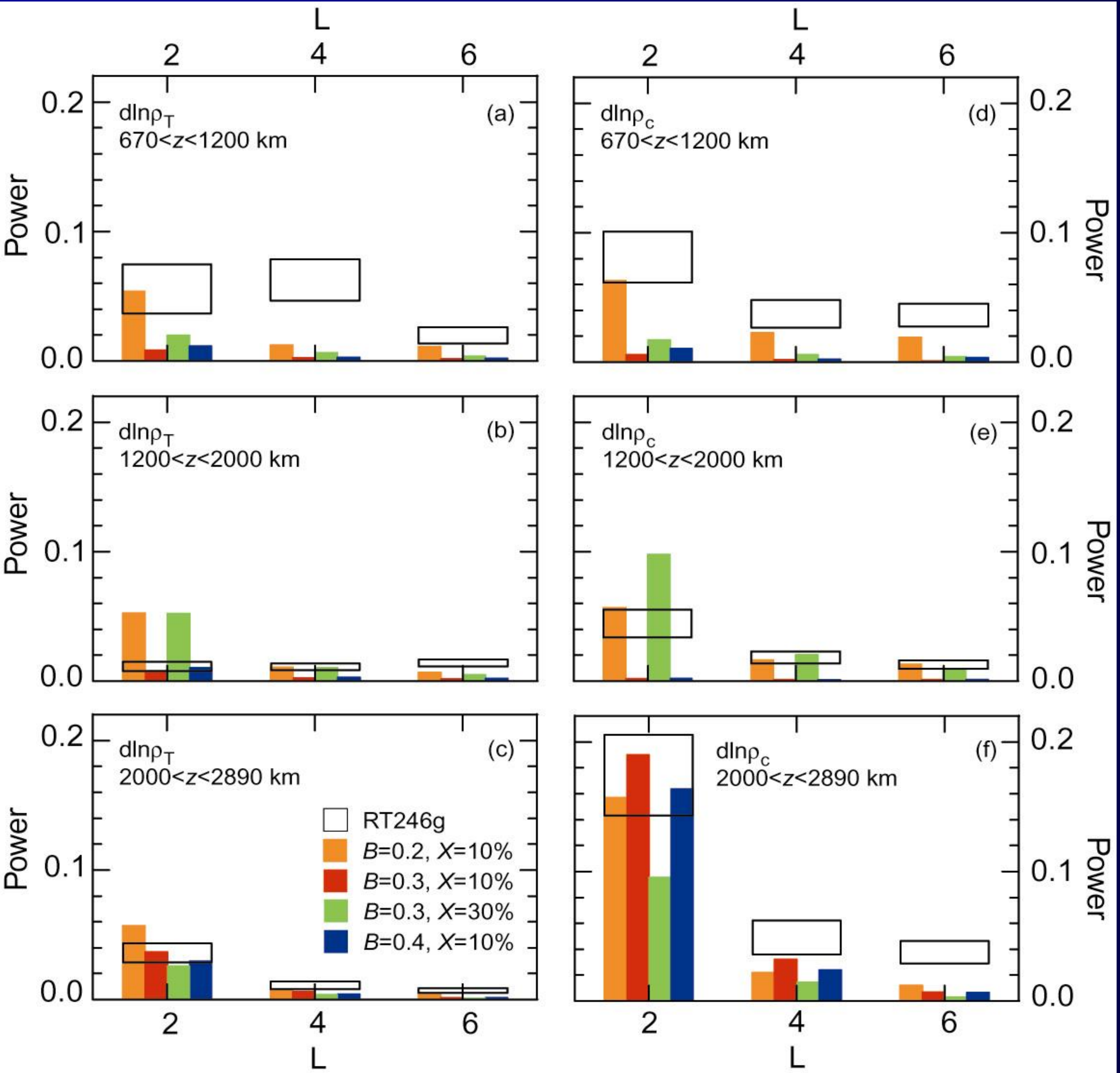





Density 2000-2891 km, gravity filtered
Density 1200-2000 km, gravity filtered
Density 670-1200 km, gravity filtere
total
total
total
-0.008
0.000
0.008-0.008
0.000
0.008-0.008
0.000
chemical
chemical
chemical
-0.008
0.000
0.008-0.008
0.000
0.008-0.008
0.000
thermal
thermal
thermal



Tackley, GGG, 2002


L
2
4
6
dlnρT
670<z<1200 km
(a)
dlnρc
670<z<1200 km
(d)
dlnρT
1200<z<2000 km
(b)
dlnρc
1200<z<2000 km
(e)
dlnρT
2000<z<2890 km
(c)
dlnρc
2000<z<2890 km
(f)
Power
0.0
0.1
0.2
RT246g
B=0.2, X=10%
B=0.3, X=10%
B=0.3, X=30%
B=0.4, X=10%

Finally let's add some SPICE

- For years to come, full Monte Carlo together with 3D wave propagation is out of reach
- NA: iterative linearized inversion using adjoint kernels (Sieminski, Liu, Tromp) → NA around model of last iteration
- NN: data base of all 3D simulations done worldwide can be used for training.