Trans-D Methods for Quantifying Uncertainty in Seismic Inversion

Anandaroop Ray

Acknowledgments:
Sam Kaplan, John Washbourne, Uwe Albertin, Anusha Sekar, Mike Hoversten

Problems with the FWI objective function

$$\arg\min \phi(\mathbf{m}) = ||\mathbf{W}(\mathbf{d} - \mathbf{f}(\mathbf{m}))||_2^2 + \lambda^2 ||\mathbf{Rm}||_p^p$$

Limitations in conventional approach
Bad choices for **R** and λ^2 lead to slow convergence
Solution requires linearization
Local minima abound
No convergence guarantees exist
Model **m** is high dimensional

Problems with the FWI objective function

$$\arg\min \phi(\mathbf{m}) = ||\mathbf{W}(\mathbf{d} - \mathbf{f}(\mathbf{m}))||_2^2 + \lambda^2 ||\mathbf{Rm}||_p^p$$

Limitations in conventional approach	Bayesian solutions
Bad choices for **R** and λ^2 lead to slow convergence	Appeal to a parsimonious model basis
Solution requires linearization	Do not linearize
Local minima abound	Sample in parallel (parallel tempering)
No convergence guarantees exist	Sample the model space (Markov chain Monte Carlo or McMC)
Model **m** is high dimensional	Appeal to parsimony (trans-D McMC)

A traditional Bayesian view

$$\text{updated belief} \propto \text{likelihood of belief} \cdot \text{prior belief}$$

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}) \cdot p(\mathbf{m})$$

Given the observed seismic data **d,** new belief in model **m**

Given the model **m,** accuracy of seismic prediction

m is a model obtained from prior notions, e.g., well data, geology, etc.

Equivalence of Bayes' theorem with optimization

$$\text{updated belief} \propto \text{likelihood of belief} \cdot \text{prior belief}$$

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}) \cdot p(\mathbf{m})$$

$$\arg\min \phi(\mathbf{m}) = ||\mathbf{W}(\mathbf{d} - \mathbf{f}(\mathbf{m}))||_2^2 + \lambda^2 ||\mathbf{R}\mathbf{m}||_p^p$$

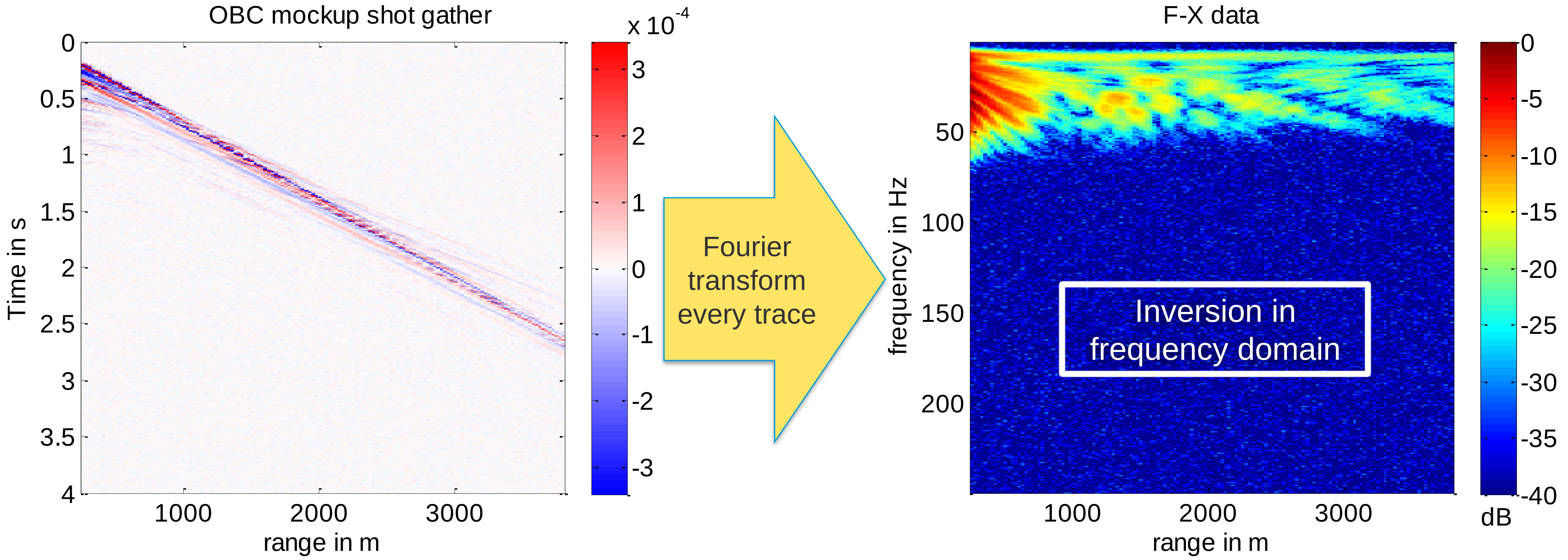
But non-linearity (i.e., non-uniqueness), high model dimension and model parametrization make equivalence more of a ***theoretical*** comfort

Trans-dimensional (trans-D) Bayesian inversion

Ordinary McMC	Change model parameters while sampling
trans-D McMC	***Add/delete*** parameters while sampling

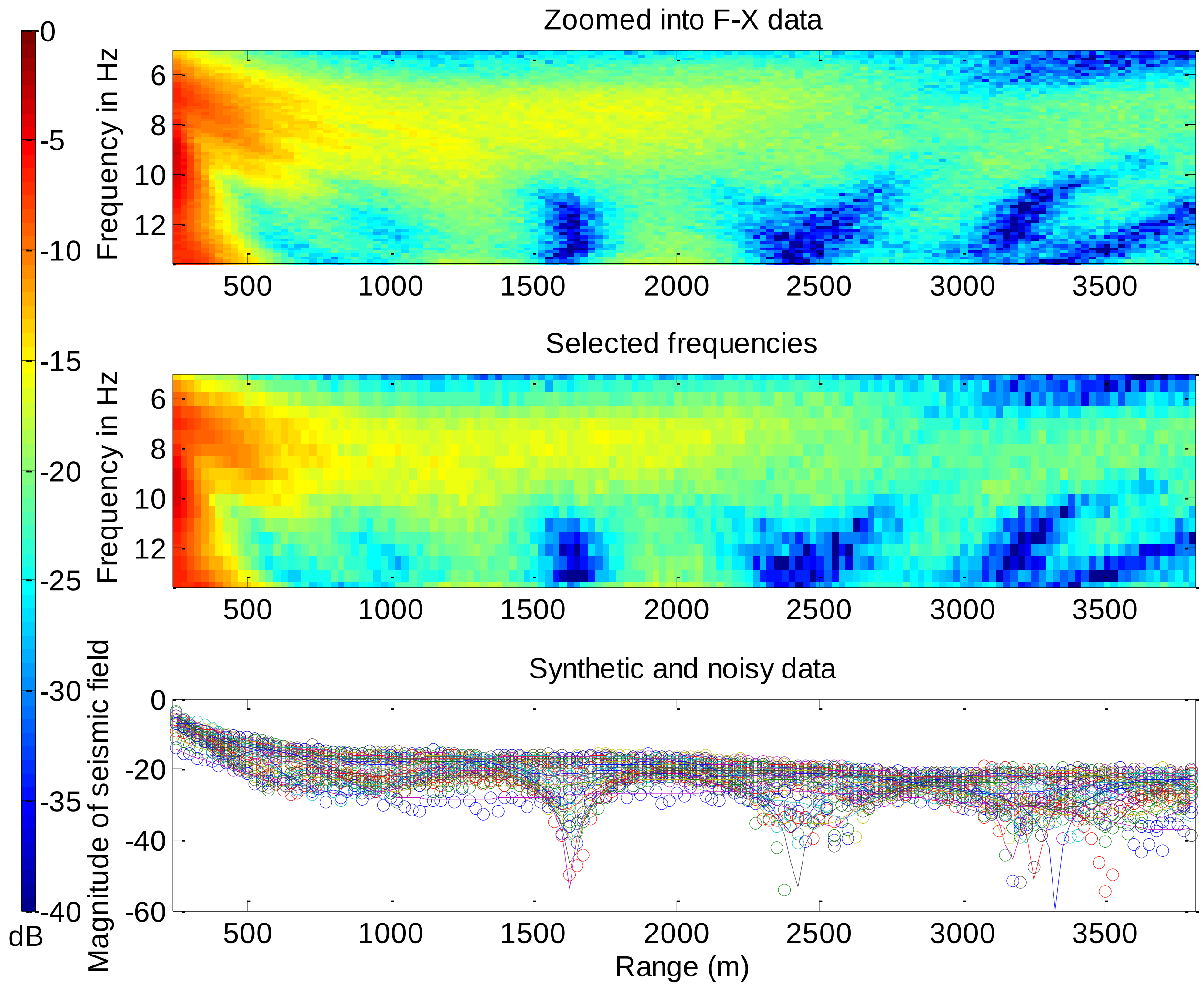
$$p(k|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m_k}, k) \cdot \left[f_1 . f_2 ... f_k \right].$$

Noisy synthetic data



Hydrophones receive mode converted data as well

F-X data for inversion
(SYNTHETIC EXAMPLE)

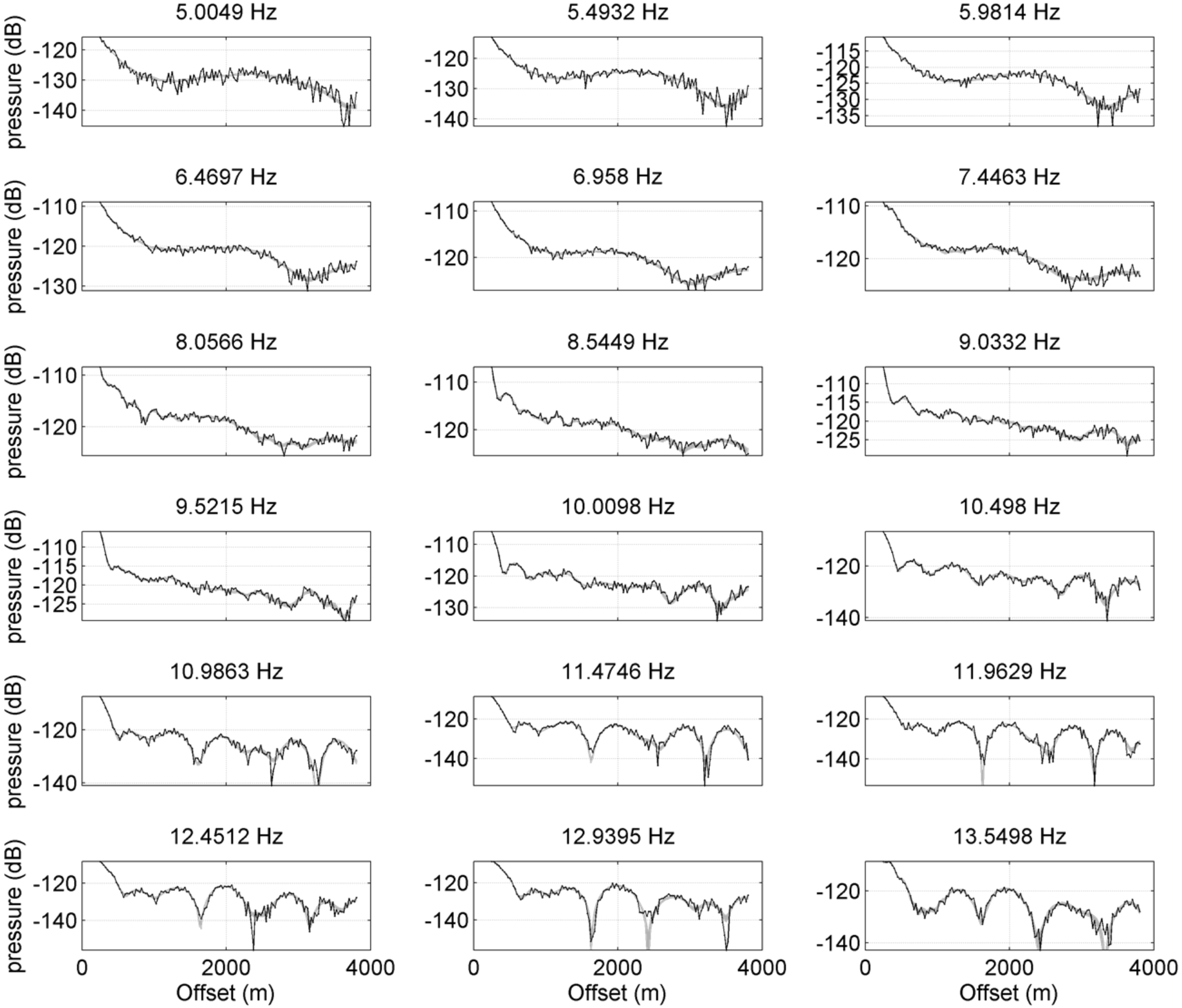


18 frequencies selected every 0.5 Hz

18 frequency inversion

Black = noisy data

Grey = 500 calculated responses, randomly selected from posterior models



Maximum likelihood source spectrum, data noise and misfit function

0.Compute Green's function

For model $\mathbf{m} = [\mathbf{z}, \mathbf{V_p}, \boldsymbol{\nu}, \boldsymbol{\rho}, \mathbf{Q_p}, \mathbf{Q_s}],$

1.Estimate wavelet

$$\hat{S}_l(\mathbf{m}) = \frac{\mathbf{G}_l^\dagger(\mathbf{m})\mathbf{d}_l}{\mathbf{G}_l^\dagger(\mathbf{m})\mathbf{G}_l(\mathbf{m})},$$

2.Calculate forward

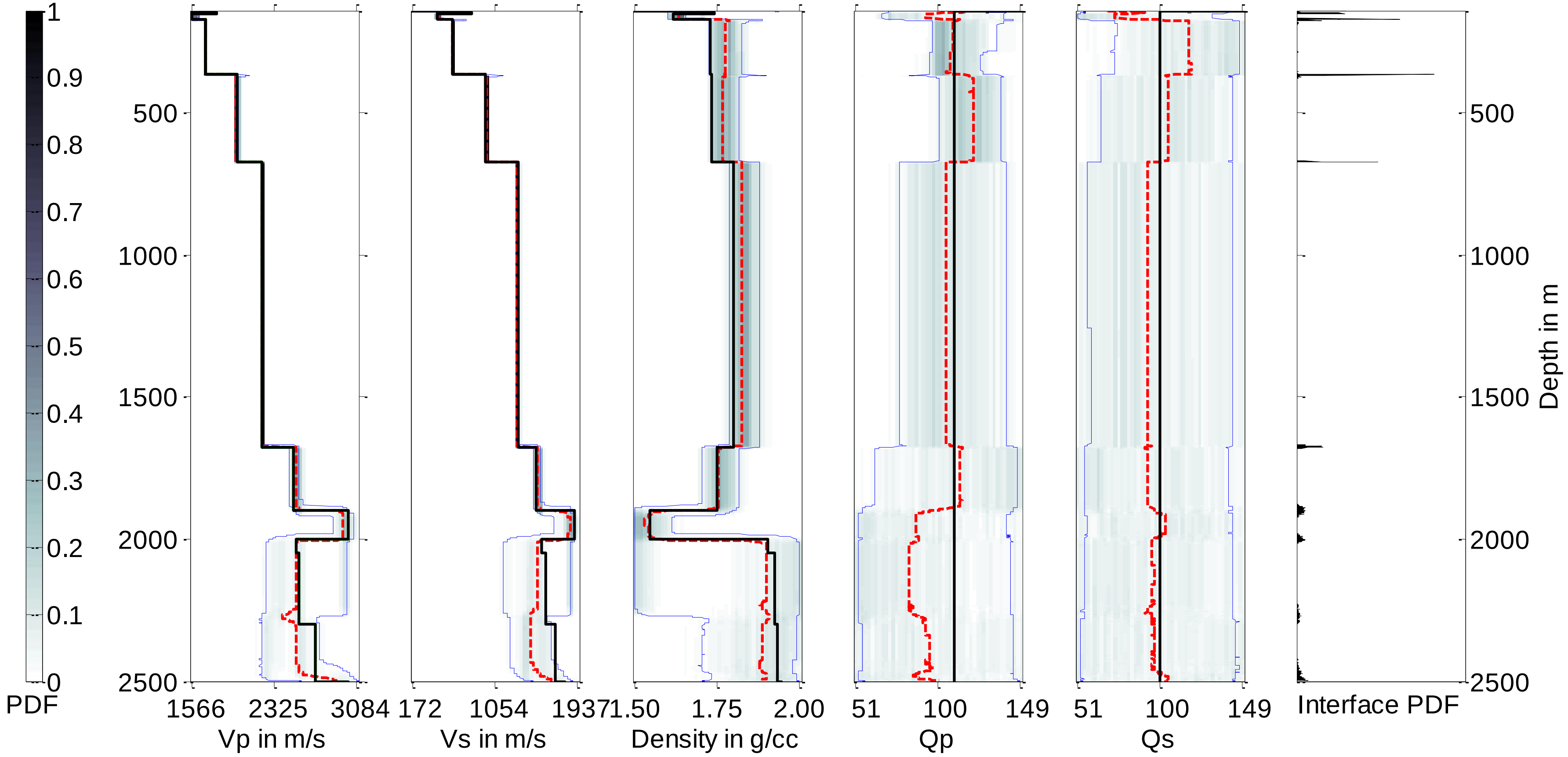
$$\mathbf{f_l}(\mathbf{m}) = \hat{S}_l(\mathbf{m}) \cdot \mathbf{G}_l(\mathbf{m}).$$

3.Calculate misfit

$$-\log \mathcal{L}(\mathbf{m}) = n_r \sum_{l=1}^{n_f} [\mathbf{d_l} - \mathbf{f_l}(\mathbf{m})]^\dagger [\mathbf{d_l} - \mathbf{f_l}(\mathbf{m})].$$

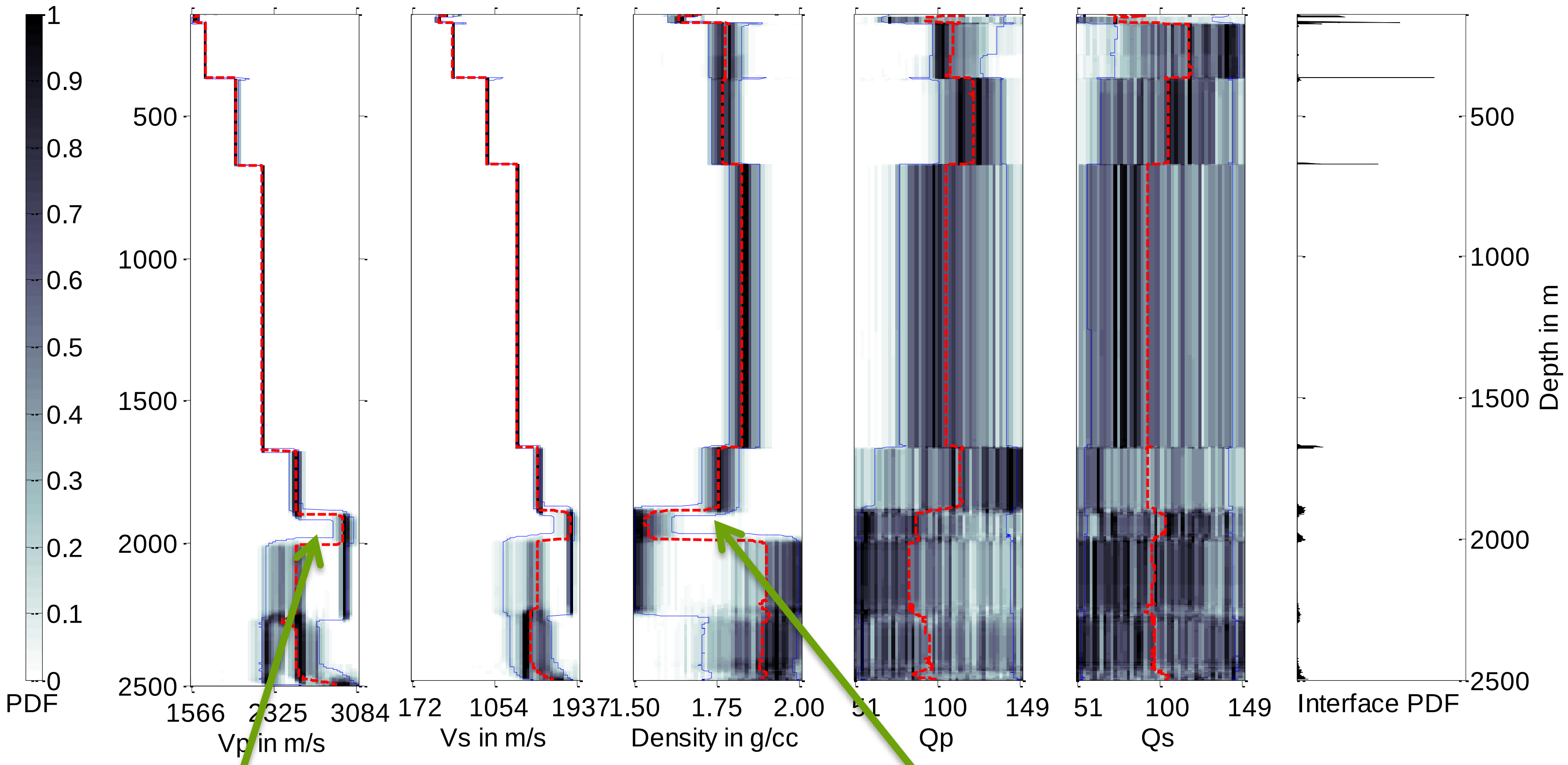
Gerstoft & Mecklenbrauker (2000)

Bayesian posterior model PDFs
(SYNTHETIC EXAMPLE)



Black = true model, Red = Median model, Blue = P5 and P95

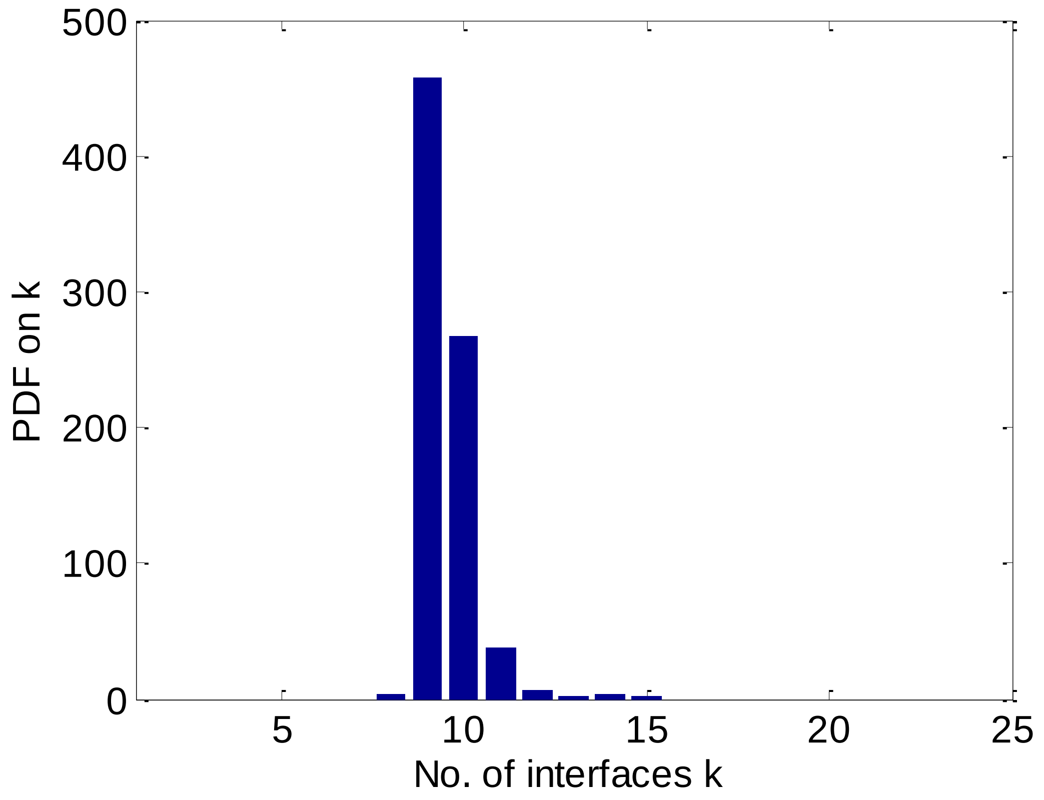
PDFs normalized at every depth
(SYNTHETIC EXAMPLE)



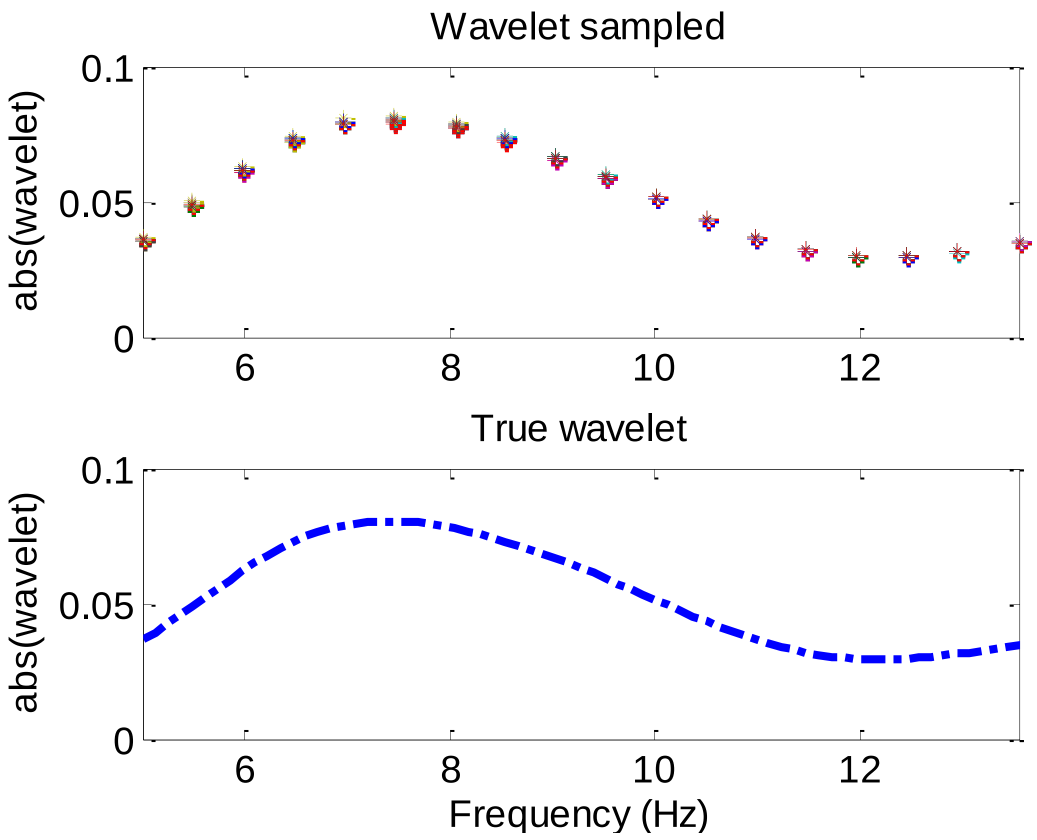
Vp PDFs cluster at high values

Rho PDFs cluster at low values

Posterior vs true wavelet
(SYNTHETIC EXAMPLE)



$$\hat{S}_l(\mathbf{m}) = \frac{\mathbf{G}_l^\dagger(\mathbf{m})\mathbf{d}_l}{\mathbf{G}_l^\dagger(\mathbf{m})\mathbf{G}_l(\mathbf{m})},$$



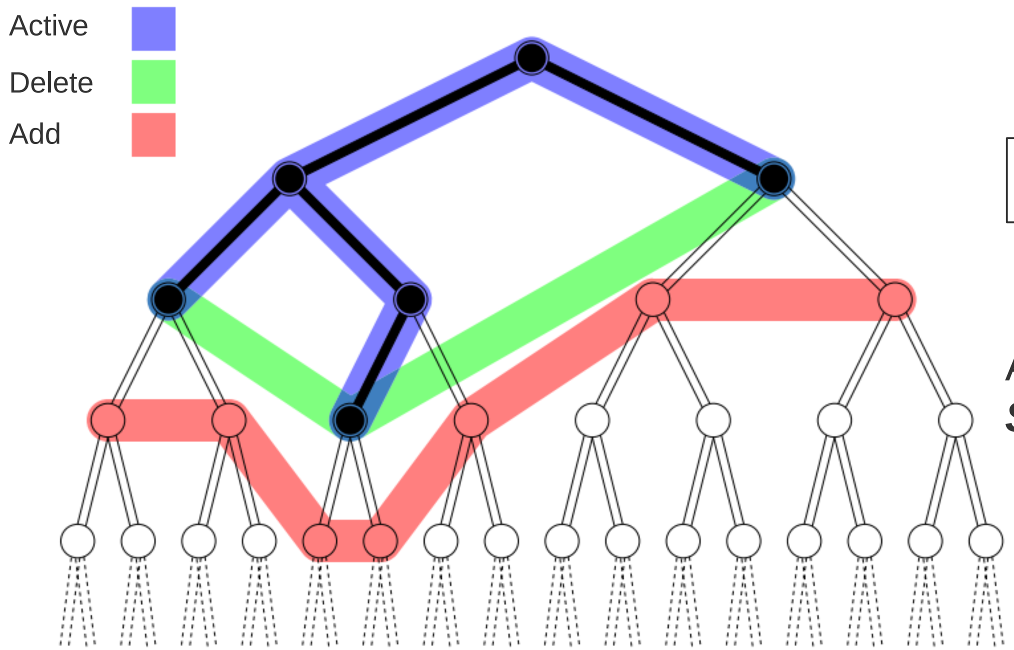
Trans-dimensional (trans-D) Bayesian inversion

Ordinary McMC	Change model parameters while sampling
trans-D McMC	***Add/delete*** parameters while sampling

$$p(k|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m_k}, k) \cdot \left[f_1 . f_2 ... f_k \right].$$

Feasible Trans-D beyond 1D

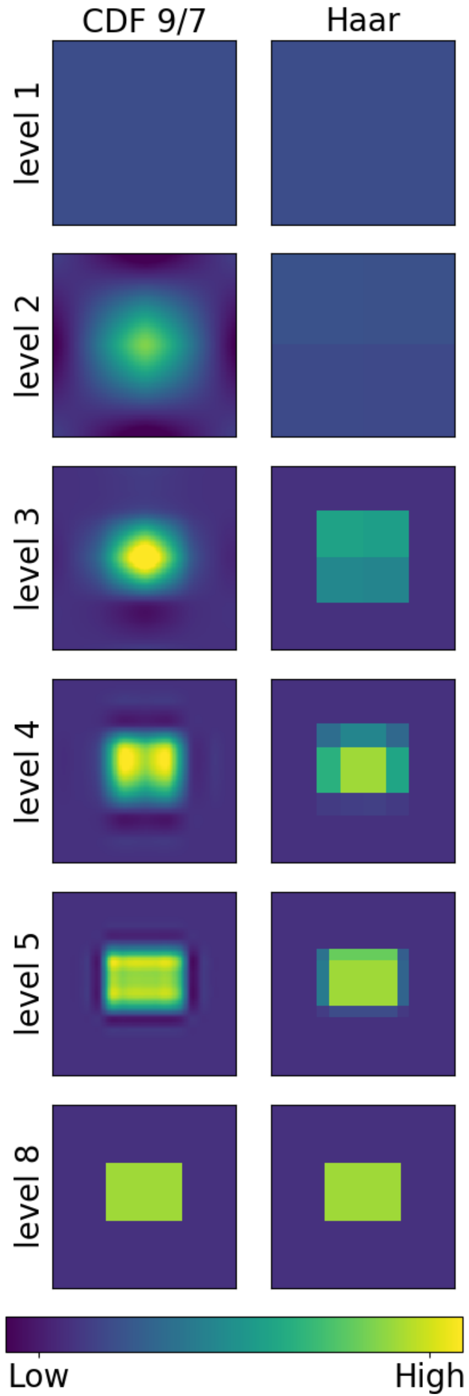
Ordinary McMC	Change model parameters while sampling
trans-D McMC	***Add/delete*** parameters while sampling
tree based trans-D McMC	Do trans-D on ***wavelet transform trees***



Works for 1D, 2D or 3D

After ***Hawkins and Sambridge (2015)***

Required coefficients

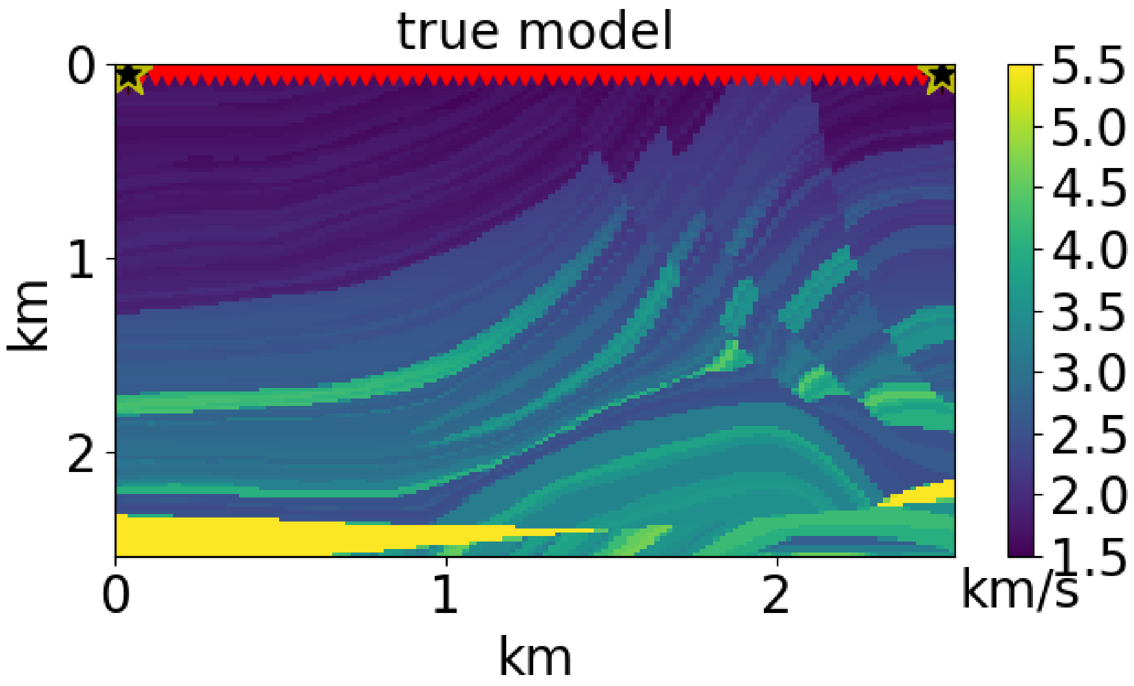


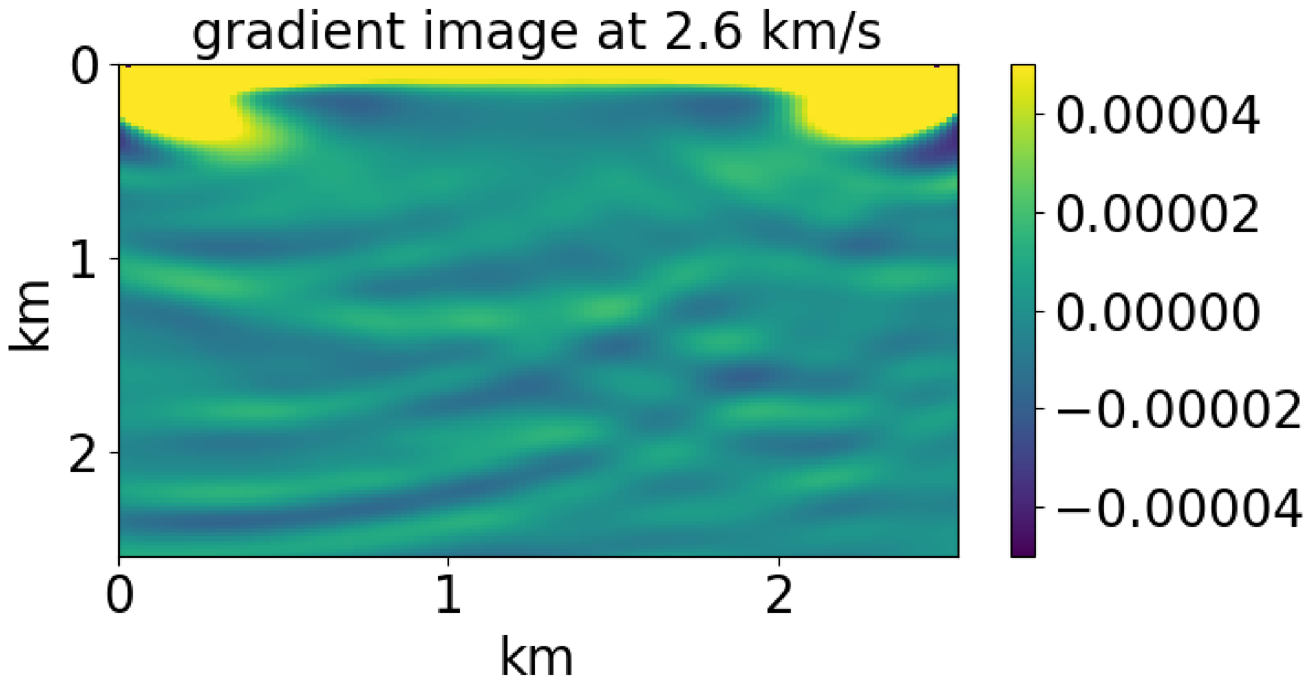
level 1	1
level 2	4
level 3	16
level 4	64
level 5	128
level 8	16384

As a sum of basis functions, this is how many coefficients we need for a given level of approximation.

Choice of basis is important!

Gauss-Newton: Fails due to cycle skipping



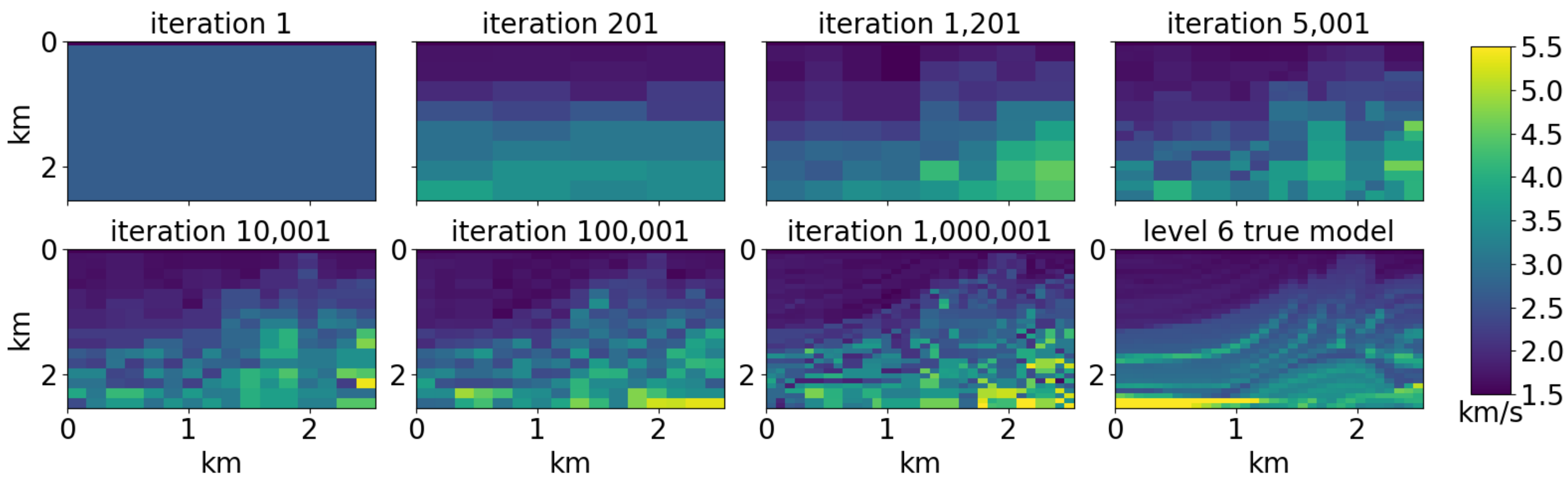


G-N was not able to update with incorrect background velocity.

$$\phi(\mathbf{m}) = \frac{1}{2}[\mathbf{d} - \mathbf{f}(\mathbf{m})]^t[\mathbf{d} - \mathbf{f}(\mathbf{m})]$$

$$\nabla_{\mathbf{m}}\phi = \mathbf{J}^t[\mathbf{f}(\mathbf{m}) - \mathbf{d}]$$

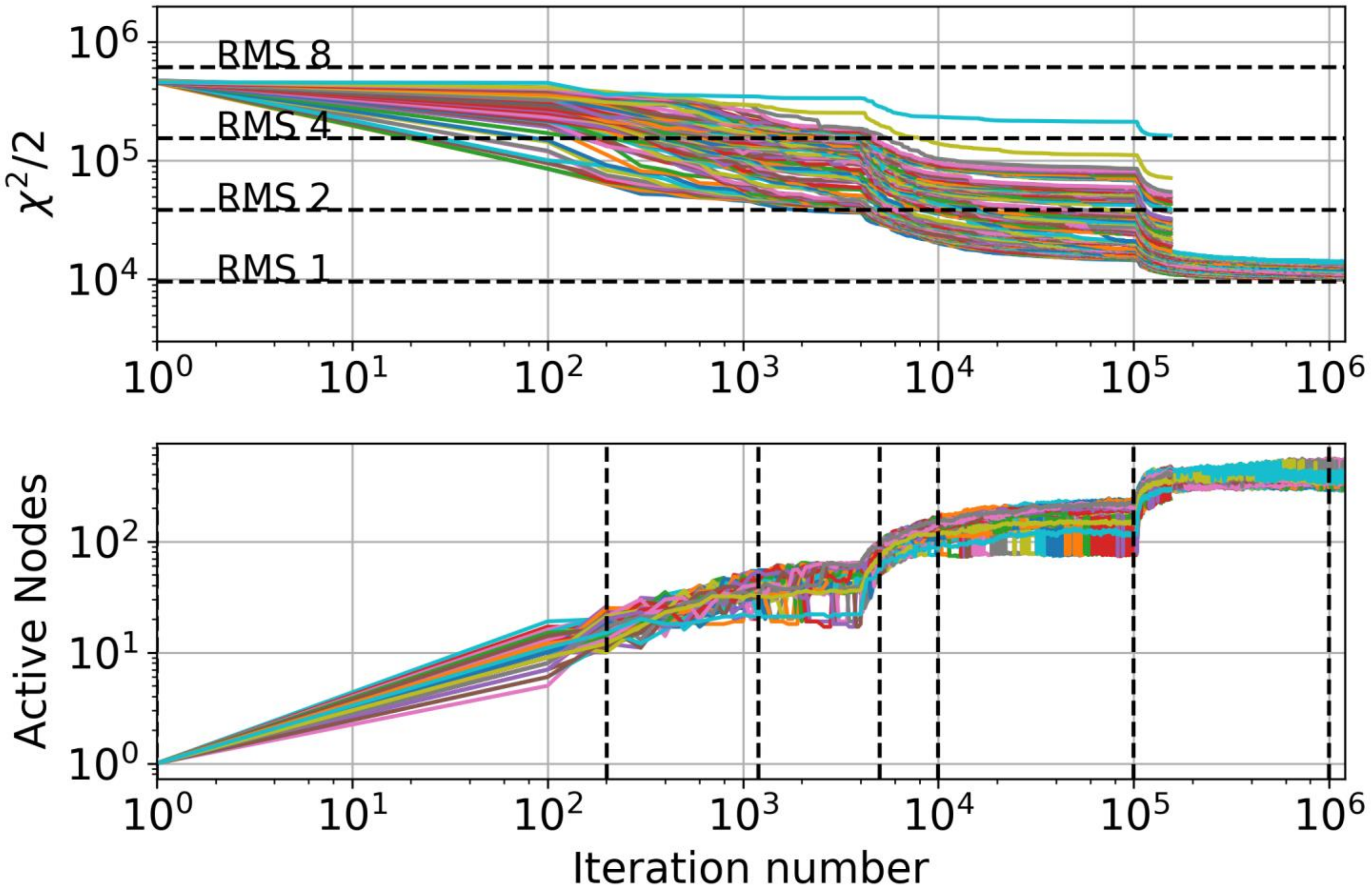
Trans-D sampling: On the other hand …



Only ***2 shots*** were used for this inversion

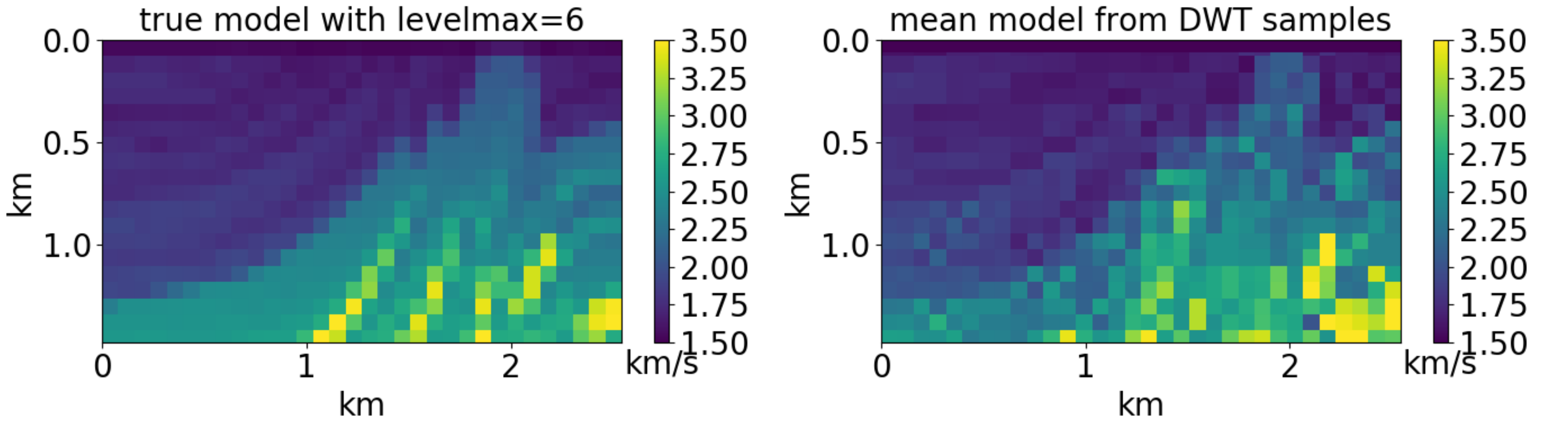
Background velocities are quickly recovered, with more detail appearing later

Trans-D sampling progress



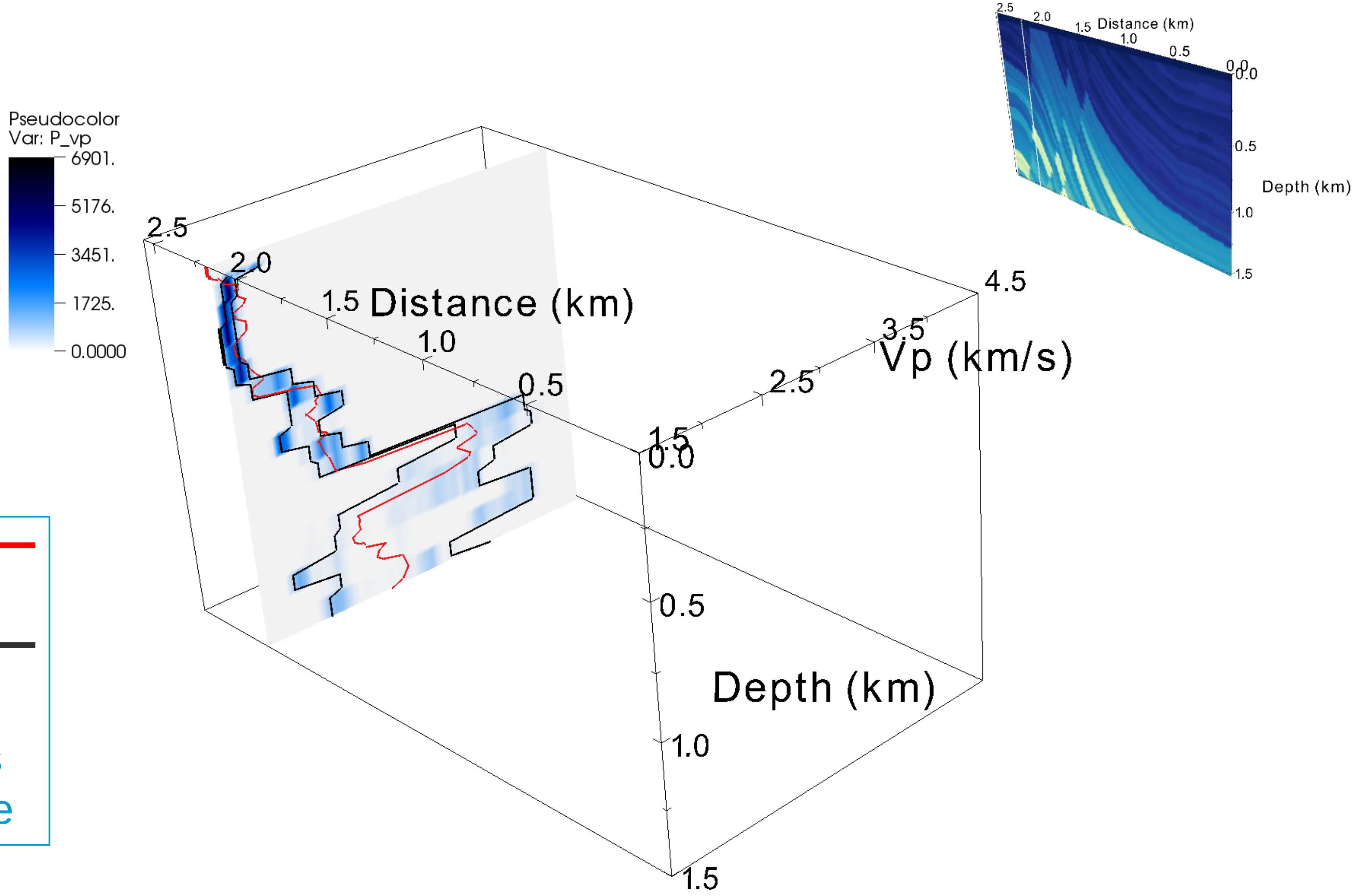
No more than 450 coefficients used

Level 6 DWT of true model and mean recovered model at same level



But why provide only a summary statistic - What if the uncertainties are multi-modal?

Uncertainty on inverted Vp

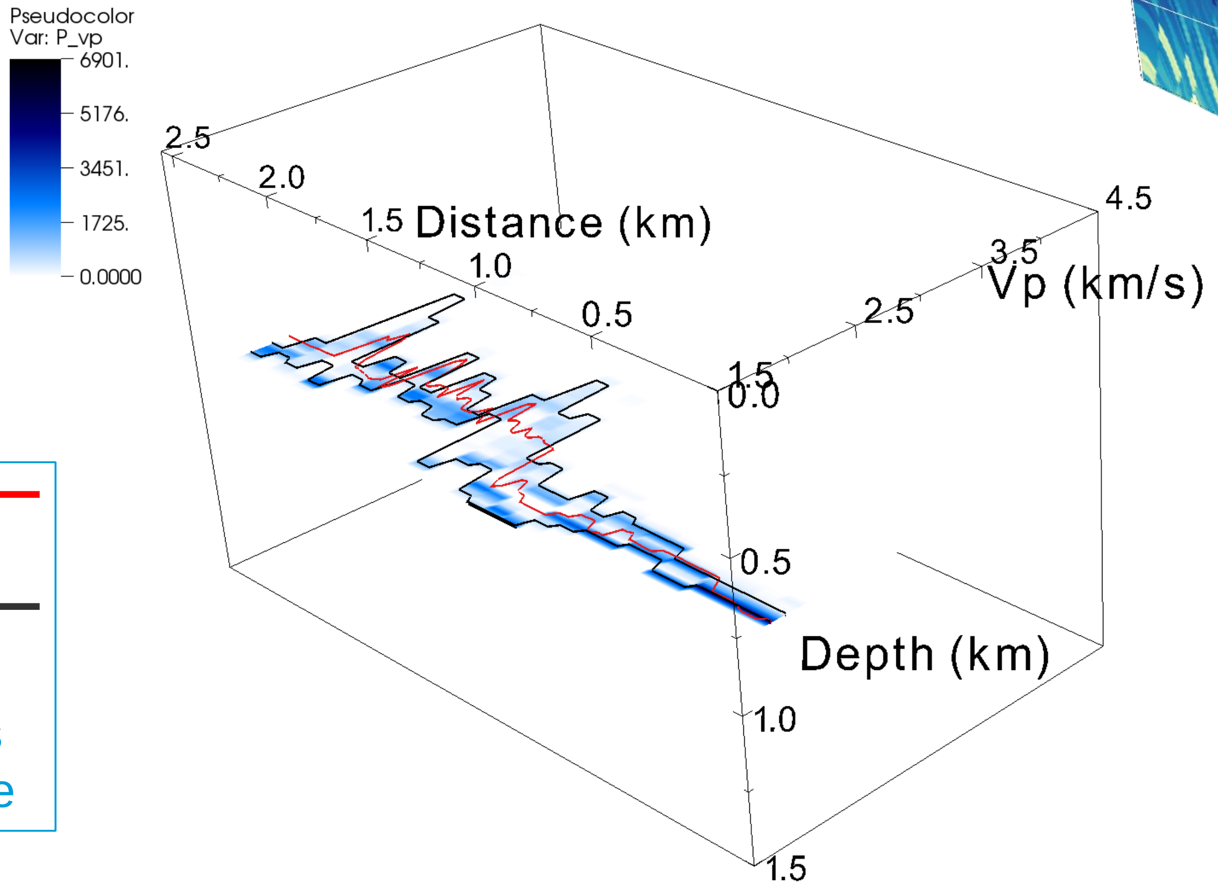


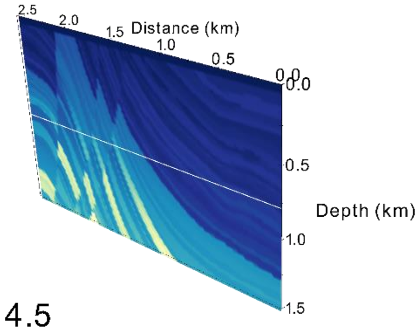
True model

P5, P95

Darker blue is more probable

Uncertainty on inverted Vp



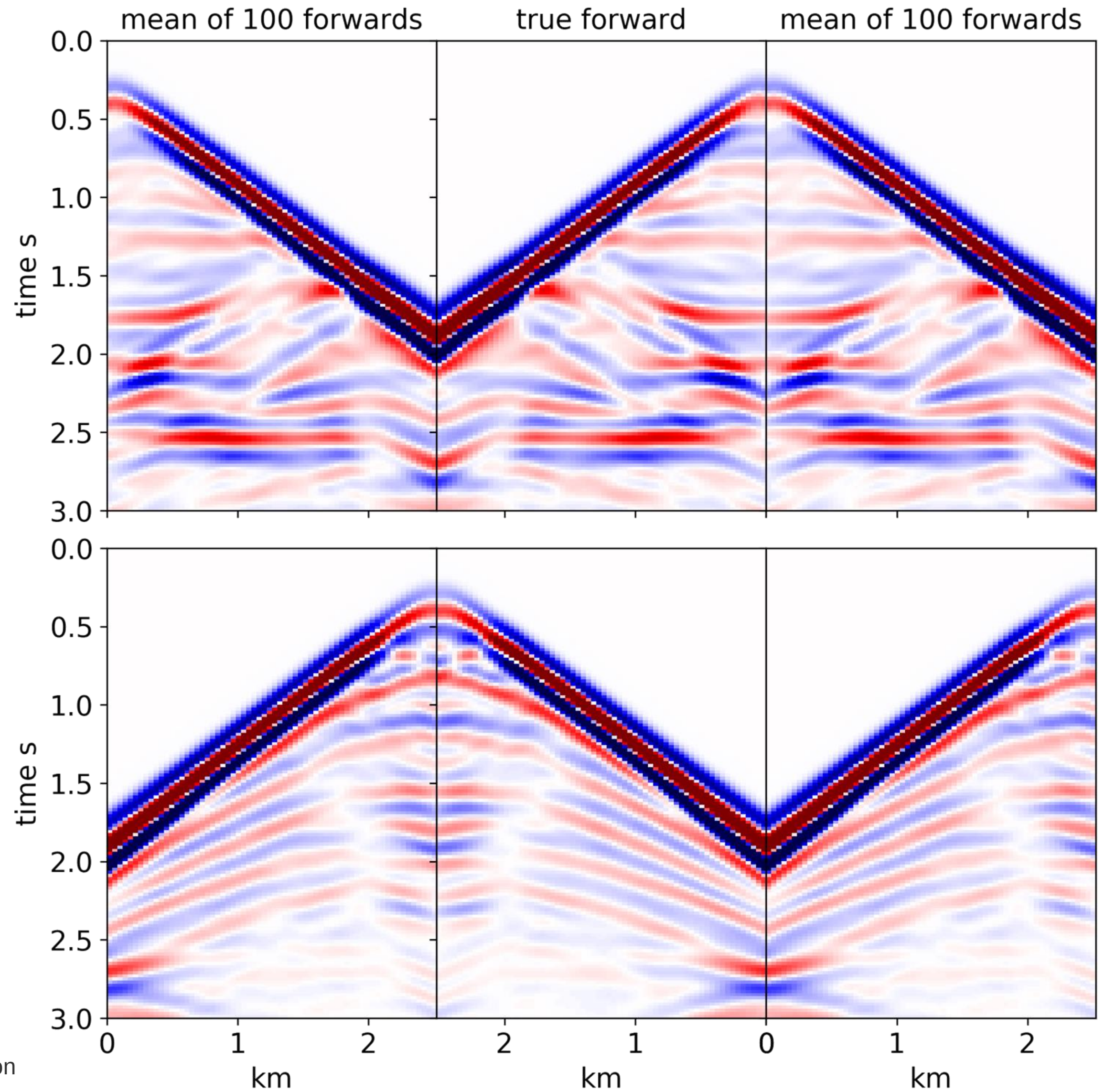


True model

P5, P95

Darker blue is more probable

Data match for 2 shots



The AVO characteristics as well as kinematics for both shot gathers are well matched

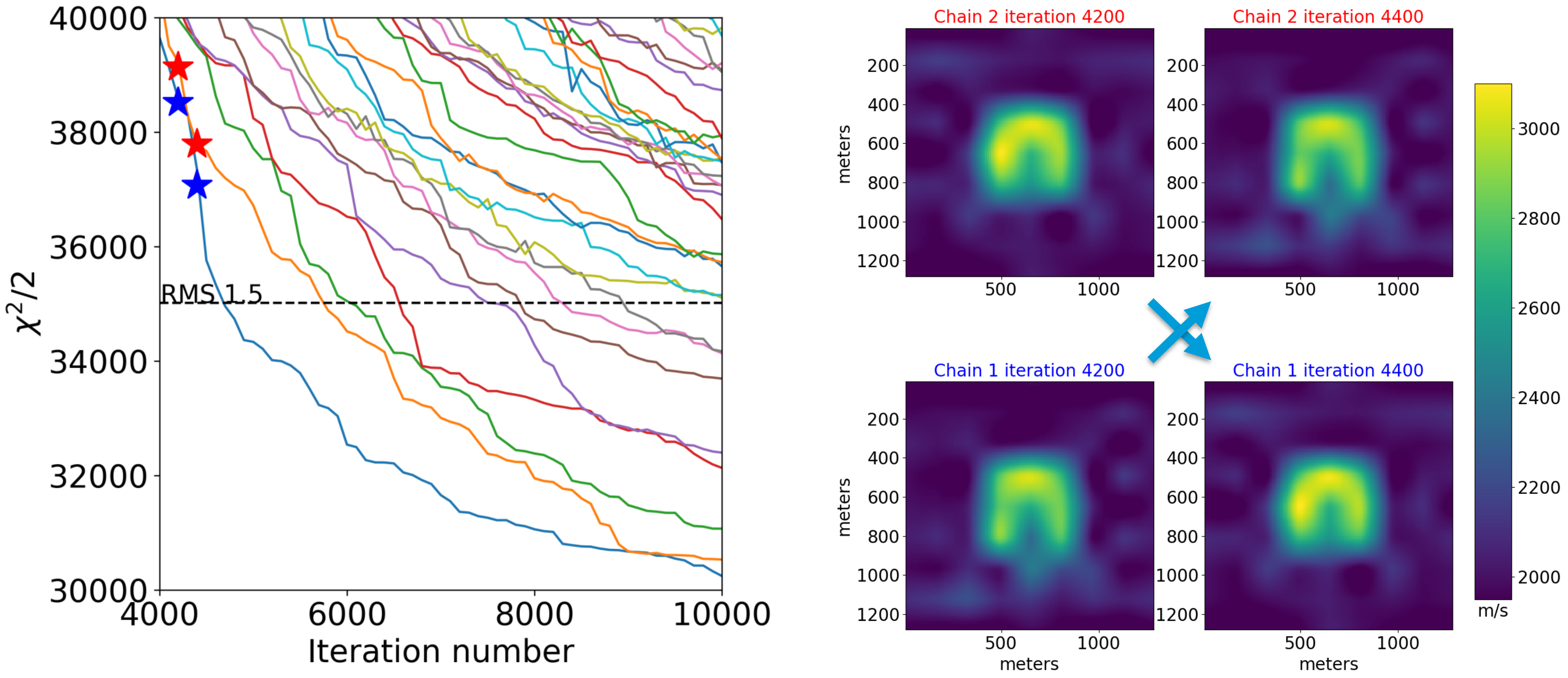
This includes multiples, refractions as well as reflections!

Conclusions

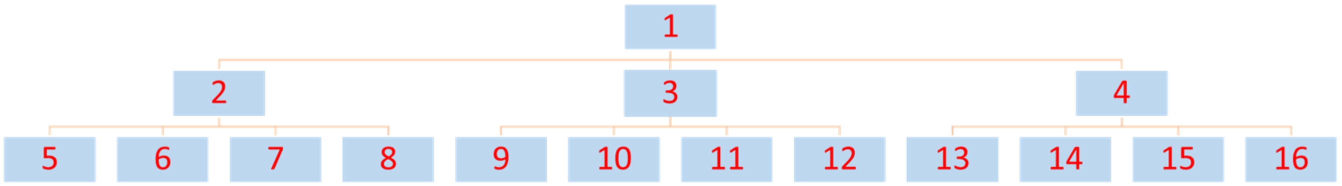
- Current **production methods** (AVA inversion) for elastic parameters have **large uncertainty**
  - May even be **Zoeppritz incompatible**
- **Standard** optimization methods (FWI) suffer from
  - Local minima problem (**cycle skips**)
  - Massive **crosstalk** problem (trade-offs)
- **Stochastic methods** can avoid these problems by
  - Dimension reduction**, Trans-D, Parallel Tempering**
- Key challenges
  - **Large number** of forwards, **cost** of forwards is very high
  - **can be addressed** using
    - **gradient** based sampling
    - using **optimized FD** engines
    - Modeling a **reduced basis** set directly
    - Using **less** shots

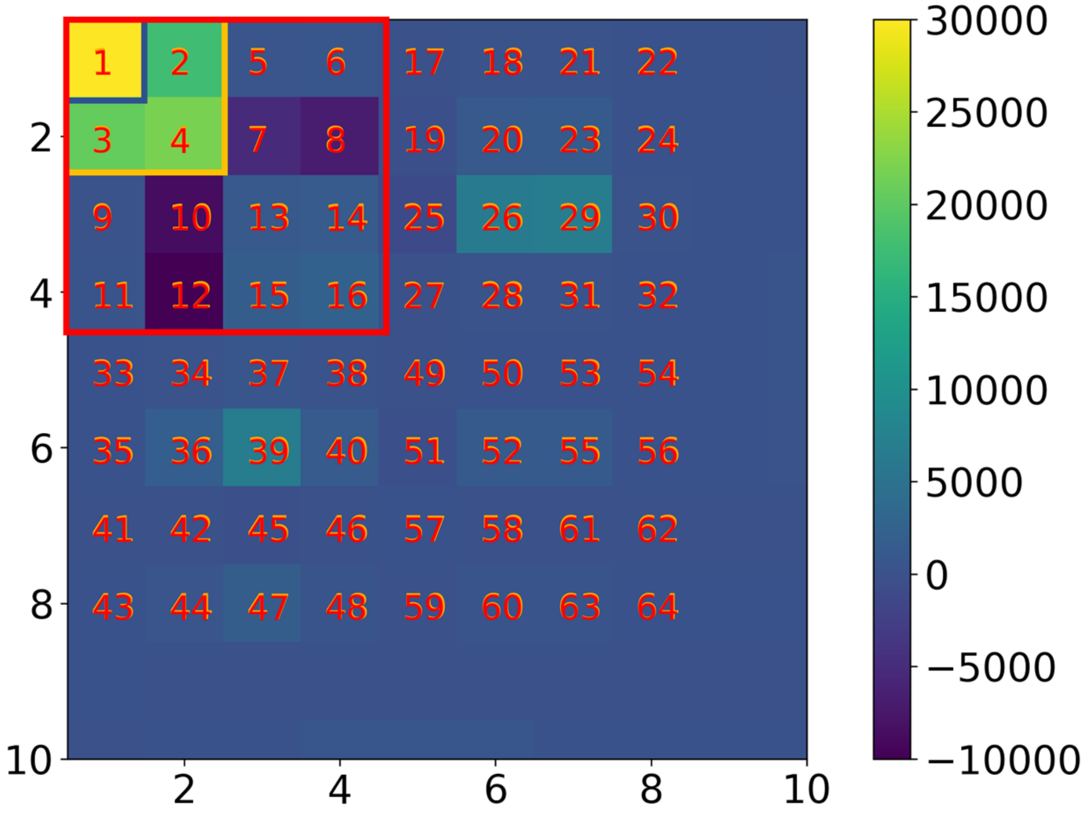
Backup

Parallel tempering in action



The discrete wavelet transform and its tree representation

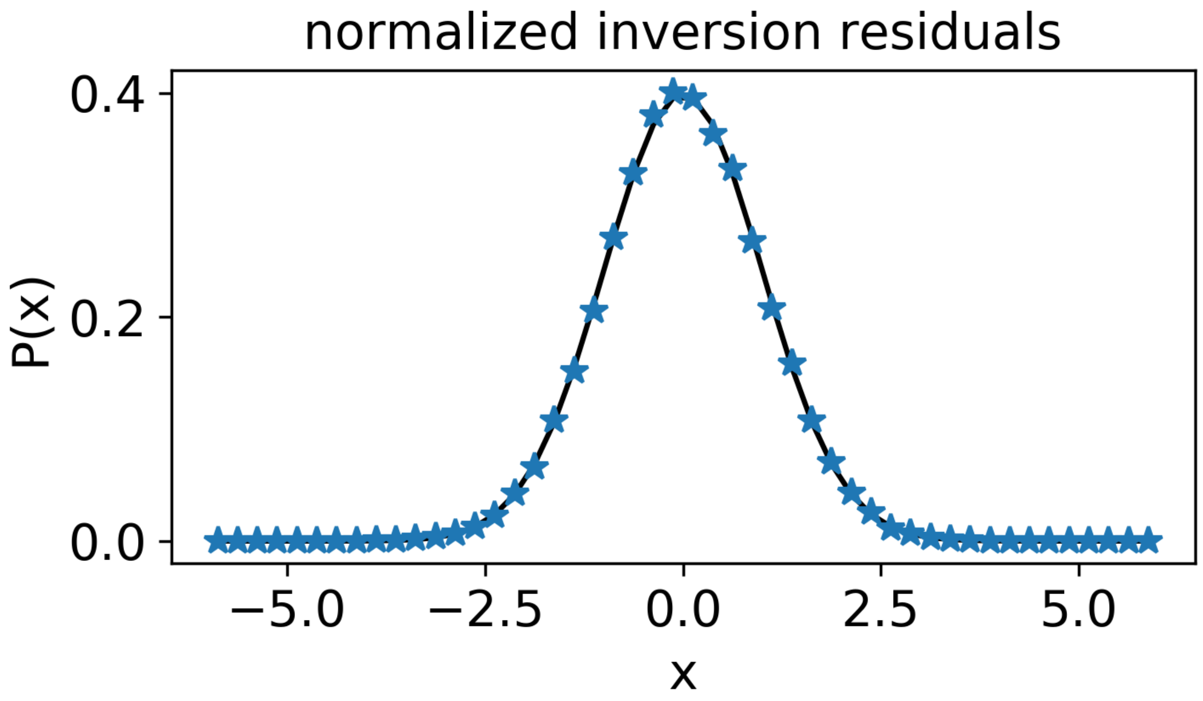




Most elements are near zero!

Possibility of sparse representations

Histogram of residuals from 100 sampled models



Gaussian assumption is met