Probabilistic Tomography: A Pragmatic Bayesian Approach

{ Jeannot Trampert

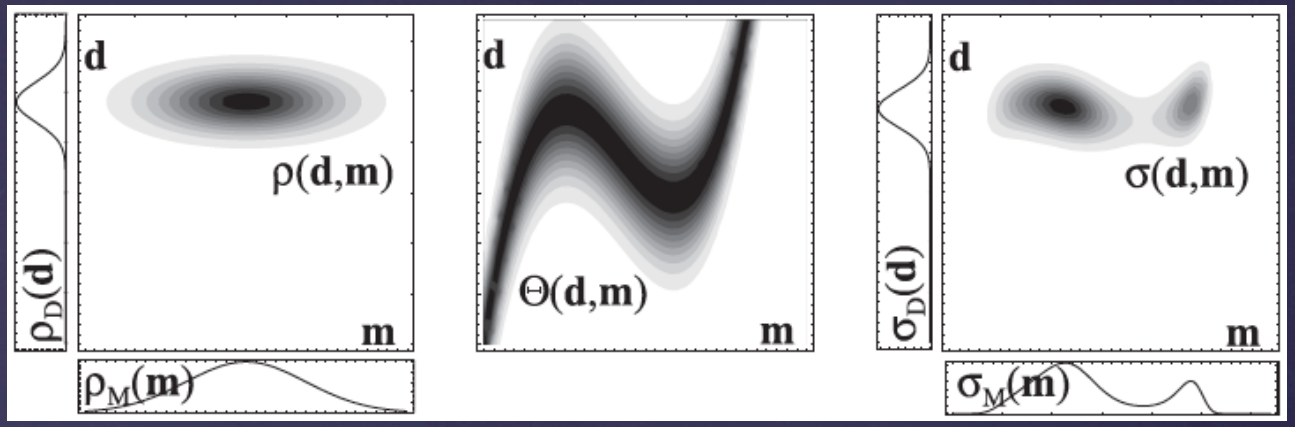
Universiteit Utrecht

erc

European Research Council

Established by the European Commission

Prior * Likelihood = Posterior



Tarantola's approach to Bayes: Conjunction of states of information

The essence of Bayesian inference

- Explicit modeling assumptions
- General framework for non-linear problems
- Handles uncertainty in a natural manner
- Embodies Occam's razor
- Given enough data, priors don't matter
- Priors are objective (non-informative) or subjective (informative) and sometimes improper

- Assume Gaussian distributions and linearity → Explicit expressions
- Sampling the posterior directly (Neighbourhood Algorithm, Metropolis-Hastings)
- Sampling the prior and predict the posterior (Neural Networks)

The technicalities of Bayesian inference

- Subjective prior and sampling the posterior (Neighbourhood Algorithm, Metropolis-Hastings)
- Informative prior and sampling of the prior only (Neural Networks)

Our pragmatic approach to Bayesian inference

→ The inference is relative: knowledge gained from data is added to prior

Explicit expressions using Gaussian statistics (e.g. Tarantola 2005)

Minimize $$C = (\mathbf{d} - \mathbf{Gm})^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{Gm}) + \mathbf{m}^T \mathbf{C}_m^{-1} \mathbf{m},$$

which gives a least-squares solution for the most likely (mean) model together with the posterior covariance

Anisotropic phase velocity maps

(Visser et al., 2008)

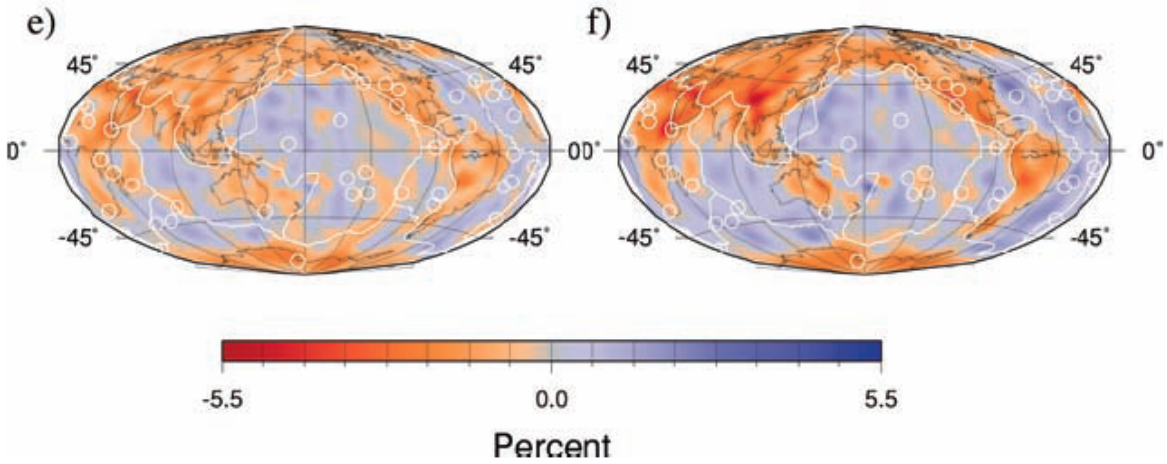


Figure 6. Relative isotropic phase velocity maps with respect to PREM for Rayleigh (a) first higher mode at 148.56 s, (b) second higher mode at 40.028 s, (c) third higher mode at 77.795 s, (d) fourth higher mode at 35.078 s, (e) fifth higher mode at 56.074 s and (f) sixth higher mode at 35.141 s.

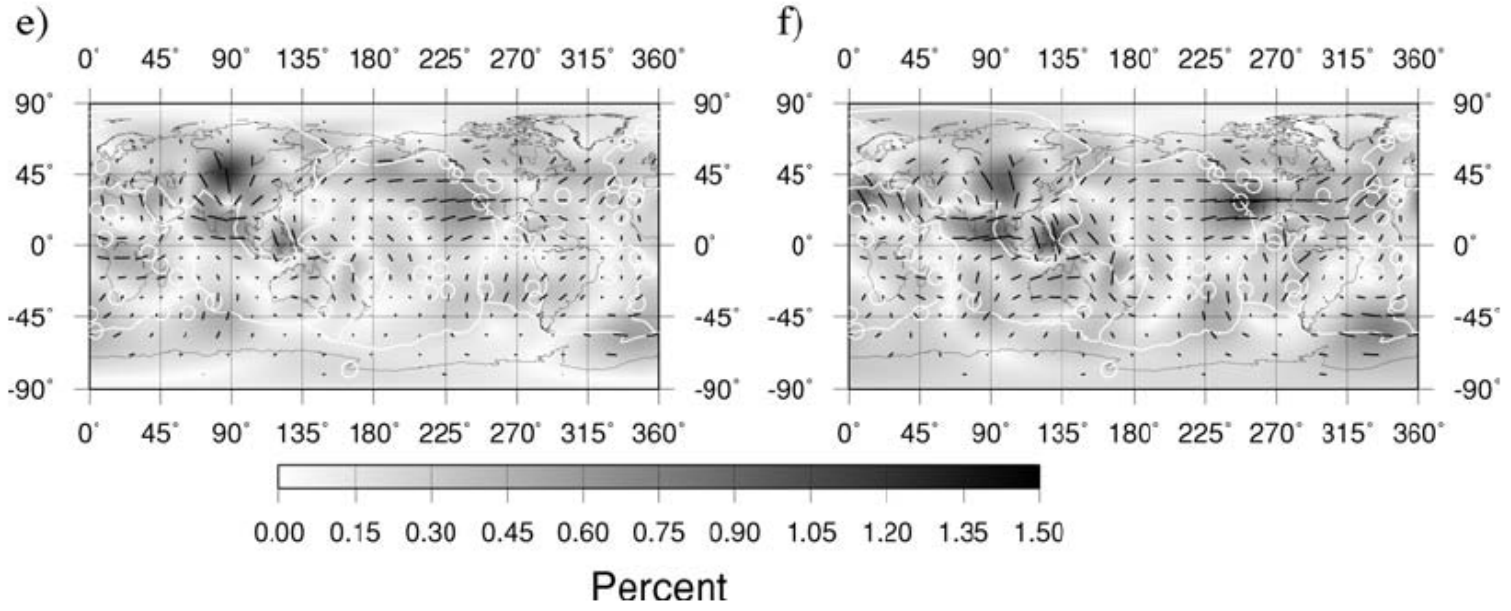
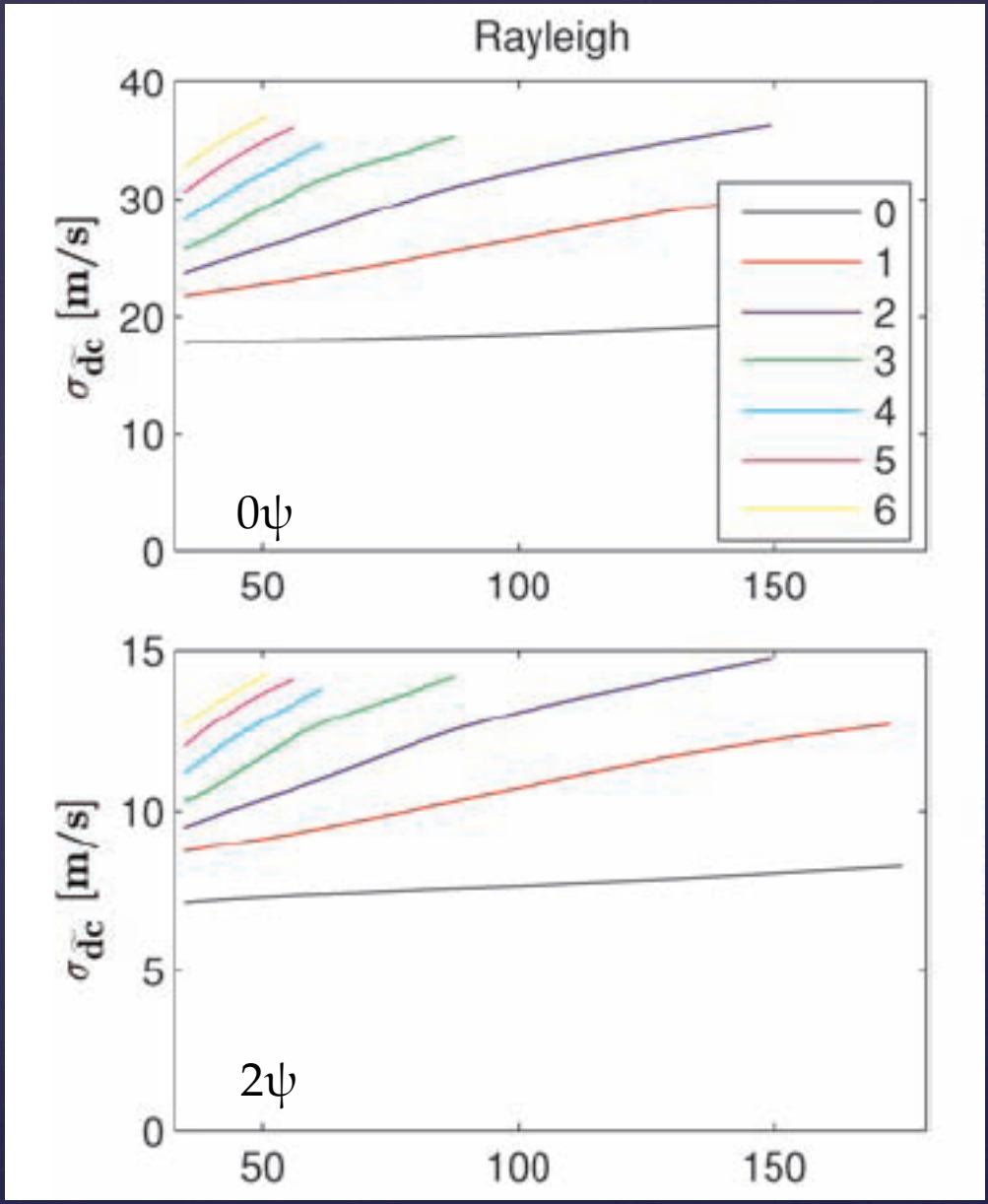


Figure 8. Azimuthally anisotropic 2ψ phase velocity maps for Rayleigh. The grey scale in the background corresponds to the peak-to-peak amplitude of anisotropy expressed relative to the average phase velocity calculated from PREM. The black lines represent the fast directions which are also scaled to the amplitude shown in the background. The plate boundaries and hotspots are indicated in white. Panels (a)–(f) show the different modes and periods as indicated in Fig. 6.



Perturbations are typically 100-250 m/s from PREM

Neighbourhood Algorithm (Sambridge, 1999)

Stage 1: Guided sampling using Voronoi cells.
For each sample the likelihood is evaluated.

Stage 2: Importance sampling (Gibbs) of the posterior using the Voronoi approximation.

Seismic models from probabilistic tomography
(Mosca et al., 2012)

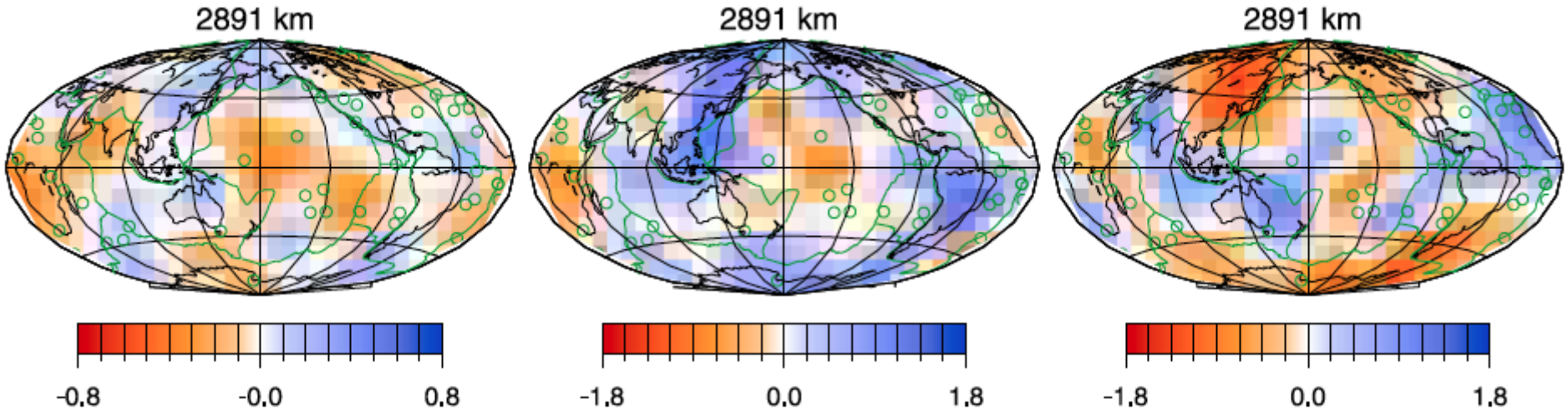


Figure 3. (left) Compressional wave speed, (middle) shear wave speed, and (right) density perturbation maps at six depths, together with hot spots (green circles) and plate boundaries (green lines). Red (blue) regions denote slower (faster) than average seismic wave speed and a decrease (increase) in density. Measurements were calculated for even degrees 0, 2, 4 and 6. Perturbations are given in percent with respect to PREM. We divided each layer into 264 nodes of 15° × 15° area and estimated the mean and standard deviation in a grid node from the 1-D marginal *pdf* provided by the Neighborhood Algorithm.

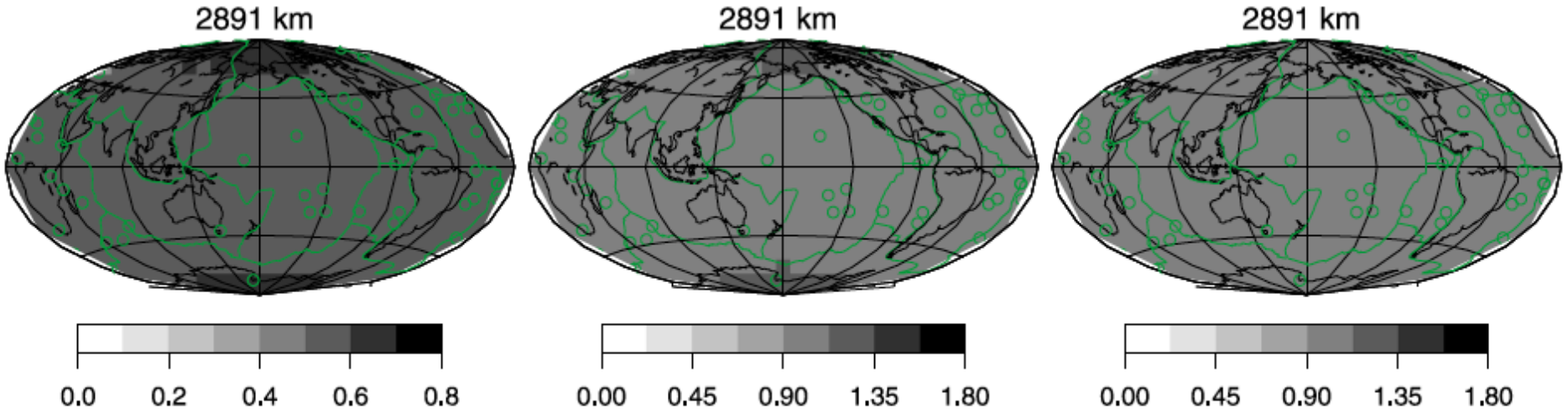


Figure 4. As Figure 3, but for the lateral distribution of the uncertainty of $d\ln v_p$, $d\ln v_s$ and $d\ln \rho$.

Metropolis -Hastings

Using 100,000 thermochemical models sampling variations of T (2300-4800K), Pv (60-100%), Fe (0-20%), Al and Si (0-15%), together with Perplex (Connolly, 2009), we match the seismic pdfs at each point.

Thermo-chemical structure from probabilistic tomography

(Mosca et al., 2012)

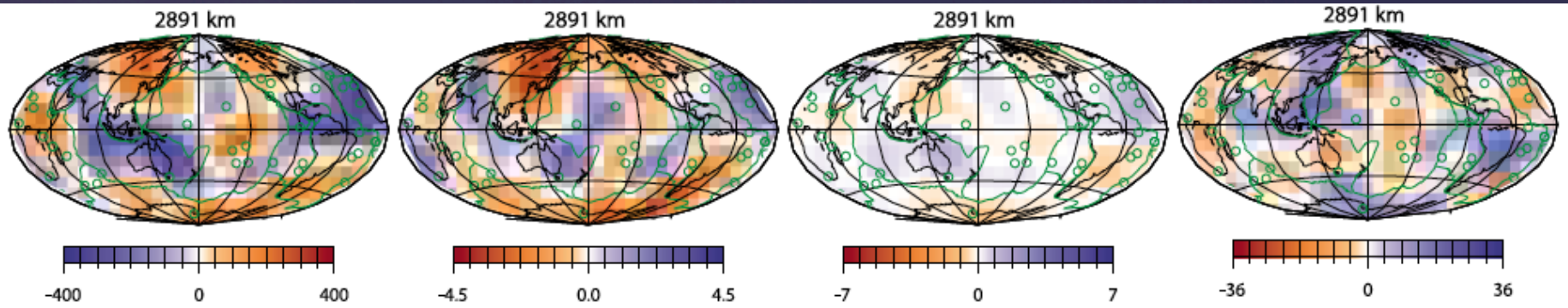


Figure 9. Lateral variations of temperature, iron, silicates and post-perovskite at various depths of the lower mantle. Temperature variations are expressed in Kelvin, while iron, perovskite and post-perovskite perturbations are given in percent. Furthermore, variations are measured relative to thermochemical models which fit PREM to within 1%. Red regions denote negative chemical anomalies and a higher than average temperature, while blue regions are associated with positive chemical variations and a lower than average temperature. Maps show hot spots (green circles) as well as plate boundaries (green lines).

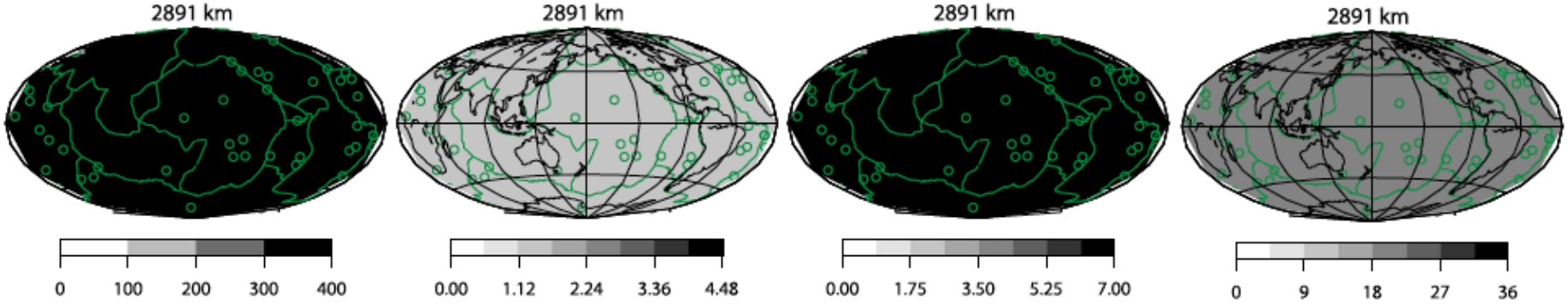


Figure 10. As Figure 9, but for the uncertainty of thermal and chemical anomalies.

Neural Networks

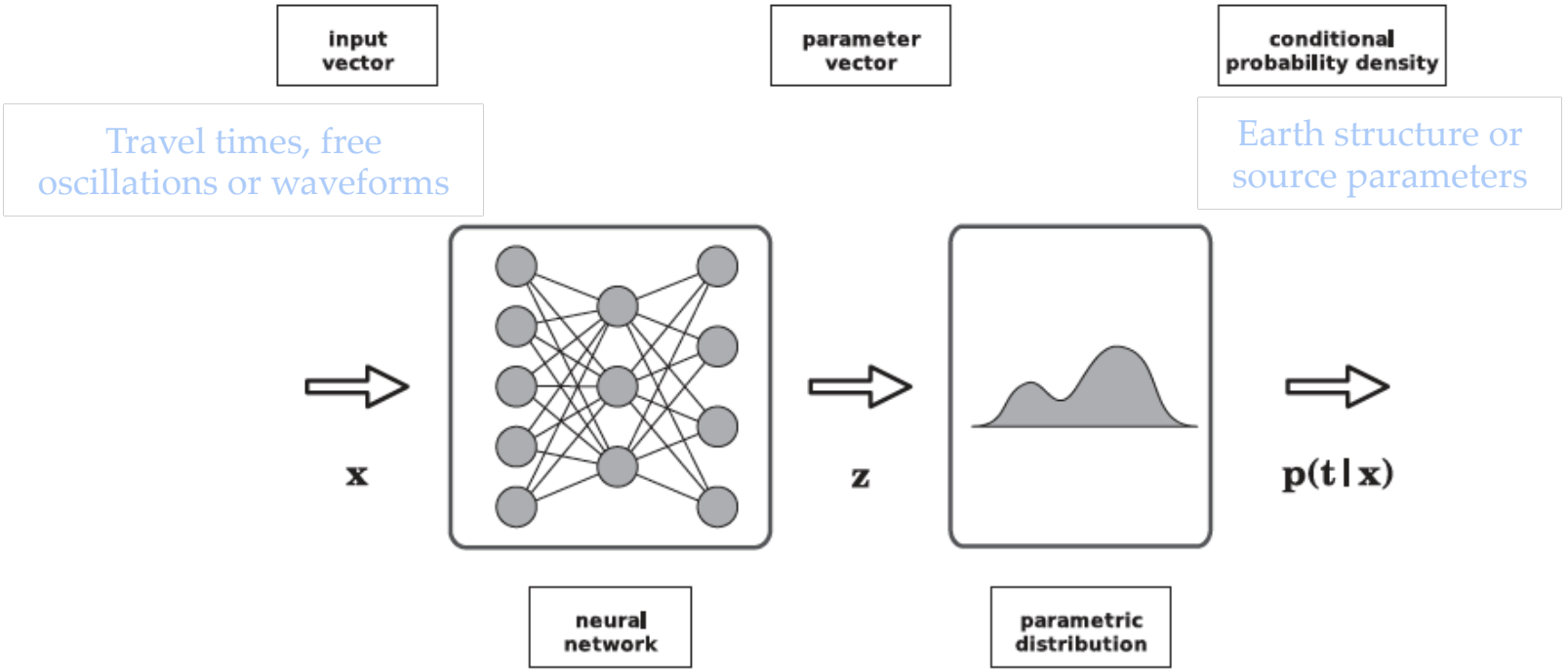


Figure 2. A Mixture Density Network (MDN), as introduced by Bishop (1995). The output of an MDN approximates a parametric distribution $p(\mathbf{t}|\mathbf{x})$ for the target $\mathbf{t}$, conditioned on the input $\mathbf{x}$. The parameters describing this distribution are given by the output $\mathbf{z}$ of a neural network, such as the MLP shown in Fig. 1.

A set is generated from the prior distributions to train the neural network parameters. Once trained, the network predicts the marginal conditional probability of a chosen parameter given the real observation.

	MCMC	NN
probability space	joint	marginal
sampling space	posterior	prior
assumption	n/a	smoothness
misfit function	required (likelihood)	n/a
interpolation	n/a	general
zooming around high likelihood	yes	no
tuned to specific data	yes	no
tuned to specific model parameter	no	yes
repeatability	no	yes

Monte Carlo versus Neural Network

Prior models

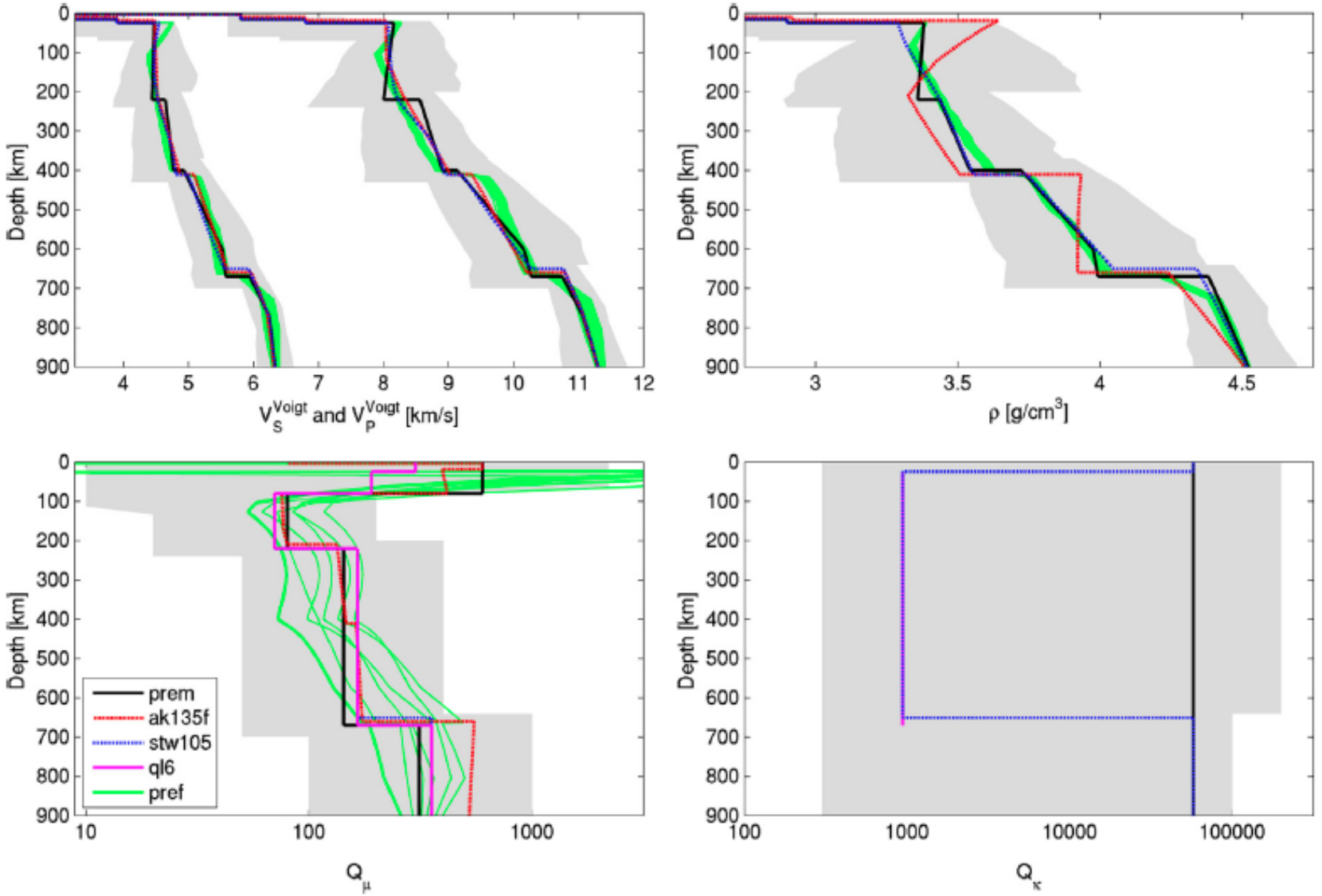


Fig. 1. Radial earth models in the upper mantle. The parameter range spanned by the prior model space is represented by the grey shaded area, along with the 1-D reference models *PREM* (black, solid), *ak135f* (red, dashed), *STW105* (blue, dotted-dashed), *QL6* (magenta, solid) and *PREF* (Cammarano et al., 2005, 99 models, green) for V_S^{Voigt} and V_P^{Voigt} (top-left panel), ρ (top-right panel), Q_μ (bottom-left panel) and Q_κ (bottom-right panel). The horizontal scale for the bottom panels is logarithmic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

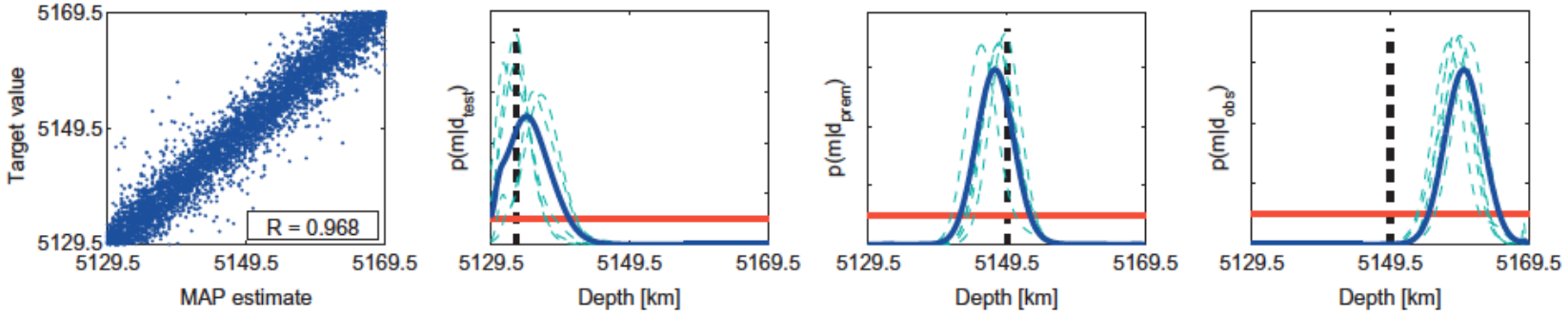


Fig. 2. Ensemble of MDNs trained on the depth of the ICB. The left-hand panel shows the performance on the 5000 test set samples, represented by the correspondence between the MAP estimate and the target value and quantified by the correlation coefficient *R*. The other panels show the 1-D marginal posterior pdf (blue line), the prior pdf (red) and the target value (black, dashed) for a test set pattern (second panel from the left), *PREM* (third panel) and the observed data (right-hand panel). Note that the *PREM* value in the right-hand panel is shown as a reference and does not represent a target for the observed data. The normalised pdfs for five individual networks in the ensemble are shown in cyan. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Posterior statistics for the seven discontinuity depths in kilometres, in terms of the MAP estimate θ and asymmetric 2σ model error bars, corresponding to $1/e^2$ levels in the unit normalised 1-D marginal posterior pdfs. The corresponding *PREM* and *ak135* values are given for comparison. The last column shows the information gain D_{KL} in bits (C).

Discontinuity	*PREM*	*ak135*	θ	$\theta - 2\sigma$	$\theta + 2\sigma$	D_{KL} [bits]
ICB	5149.5	5153.5	5160.1	5154.7	5165.7	4.7
CMB	2891.0	2891.5	2888.6	2886.5	2890.7	13.8
D″ (top)	2741.0	2740.0	2722.4	2721.0	2761.0	0.0
"660"	670.0	660.0	663.7	650.3	676.2	2.5
"410"	400.0	410.0	386.2	370.0	413.8	0.7
"220"	220.0	—	200.9	200.0	240.0	0.0
Moho	24.4	35.0	36.2	20.8	47.3	3.4

de Wit et al., 2014

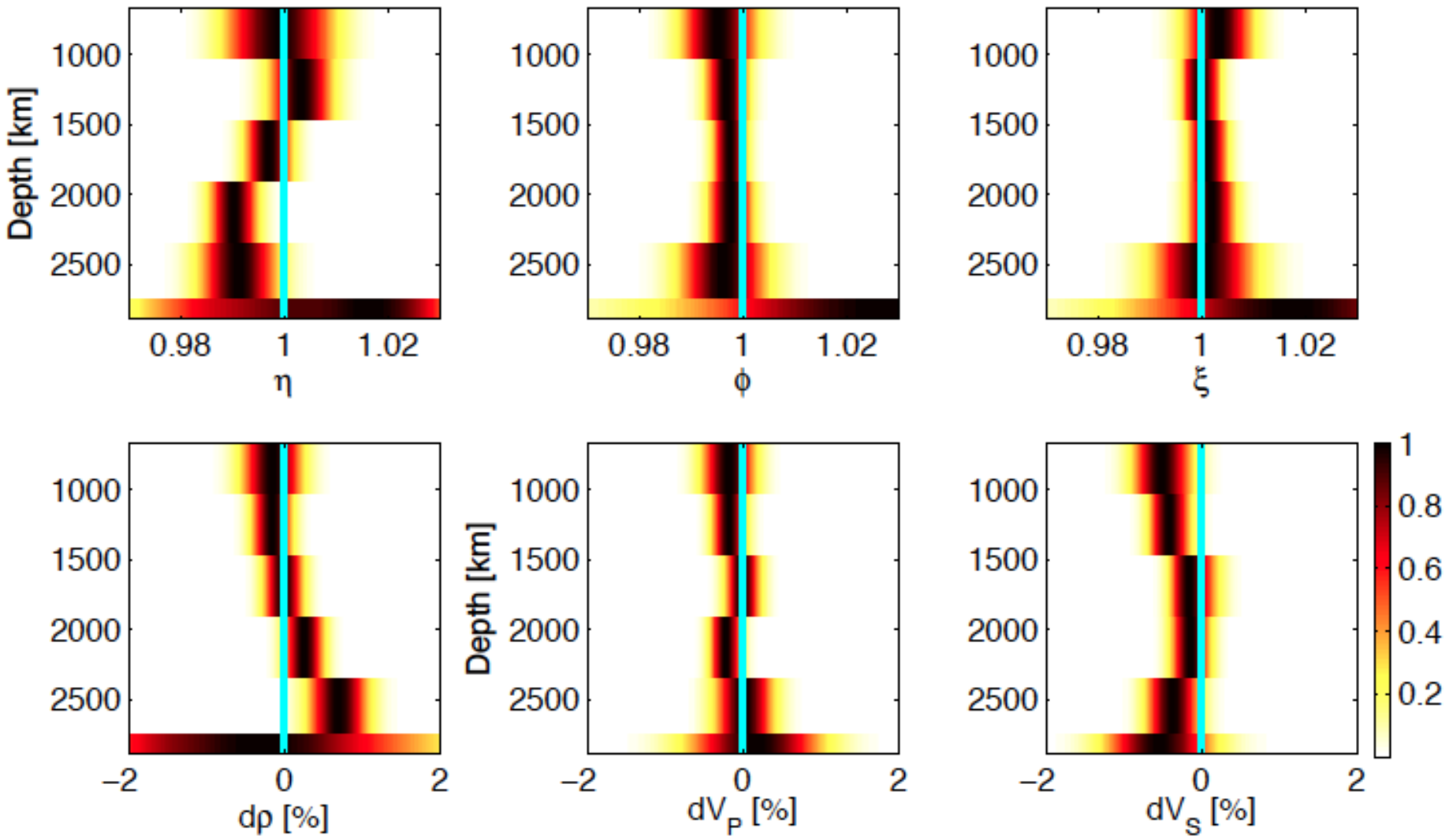
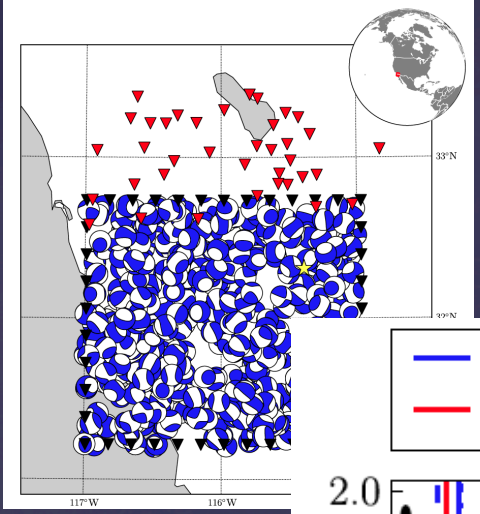
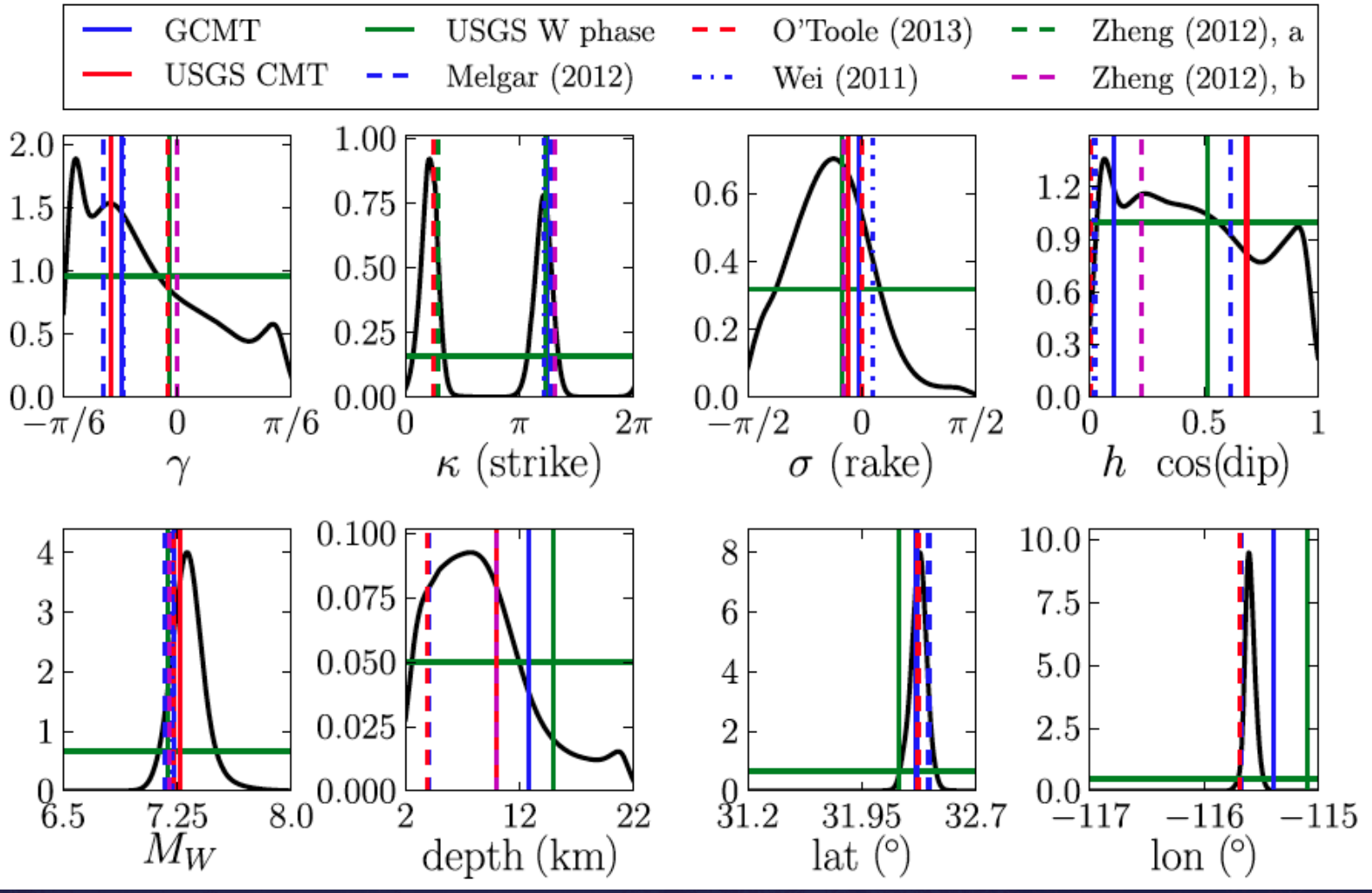


Figure 1: 1-D marginal posterior pdfs for the averages of the six seismic parameters in the six lower mantle layers (Section S1). The bottom layer in each panel represents the D" region. *PREM* (cyan line) is isotropic in the lower mantle and is given as a reference. The velocities and density are expressed as percentage deviations from *PREM*. The probability for each 1-D pdf is rescaled so that the maximum equals 1. Asymmetric 1σ and 2σ error bars correspond to the $1/e^{1/2}$ (0.61) and $1/e^{2}$ (0.14) contours, respectively.





Kaeufl et al., 2014

Thanks to

- Laura Cobden
- Paul Kaeufl
- Andrew Valentine
- Karin Visser
- Ralph de Wit