JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

The Thermo-Chemical and Physical Structure beneath the North American Continent from Bayesian Inversion of Surface-wave Phase Velocities

A. Khan[1], A. Zunino[2], F. Deschamps[3]

A. Khan, Institute of Geochemistry and Petrology, Swiss Federal Institute of Technology, Clausiusstrasse 25, 8092 Zürich, Switzerland (amir.khan@erdw.ethz.ch).

A. Zunino, Dipartimento per lo Studio del Territorio e delle sue Risorse, Sez. Geofisica, University of Genova, Viale Benedetto XV 5, 16132 Genova, Italy (andrea.zunino@dipteris.unige.it).

F. Deschamps, Institute of Geophysics, Swiss Federal Institute of Technology, Sonneggstr. 5, 8092 Zurich, Switzerland (frederic.deschamps@erdw.ethz.ch).

[1]Institute of Geochemistry and Petrology, Swiss Federal Institute of Technology, Zurich, Switzerland.

[2]DipTeRis, University of Genova, Genova, Italy.

[3]Institute of Geophysics, Swiss Federal Institute of Technology, Zurich, Switzerland.

Abstract. We jointly invert local fundamental-mode and higher-order surface-wave phase-velocities for radial models of the thermo-chemical and anisotropic physical structure of the Earth's mantle to ∼1000 km depth beneath the North American continent. Inversion for thermo-chemical state relies on a self-consistent thermodynamic method whereby phase equilibria and physical properties (P-, S-wave velocity and density) are computed as functions of composition (in the Na_2O-CaO-FeO-MgO-Al_2O_3-SiO_2 model system), pressure and temperature. We employ a sampling-based strategy to solve the non-linear inverse problem relying on a Markov chain Monte Carlo method to sample the posterior distribution in the model space. A range of models fitting the observations within uncertainties are obtained from which any statistics can be estimated. To further refine sampled models we compute geoid anomalies for a collection of these and compare with observations, exemplifying *a posteriori* filtering through the use of additional data. Our thermo-chemical maps reveal the tectonically stable older eastern parts of North America to be chemically depleted (high Mg#) and colder (>200°C) relative to the active younger regions (western margin and oceans). In the transition zone the thermo-chemical structure decouples from that of the upper mantle, with a thermal anomaly appearing beneath the cratonic area that likely extends into the lower mantle. In the lower mantle no consistent large-scale thermo-chemical heterogeneities are observed, although our results do suggest distinct upper and lower mantle compositions. Concerning anisotropy

structure, we find evidence for a number of distinct anisotropic layers pervading the mantle, including transition zone and upper-most lower mantle.

1. Introduction

Seismic tomography has proven a powerful tool in its ability to provide information on the internal structure of the Earth and has done much to advance our understanding of its dynamics. Since its advent in the late 1970s [e.g. *Aki et al.,* 1977; *Sengupta & Toksöz,* 1977; *Dziewonski et al.,* 1977], images obtained from the latter have proceeded to reveal features at an unprecedented level of resolution, laterally as radially, that is continuously being improved [e.g. *Ritsema et al.,* 2010].

While there is good reason to believe that the large-scale global velocity structure is relatively well-resolved, given current consensus among studies that employ different data and modelling techniques [e.g. *Grand et al.,* 1997; *Masters et al.,* 2000; *Trampert & Woodhouse,* 2001; *Boschi & Ekström,* 2002; *Shapiro & Ritzwoller,* 2002; *Romanowicz,* 2003; *Ritsema et al.,* 2004; *Panning & Romanowicz,* 2006; *Kustowski et al.,* 2008; *Rawlinson et al.,* 2010], there is less agreement with regard to the smaller scales [e.g. *Trampert & Van der Hilst,* 2005]. This ambiguity results partly from the use of iterative least-squares approaches that are based upon a linearized forward model to invert global or regional seismic data. This potentially biases the resulting images of the Earth in the direction of the particular starting model and regularization scheme chosen, whereby assessment of reliable model parameter uncertainty estimates become complicated, in addition to complicating direct comparison between models obtained from different studies [e.g. *Trampert,* 1998; *Boschi & Dziewonski,* 1999; *Shapiro & Ritzwoller,* 2002; *Trampert & Van der Hilst,* 2005; *Khan et al.,* 2010].

In recognition of this, a number of recent studies [e.g. *Shapiro & Ritzwoller,* 2002;

Visser et al., 2008a,b; *Bodin et al.,* 2009; *Khan et al.,* 2009, 2010; *Mosca,* 2010] have employed non-linear stochastic-based inversion methods. Stochastic approaches, which typically rely on Markov chain Monte Carlo (MCMC) methods, have proven, in spite of their computationally intensive nature, increasingly popular within the geophysical community, not only because of their versatility, but most importantly because of their ability to provide quantitative measures of model resolution, uncertainty and non-uniqueness [e.g. *Mosegaard & Sambridge,* 2002; *Sambridge & Mosegaard,* 2002 and references therein].

However, rather than invert for seismic wave speeds we propose to invert directly for the fundamental parameters of interest, namely mantle composition and thermal state. Indeed, with the present level of completion of the mineral physics database, enabling quantitative inferences to be made, a series of past studies using a variety of techniques have focused on the problem of constraining mantle chemistry and thermal state using geophysical data [e.g. *Deschamps & Trampert,* 2003; *Perry et al.,* 2003; *Trampert et al.,* 2004; *Shapiro & Ritzwoller,* 2004; *Kuskov et al.,* 2006, 2011; *Cammarano et al.,* 2009; *Khan et al.,* 2009, 2010; *Cobden et al.,* 2008; *Ritsema et al.,* 2009]. Our approach, which has been detailed previously [e.g. *Khan et al.,* 2007], makes use of a self-consistent thermodynamic methodology [*Connolly,* 2005] to systematically compute phase equilibria, seismic wave speeds and density that depend only on composition, pressure and temperature.

It is the purpose of the present study to employ the Metropolis algorithm (a type of MCMC method) to estimate thermo-chemical, physical and anisotropic structure beneath the North American continent and adjacent easternmost part of the Pacific Ocean using the global surface-wave phase-velocity maps of *Visser et al.* [2008a], which consist of

Rayleigh and Love-wave phase velocities of fundamental modes and overtones including uncertainties (see next section for further discussion). The use of phase velocity maps in place of the original phase velocity measurements, from which the former are derived, should be considered a simplifying assumption for the purpose of rendering the current study tractable. We pick the North American continent as it has been studied extensively seismically [e.g. *Grand,* 1994; *Van der Lee & Nolet,* 1997b; *Goes & Van der Lee,* 2002; *Godey et al.,* 2004; *Van der Lee & Frederiksen,* 2005; *Marone et al.,* 2007; *Nettles & Dziewonski,* 2008; *Sigloch et al.,* 2008; *Tian et al.,* 2009; *Yuan & Romanowicz,* 2010; *Yuan et al.,* 2011], providing a number of models with which to compare.

The immediate benefit of a wedlock between stochastic inversion and thermodynamic modelling as envisioned here, include: 1) inversion of seismic data directly for thermo-chemical structure, 2) quantitative assessment of model parameter uncertainties, resolution and non-uniqueness, 3) no potential bias through particular choice of initial/reference model nor damping parameter/regularization scheme and 4) depending on the specific set of seismic data considered, which typically are either sensitive to P or S-wave speed, we simultaneously constrain P-, S-wave speed and density. In particular point 1) above is all-important for unraveling the underlying nature of the processes that produce the observed variations in seismic wave speeds seen in tomography images, inasmuch as it enables us to distinguish between the relative contributions of composition and temperature, which as yet are not fully understood.

A prominent asset of seismic tomography studies is a global or regional image of seismic wave velocity. However, as it is not feasible to display images of all models sampled here, we revert to the idea initiated by *Koren et al.* [1991] and further exemplified by *Mosegaard*

& Tarantola [1995] and *Tarantola* [2005] of employing the movie strategy, which entails extracting and displaying random samples from the prior and posterior probability distributions. It is the contention here that the standard practice of looking at a single tomographic image should be abandoned in favour of an approach where several images are analyzed or interpreted concomitantly for current geoscientific implications of interest. This is a consequence of the fact that all models to be shown here are models with a high likelihood that fit data within uncertainties, but are likely to differ in terms of geodynamical implications. As a specific illustration of this we will compare observed geoid anomalies with those computed from the density (anomaly) maps obtained here.

Finally, we would like to dedicate this paper to the founding father of the probabilistic inference approach to inverse problems, the late Albert Tarantola, whose ideas on sampling-based methods for searching high-dimensional model parameter spaces and probabilisitic treatment of inverse problems in general are at the heart of the present study (for an excellent summary account of Albert Tarantola's work we refer the reader to *Mosegaard* [2011]).

2. Surface-wave Dispersion Data

As data we consider the isotropic part (azimuthally averaged) of the global azimuthal anisotropic phase-velocity maps of fundamental and higher-mode Love (to 5th order) and Rayleigh (to 6th order) waves of *Visser et al.* [2008a], with a lateral resolution of 5°×5°. The maps were obtained through an initial linear inversion of the global phase-delay database of *Visser et al.* We follow the approach of *Shapiro & Ritzwoller* [2004] and *Visser et al.* [2008b] and extract from the global surface-wave phase-velocity maps, on a 5°×5°-grid, dispersion curves (at the center of each pixel) for an area covering the North

American continent and surrounding, mostly Pacific, ocean (see figure 1). For each pixel we thus have 13 dispersion curves consisting of a total of 149 distinct Love and Rayleigh-wave phase-velocities as a function of frequency that we invert jointly for radial profiles of composition, temperature and anisotropic structure underneath each 5°×5° pixel. We limit ourselves to regional scale because of the high computational load inherent in the use of MCMC methods, which at present prohibit a global-scale study at the above resolution.

We are aware of the limitations imposed on this study through our choice of surface-wave phase-velocity maps rather than the original phase-velocity measurements ("tomographic data") from which the former are constructed. The range of models that will be mapped out here will to a large extent be controlled by the phase-velocity maps (data) and their uncertainties, which, *a priori* are not necessarily representative of the uncertainties inherent in inversion of the "tomographic data". However, we would like to note that the phase-velocity maps of *Visser et al.* [2008a] were contructed from phase-velocity measurements obtained using a model space search technique, which, as deemed by *Visser et al.* [2008a], provides consistent uncertainties on phase-velocity meaurements as well as on phase-velocity maps. Thus, although the present inversion is not a tomographic inversion in *sensu stricto*, it is an inversion for a set of (local) radial profiles of thermo-chemical and physical structure, which, when pieced together, result in a range of tomographic images that are consistent with data, i.e. *Visser et al.*'s phase-velocity maps and uncertainties.

Although the inversion of maps in itself is undesirable, adherence to Monte Carlo methods for inverting data strongly limit the amount of unknowns one can invert for. Indeed, with present computational resources available sampling-based strategies only allow for low-resolution global seismic tomography models [e.g. *Mosca,* 2010; *Khan et al.,* 2010] or models of limited geographical extent [e.g. *Bodin et al.,* 2009] and/or other modelling simplifications [e.g. *Shapiro et al.,* 2002]. However, the present approach should nonetheless be considered a step toward future full resolution of the seismic tomography problem using non-linear strategies.

With regard to sensitivity of these waves they sense well into the upper part of the lower mantle (from hereon simply lower mantle) to a depth of ∼1300 km. Moreover, while fundamental-mode surface-waves are predominantly sensitive to horizontally and vertically polarized S-wave velocity, the relative sensitivity of higher modes to compressional velocity (for Rayleigh-waves) and density grows with increasing overtone number [see *Anderson & Dziewonski,* 1982]. This difference in sensitivity of individual surface-wave modes allows us to simultaneously determine both thermal and compositional structure. Examples of dispersion curves will be shown later (see section 4.2, figure 2).

3. Parameterization and Forward Problem

Lateral variations in properties are defined over the grid shown in figure 1, with values defined at the center of each pixel, while radial variation is described using a number of layers, whose number varies depending on the particular property (this will be discussed further in section 4.1). Vertical layers beneath each pixel are described using the following set of parameters: 1) composition c, 2) temperature T, 3) anisotropy parameters ξ, ϕ and η (to be defined below), 4) seismic wave attenuation Q and 5) layer thicknesses. All the parameters are implictly assumed to be functions of radius.

In order to compute isotropic physical properties (V_s, V_p, ρ) beneath each pixel of our lateral grid, given the fundamental parameters c and T, we employ a self-consistent thermodynamic method based on Gibbs free energy minimisation. We assume mantle mineralogy to be dictated by thermodynamic equilibrium and predict mineralogy as a function of composition, pressure, and temperature by Gibbs energy minimization using the method of Connolly (2005). For this purpose we adopt the thermodynamic formalism of Stixrude & Lithgow-Bertelloni (2005) as parameterized by Xu et al. (2008) for mantle minerals in the model chemical system Na_2O-CaO-FeO-MgO-Al_2O_3-SiO_2 (abbreviated NCFMAS). The Gibbs energy minimization procedure yields the amounts, compositions, and physical properties, including elastic moduli, of the stable minerals in the model chemical system. Aggregate elastic moduli are estimated from this information by Voigt-Reuss-Hill averaging.

Although we cannot really constrain attenuation structure Q with the surface-wave data here, we follow our previous approach [*Khan et al.,* 2010] and use the following expression to estimate Q and thus anelastic contributions to the isotropic (anharmonic) P and S-wave velocities obtained using Gibbs free energy minimisation [e.g. *Anderson,* 1989; *Jackson,* 2000]

$$Q_\mu = Q_0 \exp\left[\frac{\alpha(E_a + pV_a)}{\mathsf{R}T}\right] \tag{1}$$

where Q_0 is a constant, E_a activation energy, V_a activation volume, p pressure, T temperature, R the gas constant and α an exponent, which has been determined experimentally to be between 0.15-0.25 [*Jackson et al.,* 2002].

In relation to anisotropy, we follow the standard assumption adopted in most surface-wave tomography studies [e.g. *Panning & Romanowicz,* 2006; *Kustowski et al.,* 2008; *Nettles & Dziewonski,* 2008] and assume transverse anisotropy (symmetry axis in vertical direction). In order to compute Love and Rayleigh-wave dispersion curves, we first need to compute $V_{sv}, V_{sh}, V_{pv}, V_{ph}$, which are velocities of vertically (v) and horizontally (h) polarized S-waves and vertically and horizontally propagating P-waves, respectively. Following the approach of previous surface-wave studies [e.g. *Panning & Romanowicz,* 2006] and assuming that anisotropy is small ($\eta \sim 1$), anisotropic Voigt-averaged velocities can be computed from isotropic P and S-wave velocities [*Babuska & Cara,* 1991] using

$$V_{sv} = \sqrt{\frac{3V_s^2}{2\xi}},\ V_{sh} = \sqrt{\frac{3V_s^2}{2}},\ V_{ph} = \sqrt{\frac{5V_p^2}{4\phi}},\ V_{pv} = \sqrt{\frac{5V_p^2}{4}} \tag{2}$$

where ξ and ϕ quantify S and P-wave anisotropy, respectively, and η describes the dependence of velocity on the incidence angle of a propagating wave (Dziewonski & Anderson, 1981). In summary, given values of the set of parameters $\{V_p, V_s, \xi, \phi, \eta\}$, anisotropic velocities are easily computed from expressions (2 & 3). Finally, we would like to recall that because we are considering surface-wave overtone data, sensitivity encompasses P-wave velocity and density in addition to S-wave velocity. Note also the complete absence of preassigned scaling factors between the various parameters.

4. Inverse Problem

As in our previous work we employ the probabilistic approach of *Tarantola & Valette* [1982] to solve the non-linear inverse problem. Within the Bayesian framework, the solution to the general inverse problem $\mathbf{d} = \mathrm{g}(\mathbf{m})$ ($\mathbf{d}$ is data vector and g a typically non-linear operator that maps a model parameter vector $\mathbf{m}$ into data), is given by [e.g. *Tarantola & Valette,* 1982; *Mosegaard & Tarantola,* 1995]

$$\sigma(\mathbf{m}) = kf(\mathbf{m})\mathcal{L}(\mathbf{m}), \tag{3}$$

where k is a normalization constant, $f(\mathbf{m})$ is the prior probability distribution on model parameters, i.e. information about model parameters obtained independently of data,

$\mathcal{L}(\mathbf{m})$ is the likelihood function, which in probabilistic terms can be interpreted as a measure of misfit between the observations and the predictions from model $\mathbf{m}$, and $\sigma(\mathbf{m})$ is the posterior model parameter distribution containing the solution to the inverse problem. The particular form of $\mathcal{L}(\mathbf{m})$ is determined by the observations, their uncertainties and how these are employed to model data noise.

In the current interdisciplinary context (see e.g. *Bosch* [1999] and *Khan et al.* [2007] for details) we are dealing with several different model parameters describing the system at various levels (physical - V_p & V_s, mineralogical/petrological - M (equilibrium modal mineralogy) and thermo-chemical - c & T). For present purposes we define three sets of parameters termed primary, secondary and tertiary model parameters (with the present general formulation this is easily generalised to any number of parameter vectors), which are given by $\mathbf{m}_p = \{c, T, \xi, \phi, \eta, Q\}$, $\mathbf{m}_s = \{M, V_p, V_s, \rho\}$ and $\mathbf{m}_t = \{V_{pv}, V_{ph}, V_{sv}, V_{sh}\}$, respectively. In the joint model parameter space $\mathcal{M} = \mathcal{M}_p \times \mathcal{M}_s \times \mathcal{M}_t$ we can define the joint model parameter vector $\mathbf{m} = \{\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t\}$, where $\mathcal{M}_p$, $\mathcal{M}_s$ and $\mathcal{M}_t$ are primary, secondary and tertiary model parameter spaces, respectively. Extending eq. (3) to the joint description, we obtain

$$\sigma(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t) = kf(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t)\mathcal{L}(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t) \tag{4}$$

Note that since secondary parameters are functions of the primary model parameters and tertiary parameters are functions of both primary and secondary model parameters, the joint prior probability distribution and likelihood function can be suitably decomposed and dealt with separately by the rule of conditional probabilities. This is probably also warranted from the point of view that for most real problems the complexity of the joint prior density function is such that it generally would be difficult to formulate. Decomposing the joint prior then, we have

$$f(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t) = f_t(\mathbf{m}_t|\mathbf{m}_s, \mathbf{m}_p) f_s(\mathbf{m}_s|\mathbf{m}_p) f_p(\mathbf{m}_p) \tag{5}$$

where $f_p(\mathbf{m}_p)$ is a marginal probability density function (*pdf*) describing prior information on primary parameters, $f_s(\mathbf{m}_s|\mathbf{m}_p)$ and $f_t(\mathbf{m}_t|\mathbf{m}_s, \mathbf{m}_p)$ are conditional *pdf*s containing information about secondary and tertiary parameters and their dependence on primary and secondary parameters, respectively.

Let us assume that we have performed a number (k) of different geophysical experiments to study the system, structure or region of interest, with each of these giving rise to a set of observations $\mathbf{d}_1, \ldots, \mathbf{d}_k$ belonging to the joint data parameter space $\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_k$. Since, in general, observational uncertainties among different geophysical methods are independent we can write the joint likelihood function over the joint model space as

$$\mathcal{L}(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t) = \prod_{j=1,k} \mathcal{L}_j(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t) \tag{6}$$

where the $\mathcal{L}_j$ are independent likelihood functions appropriate for each of the geophysical methods employed.

We can now summarize the posterior *pdf* in the joint model space by combining eqs. 4, 5 and 6

$$\sigma(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t) = c f_t(\mathbf{m}_t|\mathbf{m}_s, \mathbf{m}_p) f_s(\mathbf{m}_s|\mathbf{m}_p) f_p(\mathbf{m}_p) \times \prod_{j=1,k} \mathcal{L}_j(\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t) \tag{7}$$

To sample the posterior distribution (eq. (7)) in the joint model space we employ a Metropolis algorithm (a Markov chain Monte Carlo method). Although this algorithm is based on a random sampling of the model space, only models that result in a good data fit and are consistent with prior information are frequently sampled.

We employ the Metropolis-Hastings algorithm [*Metropolis et al.,* 1953; *Hastings,* 1970], which can be summarized in the following rules, to sample the joint posterior distribution (eq. 7) [e.g. *Mosegaard & Tarantola,* 1995]

1. Consider $\mathbf{m} = \{\mathbf{m}_p, \mathbf{m}_s, \mathbf{m}_t\}$ to be some current joint model in the Markov chain and randomly modify it to some candidate joint model $\mathbf{m}' = \{\mathbf{m}'_p, \mathbf{m}'_s, \mathbf{m}'_t\}$, where the candidate model is drawn from the prior using a proposal distribution.

2. Acceptance of $\mathbf{m}'$ is governed by the probability

$$\mathcal{P} = \min\left[1, \frac{\mathcal{L}(\mathbf{m}')}{\mathcal{L}(\mathbf{m})}\right] \tag{8}$$

3. If $\mathbf{m}'$ is accepted then it becomes the current joint model, otherwise the current state remains $\mathbf{m}$.

4. Return to point 1 above and reiterate.

This algorithm is capable of sampling the model space with a sampling density proportional to the given probability density without excessively sampling low-probability areas of the model space. This is particularly important when we consider high-dimensional model spaces in which a large proportion of the volume may have near-zero probability density.

Single realizations such as the mean, median or maximum likelihood model as a means of studying the solution to the general inverse problem are generally inadequate descriptors and are best replaced by looking at samples from the posterior *pdf*, for example. Another possibility is to calculate resolution measures, which are easily evaluated from [e.g. *Mosegaard,* 1998]

$$\mathcal{R}(\Omega, h) = \int_{\Omega} h(\mathbf{m})\sigma(\mathbf{m})d\mathbf{m} \approx \frac{1}{N} \sum_{\{n|\mathbf{m}_n \in \Omega\}} h(\mathbf{m}_n) \tag{9}$$

where $h(\mathbf{m})$ is a given function of the model parameters $\mathbf{m}$, Ω is an event or subset of the model space containing the models of current interest and N is the total number of samples taken from Ω. The point to note is that within the Bayesian framework we can pose any question of the sort: what is the probability for observing a given model feature ? There are generally no ill-posed questions, but just questions that have a probabilistic answer given by eq. 9.

An alternative means to quantitatively analyze the posterior *pdf* involves the standard Bayesian approach to hypothesis testing in the form of the Bayes factor, which summarizes the evidence for one hypothesis over another. The Bayes factor $\mathcal{B}_{ij}$ for hypothesis (or model) A_i against hypothesis A_j, given data and prior information, is defined as the ratio of posterior to prior odds, or equivalently, as the ratio of likelihoods, signaling the effect of data on gauging relative prior beliefs into relative posterior beliefs [e.g. *Bernardo & Smith,* 1994; *Khan et al.,* 2004]

$$\mathcal{B}_{ij} = \frac{\mathcal{L}(\mathsf{A}_i)}{\mathcal{L}(\mathsf{A}_j)} \tag{10}$$

In the following we will briefly enumerate prior information and likelihood function.

4.1. Prior Model Parameter Information

The parameters detailed below define the model parameters that describe the radial parameterization beneath the center of each pixel.

4.1.1. Crustal structure

Crustal structure is described by the physical parameters: ρ, V_p, V_s and depth to Moho. For each pixel an average four-layer crustal profile was extracted as starting model from the global crustal model CRUST2.0 (http://mahi.ucsd.edu/Gabi/rem.html). In each of the four layers ρ, V_p and V_s are variable within upper and lower bounds, where the former are ρ=1.5 g/cm^3, V_p=2.5 km/s and V_s=1.5 km/s and the latter correspond to the thermodynamically-determined parameter at the first depth node in the mantle, respectively. We assume additionally ρ, V_p and V_s to be non-decreasing as a function of depth, while Moho depth d_{cr} varies to within ±5 km (oceanic regions) and ±20 km (continental regions) of the crustal thickness of each pixel extracted from CRUST2.0. [13 parameters].

4.1.2. Temperature

Temperature T is assumed uniformly distributed with no lower or upper bounds, with the constraint that it be non-decreasing as a function of depth. Surface temperature is held constant at 0°C. Temperatures are specified in 25 uniform layers at intervals of 50 km in the depth range 0-700 km, and increasing to 100 km in the range from 700-2886 km. [25 parameters].

4.1.3. Compositional layer thickness

We model crust and mantle as consisting of three layers corresponding to a compositional division into crust, upper and lower mantle, respectively. Depths of these layers are located at the physically-determined Moho depth (see section 4.1.1 above), 660 and 2900 km depth, respectively. For purposes of simplification, only the '660-km' discontinuity is considered variable. Earth's surface and core-mantle-boundary (CMB) are fixed in accordance with values taken from PREM at 0 km and 2891 km depth, respectively. [1 parameter].

4.1.4. Silicate mantle composition

Mantle compositions were explored within the NCFMAS system, a model that accounts for more than 98% of mass of the mantle [*Irifune,* 1994]. Mantle compositions c adopted here are assumed to be uniformly distributed in all four mantle layers within the bounds given in table **??**. The bounds chosen for the upper mantle are such that our compositions are in agreement with the range of compositions of mantle peridotites derived from several geochemical studies (see table 2 of *Lyubetskaya & Korenaga* [2007]). [10 parameters].

4.1.5. Attenuation and anelasticity

Because of the few constraints that surface-wave tomographic studies are able to provide on Q, we follow our previous approach (standard procedure in surface-wave tomography is to fix attenuation structure) and compute Q according to eq. (1) by fixing E_a=5·10^2 kJ/mol, V_a=2.5·10^{-3} cm^3/mol, Q_0=1 and α=0.2 [*Sobolev et al.,* 1996; *Jackson et al.,* 2002; *Cobden et al.,* 2008], while pressure p derives from the thermodynamic method and T is a model parameter. This approach ensures variability in Q without leading to large perturbations in anelastic velocities.

4.1.6. Anisotropy

Anisotropy parameters ξ, ϕ and η are assumed to be uniformly distributed within the bounds specified in table **??**, which bracket most of the range of values of recent anisotropy estimates obtained from recent surface-wave tomography studies [e.g. *Panning & Romanowicz,* 2006; *Kustowski et al.,* 2008; *Visser et al.,* 2008b]. We also assume that anisotropy parameters are constant within the following fixed layer boundaries: 0-25, 26-50, 50-100, 100-150, 150-250, 250-350, 350-450, 450-600, 600-800, 800-1000, 1000-1200, 1200-1400, 1400-1600, 1600-1800, 1800-2000 km depth. [45 parameters].

4.1.7. Isotropic and anisotropic physical properties

No constraints apply to any of these model parameters. All physical properties, including modal equilibrium mineralogy, are computed at 65 radial nodes (layer thickness is 10 km in the depth range 0-100km; 30km in the depth ranges 100-370, 420-540 and 570-630km; 5km in the depth ranges 370-420, 420-540, 630-700km; 100km at depths of 700km and more).

4.2. Sampling the Posterior Distribution

Summarizing the model parameter setup, each pixel of our model is described by 94 parameters that have to be determined. Once these parameters have been assigned values, we compute modal mineralogy and physical properties in the crust and mantle as a function of pressure, temperature and composition from which Rayleigh and Love-wave dispersion curves are subsequently estimated.

In line with our previous study, we assume the L_2-norm for modelling data misfit, which results in a likelihood function of the form

$$\begin{aligned}\mathcal{L}(\mathbf{m}) \propto \exp(&-\sum_{\text{mode}}\sum_{\text{frequency}} \frac{[d_{obs}^R - d_{cal}^R(\mathbf{m})]^2}{2\sigma_R^2} \\ &-\sum_{\text{mode}}\sum_{\text{frequency}} \frac{[d_{obs}^L - d_{cal}^L(\mathbf{m})]^2}{2\sigma_L^2}) \qquad (11)\end{aligned}$$

where d_{obs} and $d_{cal}(\mathbf{m})$ denote observed and calculated data, respectively, superscripts indicate surface-wave type (R - Rayleigh, L - Love), and $\sigma_{R,L}$ uncertainty on either of these. With the L_2-norm, we implicitly assume that data noise can be modeled using a gaussian distribution and that observational uncertainties and calculation errors between Rayleigh and Love-waves are independent.

Convergence of the algorithm is generally reached after about 10000 iterations and only after this stage is reached, i.e. when sampled models fit the observations, did we commence retaining samples from the posterior *pdf*. To ensure convergence of the MCMC algorithm in practice, we verified that the time series of all output parameters from the algorithm were stationary throughout the entire sampling stage. In order to ensure adequate sampling of the model space we sampled until no significant changes to the characteristics of the posterior *pdf* were observed, in addition to recommencing the algorithm at a number of different places in the model space. To further ensure near-independent samples an 'elapse time' (number of iterations) between retention of samples was implemented, which was found to be 100 by analysing the autocorrelation function of the fluctuations of the likelihood function. We sampled in all 1 million models from which $\sim$10000 were retained for analysis. The overall acceptance rate ranged from 35 to 40%.

The posterior probabilities so calculated are mathematical entities based on the quantitative information used as input in the inversion. Stated differently, the probabilities are based entirely on (1) data and their uncertainties, (2) prior information as quantified here and (3) the physical law connecting data and unknown model parameters. In relation to point (2), we have to be aware of the limitations imposed by our choice of model parameterization, as any inverse problem faces a trade-off between model parameter resolution and uncertainty. No exhaustive examination of the effect of different parameterizations were attempted, except for the investigation of two different compositional parameterizations - a seven-layer model in addition to the one described here. Apart from small differences in composition, all other inverted parameters were found to agree remarkably within uncertainties. In summary, there is no unique way of parameterizing a model system and the results simply reflect the particular parameterization chosen. Also, with regard to model parameter uncertainties obtained here, we would like to note that we have not considered uncertainties related to mineral physics parameters and thermodynamic formulation as a result of which the model parameter uncertainties in reality are larger than they appear here.

Given that the posterior *pdf* is defined as the conjunction of the prior with the likelihood function (eq. 7), which are generally complex functions, the posterior *pdf* will also be complex. Typically, it will be multimodal, i.e. there are many possible solutions (secondary extrema) in addition to the most probable solution (global extremum). In order to summarize information from such a complex *pdf* standard resolution measures such as means and covariances are inadequate and we have to resort to a more general approach, which will typically depend on the questions that we are trying to address.

One-dimensional (1D) marginals are appropriate for obtaining information on single parameters and their uncertainties. Information concerning other parameters, however, is absent. For this, 2D or 3D marginal *pdf*'s are required, since these reveal the correlation that exists among several parameters. However, of most importance here is the use of the movie strategy of *Mosegaard & Tarantola* [1995] and *Tarantola* [2005], which is ideally suited for analysis of the seismic tomography problem. The main point is to display a collection of models taken randomly from the prior and posterior *pdf*s. This collection of prior and posterior models provide us with an approximate idea of the prior information used, but also, by comparison of the two, the information contained in the data. General features characteristic of the models, like those that are well-resolved, will tend to be recurring in the posterior imagines, whereas those that are ill-resolved appear much more scattered and resemble prior images. The point being that data-related structural patterns are easily separable from those that appear randomly in a non-coherent and non-recurring fashion. Although posterior models can differ significantly, they are nonetheless models with high likelihood values that predict observed data within uncertainties (see figure 2 for an example of datafit at two distinct locations). It is also the reason why the mean of such a collection of models, which itself is necessarily smooth, is *a posteriori* very unlikely as it most probably cannot fit data.

5. Results - Prior and posterior movies

5.1. Mantle temperature and composition

We start this section by showing samples from the prior *pdf*. However, rather than considering the full joint prior, we limit ourselves in this section to the fundamental parameters c (here in the form of Mg#) and T. The prior models displayed in figure 3 (plots 1-24) are thus samples from $f_p(\mathbf{m}_p)$ and constitute 6 thermal models picked at random at different depths in the mantle (100, 300, 500 and 1000 km depth). In addition to showing samples from the prior *pdf*, we are also showing 1D marginal *pdf*s for a given pixel at the depths indicated above. These sets of figures depict the sort of prior information that is employed here and at the same time allows us to verify that the prior *pdf* has been sampled in accordance with expectations. Note the large model variability at all depths in accordance with the few prior constraints imposed on parameters T and c (for prior information concerning the latter parameter see figure 5). Particular characteristics of these prior maps are the non-coherent and non-recurring small-scale features that vary randomly across the images as expected.

For comparison, figure 4 shows 6 thermal models picked randomly from the posterior *pdf* (plots 1-24) as well as 1D marginal *pdf*s at the same depths as above. The juxtaposition reveals the following: 1) the randomly varying pattern seen in the prior T maps has instead been replaced by coherent and repetitive structures and 2) comparison of prior and posterior 1D marginal *pdf*s shows that the latter have decreased significantly in width. For example, at 1000 km depth prior thermal models range from ∼1500 to ∼2300 °C, whereas posterior models are confined to the range ∼1575-1725 °C. Both points argue for well-resolved mantle temperatures.

Returning to the upper mantle, we clearly observe thermal variations that correlate well with major geological surface features. In particular, the thermal models of the North American continent reveal a strong continental contrast, dividing the tectonically active western region from the tectonically stable eastern region. At 100 km depth, we find the coldest parts to extend across the eastern old continental region, while the hot parts comprise the western margin and Pacific Ocean (East Pacific Rise). Cold anomalies are also associated with the older (150-160 Ma) parts of the atlantic lithosphere off the southeast coasts of North America, while some of the coldest anomalies are observed over the North American craton. At 300 km depth, the picture described above has changed slightly with the coldest anomalies centered somewhat more closely around the North American craton and the East Pacific Rise remaining the hottest anomaly. Also, the lithosphere beneath the southern part of NA as well as the old atlantic lithosphere to the southeast have increased significantly in temperature.

Thermal anomaly maps of the upper mantle of NA have also been obtained by *Godey et al.* [2004] and *Goes & Van der Lee* [2002] from inversion of a shear-wave velocity and density model of North America. In spite of data and modelling differences, their thermal maps at 100 and 250 km depth are found to qualitatively agree with ours, in particular as concerns the thermal division of the NA continent with surface tectonic provinces. Similar observations of the thermal structure of the lithosphere and upper mantle have been made in a number of gephysical studies [e.g. *Jaupart & Mareshal,* 1999; *Röhm et al.,* 2002; *Shapiro & Ritzwoller,* 2004; *Ritzwoller et al.,* 2004; *McKenzie et al.,* 2005; *Goes et al.,* 2005; *Faul & Jackson,* 2005; *Priestley & McKenzie,* 2006; *Kuskov et al.,* 2006, 2011; *Artemieva,* 2006, 2009; *Afonso et al.,* 2008; *Simmons et al.,* 2009].

In the TZ, the thermal anomalies have become reversed with the NA craton now being hotter than surrounding mantle, whereas the East Pacific Rise and western margin appear as the coldest parts. In the lower mantle at 1000 km depth strong thermal anomalies are less prevalent (amplitudes of observed thermal anomalies decrease from ∼300 °C at 100 km depth to ∼100 °C at 1000 km depth), in line with observations from seismic tomography studies (see section 5.2 for more discussion) that show the largest lateral variations in properties to be concentrated in the upper mantle.

Turning to the compositional results, we find that comparison of prior (figure 5) and posterior (figure 6) models of Mg#, and particularly the 1D marginal *pdf*s, generally reveal the same behaviour discussed above for mantle T. In addition, the multimodel nature of the 1D compositional posterior *pdf* in the upper mantle is also discernible. General features of the Mg# maps seem to follow the tectonic division observed in the case of the thermal maps. In particular, the old stable continental region is found to be depleted in FeO (high Mg#), whereas younger continental areas and Pacific Ocean are observed to be enriched in FeO. This pattern was also observed in our previous study [*Khan et al.,* 2010] and had been hypothesized by e.g. *Jordan* [1975, 1978] as a means of explaining the stability of continental roots. Further evidence for compositional variations of the continental lithosphere also come from a number of geochemical analyses of mantle xenoliths and cratonic peridotites [e.g. *Boyd,* 1989; *Rudnick et al.,* 1998; *Griffin et al.,* 1999; *Gaul* *et al.,* 2000; *Artemieva,* 2009; *Afonso et al.,* 2010].

With the present simplified mantle compositional parameterization, no compositional variations in the TZ are observed. In the lower mantle, we find, as in the case of mantle temperatures, composition to vary relatively little laterally. Mg# anomalies are seen to vary between 0.9 and 0.92, implying general Fe-depletion. Overall, there is a noticeable tendency for higher Mg# in the lower mantle relative to the upper mantle, suggesting compositional differences between the two as observed in our previous study [*Khan et al.,* 2010].

With regard to possible trade-offs between c and T, figure 7 shows plots of 2D marginal posterior *pdf*'s for the latter parameters in the upper and lower mantle, which reveal no noticeable correlation. However, given different depth parameterizations for the two parameters, the correlations here represent a temperature average over the layers that bracket the two compositional layers.

Finally, we would like to note that our thermo-chemical results presented here are based on the assumption of thermodynamic equilibrium. *Xu et al.* [2008] have discussed a possible alternative, the mechanical mixture model. From the point of view of geophysics there is no argument for or against either model. However, from a petrological viewpoint it can be argued that while the mechanical mixture model plausibly depicts the influence of chemical segregation on the equilibrium model, it cannot be claimed to be a more realistic endmember for the earth's mantle because it is inconsistent with mid-ocean ridge volcanism (for more discussion see *Khan et al.,* 2009).

5.2. Isotropic shear-wave velocity structure

Posterior movies (for brevity we omit prior movie) of mantle shear-wave velocity structure are shown in figure 8. For comparison, we have also included two regional seismic tomography models: the anisotropic shear-wave velocity models of North America by *Nettles & Dziewonski* [2008] and *Yuan et al.*. [2011]. The model by *Nettles & Dziewonski* (henceforth ND08) is based on a large number of global and regional measurements of the dispersion of fundamental-mode surface-waves. The model of *Yuan et al.* (YU11) was obtained from inversion of long-period fundamental-mode and overtone surface-waveforms. Model ND08 is restricted to the upper mantle, while model YU11 has some sensitivity in the upper TZ. Thus no comparison with previous models is made in the lower mantle.

Structural features in the posterior movies (figure 8, plots 1-24), repeat across the different images, implying a well-resolved S-wave velocity structure, particularly in the upper mantle and TZ. The main continental division so clearly apparent in the posterior thermal and compositional movies, is closely followed here at 100 km depth. The old eastern parts of the NA continent that were found to be cold and Fe-depleted appear as regions of fast S-wave velocity, while the younger and hotter, Fe-enriched regions (the western margin and Pacific Ocean) are observed to be relatively slow. The division between the tectonically active and tectonically stable parts of NA is seen to follow the Rocky Mountain front as first outlined by *Grand* [1994]. These features are also clearly apparent at the same depth in models ND08 and YU11 (see figure 8) as well as in many regional seismic surface-wave and travel-time tomography models of NA [e.g. *Van der Lee & Nolet,* 1997b; *Frederiksen et al.,* 2001; *Li et al.,* 2002; *Godey et al.,* 2004; *Van der Lee & Frederiksen,* 2005; *Sigloch et al.,* 2008] and many other fundamental-mode global seismic surface-wave tomography models [e.g. *Ekström et al.,* 1997; *Trampert & Woodhouse,* 2001; *Shapiro & Ritzwoller,* 2002; *Lebedev & Van der Hilst,* 2008].

At 300 km depth the ocean-continent contrast is still discernible (figure 8, plots 7-12), with the oldest stable part of the continent appearing to be somewhat faster than surrounding mantle, while the region centered on the NA craton is characterized by being distinctly faster. These features can also be perceived in ND08 and YU11. Fast S-wave velocity anomalies around the NA craton are found to persist down to ~300 km depth in all models shown here. This is further supported by the shear-wave velocity profiles shown in figure 9. From 300 km depth and deeper differences in S-wave velocities beneath the various tectonic settings disappear as a result of which this part of the upper mantle appears more homogeneous.

In the TZ (figure 8, plots 13-18) we observe a much smoother picture with peak-to-peak velocity variations being ~0.1 km/s, in comparison to the upper mantle where variations ranged from 0.2 to 0.3 km/s. This is also evident from figure 9, which indicates that the TZ is less heterogeneous than the upper mantle. Velocities are also found (figure 8, plots 13-18) to have reversed with the older continental parts now being characterized by slower velocities relative to younger areas. This reversal is also observed in model YU11 and features in several seismic tomography studies [e.g. *Ritsema et al.,* 2004; *Panning & Romanowicz,* 2006; *Visser et al.,* 2008a; *Kustowski et al.,* 2008; *Lebedev & Van der Hilst,* 2008].

In the lower mantle at 1000 km depth (figure 8, plots 19-24), lateral velocity variations are now <0.1 km/s, implying a much more homogeneous lower mantle relative to the upper mantle. Note also the general overlap of S-wave velocity profiles in figure 9 for depths

$>$ 700 km. This observation is also common among global S-wave tomography models, which are typically characterized by an absence of strong heterogeneities at long wavelengths below 650 km depth, resulting in a spectrally white distribution of mid-mantle isotropic shear-wave velocity anomalies [e.g. *Ritsema et al.,* 2004; *Panning & Romanowicz,* 2006; *Kustowski et al.,* 2008].

Comparing our models with previous regional seismic tomography models shows a high degree of correlation in the upper mantle, amplitudes aside. This level of agreement is indeed reassuring given the fundamentally different approach employed here to inverting seismic data. The problem with amplitudes of retrieved velocity anomalies is a well-known and persistent feature of all seismic tomograhy models and is mostly related to the use of different regularization schemes/damping parameters and choice of particular 1D reference model. In addition to this, discrepancies between tomography models are known to arise around the TZ, where the correlation coefficient between various global tomography models is found to decrease strongly [e.g. *Kustowski et al.,* 2008]. On the face of it such discrepancies are also palpable here at e.g. 500 km depth (compare plots 13-18 with model YU11. Yuan et al. find a large low-velocity province beneath central North America that is not imaged to the same extent in our maps, which could be interpreted as an inconsistency of our data set and the measurements of Yuan et al. However, a closer look at sampled velocities at 500 km depth beneath several pixels in the form of 1D marginal *pdf*'s (figure 10) reveals a considerable degree of consistency inasmuch as model YU11 generally lies within the range of presently sampled shear-wave velocities. In addition, if uncertainty estimates for model YU11 could be taken into account any remaining discrepancies would most likely disappear (pers. comm. H. Yuan, 2011). Note that the marginal 1D *pdf*'s appear gaussian, whereby the mean shear-wave velocity ($\bar{V}_S$) and its standard deviation (σ_{V_S}) can be computed. Maps of $\bar{V}_S$ and $\bar{V}_S \pm \sigma_{V_S}$ are shown as online supporting material figure S1.

We have so far abstained from discussing mantle attenuation structure because it is less well-constrained. The latter is a result of the large uncertainties that exist on the attenuation-related parameters Q_o, V_a, E_a and α that are employed here for calculating attenuation structure. However, we made several tests, where we varied the aforementioned parameters by as much as 10-20%, to verify that the results did not change.

Given the fundamental approach of inverting directly for c and T, we also constrain P-wave velocity and density. However, for reasons of brevity these results are not shown here, but can be found as online supporting material (see online supporting figures S2-S3).

5.3. Anisotropic shear-wave velocity structure

Prior and posterior anisotropic shear-wave velocity models are shown in figures 11 - 12, and as in the case of S-wave velocity structure, model features appear fairly robust across much of the mantle shown here. At 100 km depth (figure 12, plots 1-6), most of the Pacific Ocean and NA continent are characterized by positive anisotropic anomalies (ξ >1), in agreement with what is seen in the previous regional tomographic models ND08 and YU11. At a depth of 150 km (for brevity images at intermediate depths are not shown) these features persist across all models, except for the tectonically young areas centered on the west coast, for which $\xi < 1$ now. This pattern is reinforced at 200 km depth, with all parts of the western margin, the southeast and a large part of the Pacific Ocean having $\xi < 1$, i.e. $V_{sv} > V_{sh}$ as can be seen from figure 12 (plots 7-12) at 300 km depth. This largely tectonically-driven signal is seen to extend to 350 km depth here (not shown), with old stable continental areas characterized by $\xi > 1$, in contrast to younger regions where $\xi < 1$. Parts of these features recur to some extent in ND08 and YU11.

The upper mantle anisotropy structure retrieved here generally agrees with the global anisotropic tomography model of *Gung et al.* [2003], who found the East Pacific Rise, western margin and southeastern part of NA to be characterized by ξ <1, while for the stable continental regions ξ >1. In particular, they observed that at 300 km depth, the roots of most cratons were characterized by ξ >1, which extend down to 400 km depth, whereas for the East Pacific Rise ξ <1 down to 300 km depth.

As we cross into the TZ a general change in anisotropic signal is observed that mimicks the reversal in isotropic shear-wave velocity pattern observed in the maps shown earlier (see figure 8, plots 13-18). This change commences around 400 km depth and grows more coherent as we transcend deeper into the TZ. In particular, the area centered on the NA craton and eastern part of NA as well as East Pacific Rise are regions where negative ξ anomalies predominate, i.e. $V_{sv} > V_{sh}$. In the lower mantle another subtle change in anisotropy occurs and most of the lower mantle appears to be relatively homogeneous, characterized by predominantly positive anisotropy anomalies.

The question of anisotropy in and below the TZ has been studied for some time now [e.g. *Montagner & Kennett,* 1996; *Trampert & Van Heijst,* 2002; *Wookey et al.,* 2002; *Panning & Romanowicz,* 2006; *Visser et al.,* 2008b; *Kustowski et al.,* 2008], although little agreement has emerged. *Kustowski et al.* [2008], for example, correlated their whole-mantle anisotropic model S362WANI with the one derived by *Panning & Romanowicz* [2006] SAW642AN and found anisotropic variations only to be consistent in the upper-most (150 km) and lower-most mantle (2800 km). Our model comparisons here support the contention that anisotropic models generally only agree in the upper-most mantle.

To further investigate the robustness of the TZ and lower mantle anisotropy signal retrieved here (figure 12, plots 13-24), we analyzed the correlation between ξ and all other parameters that might potentially be interfering, such as c, T and V_s. However, no trade-offs were observed (not shown for brevity), as expected. Additionally, we looked at shear-wave anisotropy maps at 1800 km depth and found prior (figure 11, plots 25-30) and posterior (figure 12, plots 25-30) movies to be similar in character with small-scale features varying randomly across the maps, typical of all prior plots shown hitherto. Again, this follows our expectation, inasmuch as the surface-wave data only have sensitivity to $\sim$1300 km depth. This suggests that the structural patterns seen at 1000 km depth are data-related as these are easily separable from those that appear randomly in a non-coherent and non-recurring fashion, as pointed out previously.

Anisotropy plays an important role in seismic tomography, because of the potential constraints that it provides on mantle flow. Anisotropy is thought to be an indicator of present-day mantle strain field or past deformation frozen in the lithosphere [e.g. *Tanimoto & Anderson,* 1984; *Montagner & Tanimoto,* 1991; *Karato,* 1998; *Montagner,* 1998; *Becker et al.,* 2008; *Long & Becker,* 2010]. Changes in sign of anisotropy can thus be interpreted as indicating changes from horizontal to vertical flow under the assumption that anisotropy is the result of a preferred orientation of the crystal lattice of the anisotropic mantle minerals as these are subjected to strains due to mantle flow. With this in mind, our results suggest a prevailing horizontal shear flow in the asthenosphere beneath continents, while the reverse is the case beneath oceanic regions and likely also younger continental areas. Several changes in sign of anisotropy are observed here, that might be indicative of the presence of a number of distinct anisotropic layers for the lithosphere, asthenosphere, TZ and possibly lower mantle. A division of the upper mantle beneath NA into distinct anisotropic lithospheric and asthenospheric layers has been proposed ealier by *Gaherty* [2004], *Marone et al.* [2007], *Deschamps et al.* [2008] and *Yuan et al.* [2011], as well as other cratonic areas in general [*Debayle et al.,* 2005].

So far we have exclusively discussed shear-wave anisotropy, leaving the other anisotropy parameters, i.e. ϕ and η, aside. P-wave anisotropy has been studied by e.g. *Anderson & Dziewonski* [1982], *Boschi & Dziewonski* [2000] and *Beghein & Trampert* [2003], although as in the case of S-wave anisotropy in the TZ and lower mantle consensus is yet to emerge, which is due to the limited sensitivity of the surface-wave data to P-wave anisotropy. As a result global seismic tomography studies simply scale P to S-wave anisotropy [e.g. *Panning & Romanowicz,* 2006; *Kustowski et al.,* 2008; *Visser et al.,* 2008a,b]. Although we have taken a somewhat more lenient approach in that we also inverted for ϕ, we will not discuss the results in any detail, given that ϕ is less well-constrained. The same arguments apply to η. However, we did verify that our particular parameterization did not lead to perturbation of the radial shear-wave anisotropic signal found here.

6. Posterior filtering of tomographic models using geoid anomalies

Additional geophysical data can be employed as a tool to refine and narrow the collection of tomographic models. As auxiliary geophysical data we consider geoid anomalies, since these are directly related to the density structure.

A major difficulty with modeling geoid anomalies from a prescribed density distribution is to correctly estimate the dynamic contribution to the geoid anomalies [e.g., *Ricard et al.,* 1984; *Forte and Peltier,* 1987; *Hager & Richards,* 1989]. Furthermore, reconstructing geoid anomalies from regional models that are not developed in spherical harmonics, requires 1) prescribing an appropriate band-pass window and 2) accounting for lateral density distributions [*Kogan & McNutt,* 1993]. Here, we compute geoid anomalies δN for the six density models shown in the posterior movie (see online supporting material, figure S4) following the regional approach described in van *Gerven et al.* [2004],

$$\delta N(\theta,\phi) = \frac{3}{4\pi\rho_m}\int_{r_{\mathrm{CMB}}}^{R}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} K_g(\Delta,r)\delta\rho(r,\theta',\phi')\sin\theta' d\theta' d\phi' dr \qquad (12)$$

where ρ_m, R and r_{CMB} are Earth's mean density, surface and core radius, respectively and Δ angular distance between the point where the geoid is measured (θ,ϕ) and the location of the density anomaly (θ',ϕ'). The local geoid kernels $K_g(\Delta,r)$ describe the response of the geoid to a density anomaly located at the position (r,θ',ϕ') in the band-pass filter $l_1 \leq l \leq l_2$ and are given by

$$K_g(\Delta,r) = \sum_{l=l_1}^{l_2} G_l(r)\mathrm{P}_l^0(\cos\Delta) \qquad (13)$$

where $\mathrm{P}_l^0(\cos\Delta)$ and $G_l(r)$ are Legendre Polynomials and radial geoid kernels, respectively. We calculated the radial kernels using the method of *Forte* [2000], which assumes viscosity to vary radially. It should also be noted that the radial geoid kernels depend strongly on the choice of viscosity profile, although this becomes less important with increasing spherical harmonic degree. To compute geoid anomalies from eq. (12) requires an appropriate viscosity profile, a cut-off angular distance Δ_c and a spherical harmonic band-pass filter.

For each pixel of our posterior models viscosity profiles for the upper mantle were calculated in a consistent manner as a function of temperature and pressure following *Korenaga & Karato* [2008]. From these, we define two viscosity models, 1) a continental average where all viscosity profiles have been averaged over the entire study area (henceforth model CAV) and 2) a tectonic average (model RAV) where viscosity profiles are averaged within four tectonically distinct regions (labeled s, q, p and r in *Khan et al.* [2011] and corresponding to the North American craton, stable platforms, tectonically active areas, and surrounding oceans, respectively). For model RAV geoid kernels were computed for each tectonic viscosity profile, which implies that the radial geoid kernels in eq. (13) depend implictly on the angular distance ($G_l(r, \Delta)$). In the lower mantle (for depths>1200 km), where our thermo-chemical distributions are less well-constrained, we employ the recent viscosity model of *Soldati et al.* [2009], which was obtained by inverting global gravity data (GRACE). Additionally, at 660 km depth a viscosity ratio of 30 was imposed. The viscosity models so computed are shown in figure S4 (online supporting material). Note that in the case of the regionally averaged viscosity model (Figure S5b), most of the discrepancy between the different tectonic regions appears in the depth range 100-500 km.

Amplitudes of local geoid kernels rapidly decrease with angular distance even in the case of small values (<20°) of the cut-off angular distance Δ_c. here, we fixed Δ_c at 60°, which is sufficient for present purposes. We also tested other values of Δ_c, but did not observe any significant changes in the computed geoid for Δ_c >25°.

As upper (l_2) and lower (l_1) bounds on the band-pass filter, we set l_1=6 and l_2=20, respectively. Low spherical harmonic degrees (<5) mostly sample the deep (>1000 km) mantle (online supporting material figure S5), and can thus safely be discarded. In contrast, degrees l >10 are mostly sensitive to the upper mantle. Presently, we limited our expansion to l_2=20 to account for the fact that small-scale anomalies may be less well-resolved. Intermediate (6≤ l ≤9) degrees sample both the upper and lower mantle.

Keeping this intermediate range is thus likely to introduce a deep-mantle signal. On the other hand, removing this range would discard continental-scale variations that originate in the upper mantle, which likely introduces a stronger bias.

In Figure 13, we plotted the geoid anomalies (filtered in the band-pass $6 \leq l \leq 20$) predicted from our six posterior density models together with the geoid anomalies from model GGM02 [*Tapley et al.,* 2005] (plot A), which is based on GRACE data. Plots 1-6 were obtained with viscosity profile CAV discussed previously, whereas plots 7-12 were constructed using RAV viscosity profile. Observed geoid anomalies (plot A) show a clear continental division, with geoid lows (down to -30 m) in the north-east, and geoid highs (up to 30 m) along the Rocky Mountains and the Basin and Range. Geoid highs (lows) with smaller amplitude are also present over the Central plains (adjacent Pacific Ocean). Clearly, strong discrepancies exist between the geoid anomalies predicted by the different posterior models, suggesting that the geoid is potentially a useful filter to refine the collection of posterior models.

Overall, geoid anomalies obtained with the regionally averaged viscosity profile explain the observations better. The geoid anomaly model that agrees best with GGM02 is model 4 (plots 4 and 10). For geoid anomalies reconstructed with the regionally averaged viscosity profiles, correlation and variance reduction reach 0.63 and 0.40, respectively. A striking discrepancy of all our geoid anomaly maps, however, is the small amplitude of the geoid lows over the North American craton (around -10 m instead of -30 m in GGM02). This difference may be due to the low-resolution compositional parameterization in radial direction that we employed. The posterior density models may thus fail to capture the entire compositional signal, which is expected to be relatively strong and to vary with depth on a continental scale. Posterior models 1 (plots 1 and 7) and 2 (plots 2 and 8) also correlate reasonably well with the observed geoid (correlation of 0.42), but the relatively low amplitude geoid anomalies lead to poor variance reduction (0.16 and 0.10 for models 1 and 2, respectively). In contrast, model 6 is found to disagree strongly with GGM02 and may thus be removed from the collection of posterior models.

Testing posterior tomographic models against gravity data is a promising tool for further refining tomographic models based entirely on seismic data. However, the method employed here to reconstruct geoid anomalies from posterior density models suffers two important limitations that should be kept in mind. First, it strongly depends on the assumed radial viscosity structure, which is not well constrained. To illustrate this, we conducted additional calculations, in which we used three global radial viscosity models from the studies of *Ricard et al.* [1993]; *Forte & Mitrovica* [2004]; *Yoshida & Nakakuki* [2009]. For all three cases, we obtained geoid anomalies that correlate to a reasonably extent with GGM02 (from 0.40 to 0.60, depending on the particular viscosity model), but strongly differ in amplitude, resulting in negative variance reductions. The continental and regional viscosity models based on the posterior thermo-chemical distributions give better results, but uncertainty and errors exist in the rheological parameters we used [*Korenaga & Karato,* 2008] that are propagated to the viscosity profiles. Second, the spectral method we employ here to model mantle viscous flow neglects the toroidal part of the flow, i.e. lateral viscosity variations are not accounted for. Because viscosity controls the dynamic topography, this may have strong implications for the geoid kernels and anomalies. Our results suggest that the use of regional viscosity models partially compensates for this neglect of the toroidal flow. Furthermore, geoid reconstructions based on a finite-volume model of thermal convection that account for lateral viscosity variations, suggest that the effect of lateral viscosity variations on geoid anomalies is moderate and vary with location [*Cammarano et al.,* 2011]. Additional studies, including the calculation of geoid kernels from finite-volume convection models, should be conducted in order to obtain more detailed insights into the effect of including lateral viscosity variations.

7. Conclusion

The primary purpose of the present study has been to describe an alternative means of inverting seismic surface-wave phase velocities and their uncertainties, through which relatively robust measures of resolution and uncertainty can be obtained. Specfically, this was facilitated by the use of a Markov chain Monte Carlo method that works by performing a random walk in a multidimensional model space. It combines prior information with information from measurements and from the theoretical relationship between data and model parameters. This was effectuated using the Metropolis algorithm. As output we assimilated random realizations of the posterior *pdf*, which contains all the information about our parameterized physical system. We presented the outcome as a collection of tomographic images that all fit data within uncertainties. The emphasis here is on drawing inferences from such an assembly of models, rather than just a single image.

As outlined, our method also goes beyond the traditional approach of inverting seismic data for seismic wave speeds, by employing a self-consistent thermodynamic technique in order that the former can be inverted directly for thermo-chemical structure of the Earth's mantle. The obvious advantage of inverting for a set of parameters that describe the system being studied at the fundamental level of chemical composition and temperature, is that all physical properties are derived from these parameters. As a result, the use of simplified scaling relationships that seek to bridge the limited sensitivity of a given data set with regard to other structural parameters is entirely obviated. Moreover, the approach allows us to naturally link geophysical data that are not *a priori* related such as seismic, gravity and electromagnetic sounding data, thus opening the avenue for joint inversions across different geophysical fields.

To render the current study feasible, we considered phase velocity maps as data in place of phase velocity measurements, from which the former are derived. As a result, we inverted for a set of local 1D radial profiles spanning the North American continent and parts of the adjacent Pacific Ocean. Data uncertainties derive from a model space search technique to fit phase velocity measurements of fundamental-mode and higher-order Rayleigh and Love waves, which is deemed to provide an adequate estimate of uncertainties on the resultant phase velocity maps [*Visser et al.,* 2008a]. Inspite of this, we nonetheless consider the present approach successfull, inasmuch as there is considerable agreement between present and previous seismic shear-wave tomography models, which, given the fundamentally different approaches, is considered strong evidence in support of our method. Keeping in mind that we presently do not consider uncertainties on thermodynamic parameters, in addition to assuming a thermodynamically equilibrated mantle, we observe that

1. the North American upper mantle thermo-chemical and physical structure follows the surface tectonic age-division closely,

2. the old stable continental parts were found to be cold and Fe-depleted, while the tectonically younger continental regions and oceanic lithosphere appeared relatively hot and Fe-enriched,

3. shear-wave velocity differences between oceans and continents disappeared around 300 km depth,

4. within the transition zone a decoupling of the structure (thermo-chemical and anisotropic) from that of the upper mantle occured, accompanied by an overall decrease of amplitudes of velocity anomalies,

5. the lower mantle is characterized by an overall absence of strong heterogeneities so prominent in the upper mantle, although there is evidence for compositionally distinct upper and lower mantles and that

6. the anisotropic upper mantle structure is akin to what has been observed in some previous studies, in paricular a significant positive ξ signal is present beneath the old stable continental part, whereas younger areas are typically characterized by negative shear-wave anisotropy. In the transition zone a general reversal of the anisotropy signal compared to above is observed, which seems to repeat, albeit to a lesser extent, in the lower mantle at 1000 km depth. This likely reflects the presence of distinct anisotropic layers in the mantle, and finally

7. testing posterior tomographic models using geoid anomalies, which are sensitive to density, presents a promising tool for refining the collection of sampled tomographic and thermo-chemical models. A current limitation, however, is the accuracy of the reconstructed geoid, which requires a good knowledge of the mantle viscosity structure.

Acknowledgments. We would like to thank K. Visser and J. Trampert for distributing their global surface-wave phase-velocity maps. We are also grateful to M. Nettles and H. Yuan for sharing their tomography models with us. Comments by J. Ritsema on an earlier manuscript were much appreciated. As were discussions with L. Boschi. This work was supported by Swiss National Science Foundation grants 200021-130411 (AK) and 200021-111870 (FD). We would also like to thank G. Nolet for a copy of his code computing surface-wave phase velocity dispersion curves. Numerical computations were performed on the ETH cluster Brutus. Figures were generally prepared using GMT (Wessel & Smith, 1998).

References

Afonso, J. C., M. Fernàndez, G. Ranalli, G., W. L. Griffin, J. A. D. Connolly, Combined geophysical-petrological modelling of the lithospheric-sublithospheric upper mantle: methodology and applications, *Geochem. Geophys. Geosyst., 9*, Q05008, doi:10.1029/2007GC001834, 2008.

Afonso, J.C., G. Ranalli, M. Fernàndez, W. L. Griffin, S. Y. O'Reilly & U. Faul, On the Vp/Vs - Mg# correlation in mantle peridotites: Implications for the identification of thermal and compositional anomalies in the upper mantle, *Earth Planet. Sci. Lett., 289*, 606, 2010.

Aki, K., A. Christoffersen & E. Husebye, Determination of the three-dimensional structure of the lithosphere, *J. Geophys. Res., 82*, 277, 1977.

Anderson, D. L., *Theory of the Earth*, Blackwell, Oxford, 1989.

Anderson, D. L. & A. M. Dziewonski, Upper mantle anisotropy: Evidence from free oscillations, *Geophys. J. R. Astron. Soc., 69*, 383, 1982.

Artemieva, I. M., Global 1°C×1°C thermal model TC1 for the continental lithosphere: implications for lithosphere secular evolution, *Tectonophysics, 416*, 245, 2006.

Artemieva, I. M., The continental lithosphere: Reconciling thermal, seismic and petrologic data, *Lithos, 109*, 23, 2009.

Babuska, V. & Cara, M., *Seismic anisotropy in the Earth*, Kluwer Academic Press, Boston, USA, 1991.

Becker, T. W., B. Kustowski & G. Ekström, Radial seismic anisotropy as a constraint for upper mantle rheology *Earth Planet. Sci. Lett., 267*, 213, 2008.

Beghein, C. & J. Trampert, Probability density function for radial anisotropy: implications for the upper 1200 km of the mantle, *Earth Planet. Sci. Lett., 217*, 151, 2003.

Bernardo, J. M. & A. F. M. Smith, *Bayesian Theory*, Chichester: Wiley, 1994.

Bodin, T., Sambridge, M. & Gallagher, K., A self-parameterising partition model approach to tomographic inverse problems, *Inverse Problems, 25*, 055009, doi:10.1088/0266-5611/25/5/055009, 2009.

Bosch, M., Lithologic tomography: From plural geophysical data to lithology estimation, *J. Geophys. Res., 104*, 749, 1999.

Boschi, L. & A. M. Dziewonski, bigh and low resolution images of the Earth's mantle - Implications of different approaches to tomographic modeling, *J. Geophys. Res., 104*, 25567, 1999.

Boschi, L. & A. M. Dziewonski, Whole Earth tomography from delay times of P, PcP, PKP phases: lateral heterogeneities in the outer core, or radial anisotropy in the mantle?, *J. Geophys. Res., 105,* 675, 2000.

Boschi, L. & G. Ekström, New images of the Earth's upper mantle from measurements of surface-wave phase velocity anomalies, *J. Geophys. Res., 107*, 10.1029/2000JB000059, 2002.

Boyd, F. R., Compositional distinction between oceaninc and cratonic lithosphere, *Earth Planet. Sci. Lett., 96*, 15, 1989.

Cammarano, F., B. Romanowicz, L. Stixrude, C. Lithgow-Bertelloni & W. Xu, Inferring the thermochemical structure of the upper mantle from seismic data, *Geophys. J. Int., 179*, 1169, doi: 10.1111/j.1365-246X.2009.04338.x.

Cammarano, F., P.J. Tackley, and L. Boschi, Seismic, petrological and geodynamical constraints on thermal and compositional structure of the upper mantle: global thermo-chemical models, *submitted to Geophys. J. Int.*, 2011.

Cobden, L., S. Goes, F. Cammarano & J. A. D. Connolly, Thermochemical interpretation of one-dimensional seismic reference models for the upper mantle: evidence for bias due to heterogeneity, *Geophys. J. Int., 175*, 627, 2008.

Connolly, J. A. D., Computation of phase equilibria by linear programming: A tool for geodynamic modeling and an application to subduction zone decarbonation, *Earth Planet. Sci. Lett., 236*, 524, 2005.

Debayle, E., B. L. N. Kennett & K. Priestley, Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia, *Nature, 433*, 509, 2005.

Deschamps, F., R. Snieder & J. Trampert, The relative density-to-shear velocity scaling in the uppermost mantle, *Phys. Earth Planet. Int., 124*, 193, 2001.

Deschamps, F. & J. Trampert, Mantle tomography and its relation to temperature and composition, *Phys. Earth Planet. Int., 140*, 277-291, 2003.

Deschamps, F., S. Lebedev, T. Meier & J. Trampert, Stratified seismic anisotropy reveals past and present deformation beneath the East-central United States, *Earth Planet. Sci. Lett., 274*, 489, doi: 10.1016/j.epsl.2008.07.058, 2008.

Dziewonski, A. M. & D. L. Anderson, Preliminary reference Earth model, *Phys. Earth Planet. Int., 25*, 297, 1981.

Dziewonski, A. M., B. H. Hager & R. J. O'Connell, Large-scale heterogeneity in the lower mantle, *J. Geophys. Res., 82*, 239, 1977.

Ekström, G., J. Tromp & E. W. F. Larson, Measurements and global models of surface wave propagation, *J. Geophys. Res., 102*, 8137, 1997.

Ekström, G. & A. M. Dziewonski, The unique anistropy of the Pacific upper mantle, *Nature, 394*, 168, 1998.

Forte, A.M., Seismic-geodynamic constraints on mantle flow: implications for layered convection, mantle viscosity, and seismic anisotropy in the deep mantle, in S.-I. Karato et al. (Eds), Earth's deep interior: mineral physics and tomography from the atomic to the global scale, *Geophysical Monograph Ser.*, **117**, American Geophysical Union, Washington, DC, pp. 3-36, 2000.

Forte, A.M. and W.R. Peltier, Plate tectonics and aspherical Earth's structure: the importance of poloidal-toroidal coupling, *J. Geophys. Res., 92*, 3645, 1987.

Faul, U.H. & I. Jackson, The seismological signature of temperature and grain size variations in the upper mantle, *Earth Planet. Sci. Lett., 234*, 119, 2005.

Frederiksen, A. W., M. G. Bostock & J. F. Cassidy, S wave velocity structure of the Canadian upper mantle, *Phys. Earth Planet. Int., 124*, 175, 2001.

Gaherty, J., A surface wave analysis of seismic anisotropy beneath eastern North America, *Geophys. J. Int., 158*, 1053, 2004.

Gaul, O. F., W. L. Griffin, S. Y. O'Reilly & N. J. Pearson, Mapping olivine composition in the lithospheric mantle, *Earth Planet. Sci. Lett., 182*, 223, 2000.

Godey, S., F. Deschamps F., J. Trampert & R. Snieder, Thermal and compositional anomalies beneath the North American continent, *J. Geophys. Res., 109*, B01308, doi: 10.1029/2002JB002263, 2004.

Goes, S. & S. van der Lee, Thermal structure of the North American uppermost mantle inferred from seismic tomography, *J. Geophys. Res., 107*, 11153, 2002.

Goes, S., F. J. Simons & K. Yoshizawa, Seismic constraints on temperature of the Australian uppermost mantle, *Earth Planet. Sci. Lett., 236*, 227, 2005.

Grand, S. P., Mantle shear structure beneath the americas and surrounding oceans, *J. Geophys. Res., 99*, 11591, 1994.

Grand, S. P., R. D. Van der Hilst & S. Widiyantoro, *GSA Today, 7*, 1, 1997.

Griffin, W. L et al., Layered mantle lithosphere in the Lac du Gras Area, Slave Craton: Composition, structure and origin, *J. Petrol., 40*, 705, 1999.

Gung, Y, M. Panning & B. Romanowicz, Global anisotropy and the thickness of continents, *Nature, 442*, 707, 2003.

Hager, B.H., and M.A. Richards, Long wavelength variations in Earth's geoid: physical models and dynamical implications, *Philos. Trans. R. Soc. Lon. A, 328*, 309-327.

Hastings, W. K., Monte Carlo sampling methods using Markov chains and their applications, *Biometrika, 57*, 97, 1970.

Irifune, T., Absence Of An Aluminous Phase In The Upper Part Of The Earths Lower Mantle, *Nature, 370*, 131, 1994.

Jackson, I., Laboratory Measurements of Seismic Wave dispersion and Attenuation: Recent Progress, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale*, Geophys. Monogr. Ser., vol. 117, eds. S. Karato et al., AGU, Washington, D. C., 265, 2000.

Jackson, I., J. D. Fitz Gerald, U. H. Faul & B. H. Tan, Grain-size-sensitive seismic wave attenuation in polycrystalline olivine, *J. Geophys. Res., 107*, doi:10.1029/2001JB001225, 2002.

Jaupart, C. & J.-C. Mareschal, Thermal structure and thickness of continental roots, *Lithos, 48*, 93, 1999.

Jordan, T. H., The continental tectosphere, *Rev. Geophys. Space Phys., 13*, 1, 1975.

Jordan, T. H., Composition and development of the continental tectosphere, *Nature, 274*, 544, 1978.

Karato, S.-I., Seismic anisotropy in the deep mantle, boundary layers and the geometry of mantle convection, *Pure Appl. Geophys., 151*, 565, 1998.

Khan, A., Mosegaard, K., Williams, J. G. & Lognonné, P., Does the Moon possess a molten core? Probing the deep lunar interior using results from LLR and Lunar Prospector, *J. Geophys. Res., 109*, E09007, doi:10.1029/2004JE002294, 2004.

Khan, A., J. A. D. Connolly, J. Maclennan, and K. Mosegaard, Joint inversion of seismic and gravity data for lunar composition and thermal state, *Geophys. J. Int., 168*, 243, doi:10.1111/j.1365-246X.2006.03200.x, 2007.

Khan, A., L. Boschi & J. A. D. Connolly, On mantle chemical and thermal heterogeneities and anisotropy as mapped by inversion of global surface wave data, *J. Geophys. Res., 114*, B09305, doi:10.1029/2009JB006399, 2009.

Khan, A., L. Boschi & J. A. D. Connolly, Mapping the Earth's thermo-chemical and anisotropic structure using global surface wave data, *J. Geophys. Res., 116*, B01301, doi:10.1029/2010JB007828, 2011.

Kogan, M.G., and M.K. McNutt, Gravity field over northern Eurasia and variations in the strength of the upper mantle, *Science, 259*, 473, 1993.

Koren, Z., K. Mosegaard, E. Landa, P. Thore and A. Tarantola, Monte Carlo Estimation and Resolution Analysis of Seismic Background Velocities, *J. Geophys. Res., 96*, B12, 1991.

Korenaga, J., and S.-I. Karato, A new analysis of experimental data on olivine rheology, *J. Geophys. Res., 113*, B02403, doi: 10.1029/2007JB005100, 2008.

Kuskov, O. L., V. A. Kronrod & H. Annersten, Upper mantle temperatures from seismic and geochemical constraints: Implications for the Kaapvaal craton, *Earth Planet. Sci. Lett., 244*, doi:10.1016/j.epsl.2006.02.016, 2006.

Kuskov, O. L., V. A. Kronrod and A. A. Prokof'ev, Thermal Structure and Thickness of the Lithospheric Mantle Underlying the Siberian Craton from the Kraton and Kimberlit Superlong Seismic Profiles, *Phys. Solid Earth, 47*, doi:10.1134/S1069351310111011, 2011.

Kustowski, B., G. Ekström & A. M. Dziewonski, Anisotropic shear-wave velocity structure of the Earth's mantle: A global model *J. Geophys. Res., 113*, B06306, doi: 10.1029/2007JB005169, 2008.

Lebedev, S. & R. D. van der Hilst, Global upper-mantle tomography with the automated multimode inversion of surface and S-wave forms, *Geophys. J. Int., 173*, 505, 2008.

Li, A., D. W. Forsyth & K. M. Fischer, Shear velocity structure and azimuthal anisotropy beneath eastern North America from Rayleigh wave inversion, *J. Geophys. Res., 108*, doi:10.1029/2002JB002259, 2002.

Long, M. D. & T. W. Becker, Mantle dynamics and seismic anisotropy, *Earth Planet. Sci. Lett., 297*, 341, 2010.

Lyubetskaya, T. & J. Korenaga, Chemical composition of Earth's primitive mantle and its variance: 1. Method and results, *J. Geophys. Res., 112*, doi:10.1029/2005JB004223, 2007.

McKenzie, D. P., J. A. Jackson & K. F. Priestley, Thermal structure of oceanic and continental lithosphere, *Earth Planet. Sci. Lett., 233*, 337, 2005.

Marone, F, Y. Gung & B. Romanowicz, Three-dimensional radial anisotropic structure of the North American upper mantle from the inversion of surface waveform data, *Geophys. J. Int., 171*, 206, doi:10.1111/j.1365-246X.2007.03465, 2007.

Masters, G., G. Laske, H. Bolton & A. M. Dziewonski, The relative behaviour of shear velocity, bulk sound speed, and compressional velocity in the mantle: Implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography From the Atlantic to the Global Scale, Geophys. Monogr. Ser.*, **117**, ed. S. Karato et al., AGU, Washington, D. C, 63, 2000.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E., Equation of state calculations by fast computing machines, *J. Chem. Phys., 21*, 1087, 1953.

Mitrovica, J.X. and A.M. Forte, A New Inference of Mantle Viscosity Based Upon a Joint Inversion of Convection and Glacial Isostatic Adjustment Data, *Earth Planet. Sci. Lett., 225*, 177, 2004.

Montagner, J. P., Where can seismic anisotropy be detected in the Earth's mantle ? *Pure Appl. Geophys., 151*, 223, 1998.

Montagner, J. P. & T. Tanimoto, Global upper mantle tomography of seismic velocities and anisotropies, *J. Geophys. Res., 96*, 20337, 1991.

Montagner, J. P. & B. L. N. Kennett, bow to reconcile body-wave and normal-mode reference earth models, *Geophys. J. Int., 125*, 229, 1996.

Mosca, I., *Probabilistic tomography using body-wave, normal-mode and surface-wave data*, Ph.D. Thesis, University of Utrecht, The Netherlands, 2010.

Mosegaard, K., Resolution analysis of general inverse problems through inverse Monte Carlo sampling, *Inverse Problems, 14*, 405, 1998.

Mosegaard, K., Quest for consistency, symmetry and simplicity - The legacy of Albert Tarantola, *Geophysics*, in press, 2011.

Mosegaard, K. & A. Tarantola, Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res., 100*, 12431, 1995.

Mosegaard, K. & M. Sambridge, Monte Carlo analysis of inverse problems *Inverse Problems, 18*, R29, 2002.

Nettles, M. & A. M. Dziewonski, Radially anisotropic shear-velocity structure of the upper mantle globally and beneath North America, *J. Geophys. Res., 113*, B02303, doi:10.1029/2006JB004819, 2008.

Panning, M. & B. Romanowicz, A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int., 167*, 361, 2006.

Priestley, K. F. & D. P. McKenzie, The thermal structure of the lithosphere from shear wave velocities, *Earth Planet. Sci. Lett., 244*, 285, 2006.

Rawlinson, N., S. Pozgay & S. Fishwick, Seismic tomography: A window into deep Earth, *Phys. Earth Planet. Int., 178*, 101, 2010.

Ricard, Y., L. Fleitout, and C. Froideveaux, Geoid heights and lithospheric stresses for a dynamic Earth, *Ann. Geophys., 2*, 267, 1984.

Ricard, Y., M. Richards, C. Lithgow-Bertelloni, and Y. Le Stunff, A geodynamic model of mantle density heterogeneity, *J. Geophys. Res., 98*, 21895, 1993.

Ritsema, J, H. J. Van Heijst & J. H. Woodhouse, Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science, 286*, 1925, 1999.

Ritsema, J, H. J. Van Heijst & J. H. Woodhouse, Global transition zone tomography, *J. Geophys. Res., 109*, B02302, doi:10.1029/2003JB002610, 2004.

Ritsema, J, W. Xu, L. Stixrude & C. Lithgow-Bertelloni, Estimates of the transition zone temperature in a mechanically mixed upper mantle, *Earth Planet. Sci. Lett., 244*, doi:10.1016/j.epsl.2008.10.024, 2009.

Ritsema, J., H. J. van Heijst, A. Deuss, and J. H. Woodhouse, S40RTS: a degree-40 shear velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltimes, and normal-mode splitting function measurements, *Geophys. J. Int., 184*, doi:10.1111/j.1365-246X.2010.04884.x, 2011.

Ritzwoller, M. H., N. M. Shapiro & S.-J. Zhong, Cooling history of the Pacific lithosphere, *Earth Planet. Sci. Lett., 226*, 69, 2004.

Romanowicz, B., Global mantle tomography: Progress status in the past 10 years, *Ann. Rev. Earth Planet. Sci., 31*, 303, 2003.

Rudnick, R. L., W. F. McDonough & R. J. O'Connell, Thermal structure, thickness and composition of continental lithosphere, *Chem. Geol., 145*, 395, 1998.

Röhm, A. H. E., R. Snieder, S. Goes & J. Trampert, Thermal structure of continental upper mantle inferred from S-wave velocity and surface heat flow, *Earth Planet. Sci.*

Lett., 181, 395, 2000.

Sambridge, M. & K. Mosegaard, Monte Carlo methods in Geophysical inverse problems, *Rev. Geophys., 40*, doi:10.1029/2000RG000089, 2002.

Sengupta, M. K. & M. N. Toksöz, Three-dimensional model of seismic velocity variation in the Earth's mantle, *Geophys. Res. Lett., 3*, 84, 1977.

Shapiro, N. M. & M. H. Ritzwoller, Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int., 151*, 88, 2002.

Shapiro, N. M. & M. H. Ritzwoller, Thermodynamic constraints on seismic inversions, *Geophys. J. Int., 157*, doi: 10.1111/j.1365-246X.2004.02254.x, 2004.

Sigloch, K., N. McQuarrie & G. Nolet, Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nature Geoscience*, doi:10.1038/ngeo231, 2008.

Simmons, N. A., A. M. Forte & S. P. Grand, Joint seismic, geodynamic and mineral physical constraints on three-dimensional mantle heterogeneity: Implications for the relative importance of thermal versus compositional heterogeneity *Geophys. J. Int., 177*, 1284, 2009,

Smith, W. H. F. & P. Wessel, Gridding with continuous curvature splines in tension, *Geophysics, 55*, 293, 1990.

Sobolev, S. V. et al., Upper mantle temperatures from teleseismic tomography of French Massif Central including effects of composition, mineral reactions, anharmonicity, anelasticity and partial melt, *Earth Planet. Sci. Lett., 157*, 193, 1996.

Soldati, G., L. Boschi, F. Deschamps, and D. Giardini, Inferring radial models of mantle viscosity from gravity (GRACE) data and an evolutionary algorithm, *Phys. Earth*

Planet. Inter., 176, 19, 2009.

Stixrude, L., and C. Lithgow-Bertelloni, Mineralogy and elasticity of the oceanic upper mantle: Origin of the low-velocity zone, *J. Geophys. Res., 110*, B03204, doi:10.1029/2004JB002965, 2005.

Tanimoto, T. & D. L. Anderson, Mapping convection in the mantle, *Geophys. Res. Lett., 11*, 287, 1984.

Tapley, B., J. Ries, S. Bettadpur, D. Chambers, M. Cheng, F. Condi, B. Gunter, Z. Kang, P. Nagel, R. Pastor, T. Pekker, S. poole, and F. Wang, GGM02 - An improved Earth gravity field model from GRACE, *J. Geodesy, 79*, 467, 2005.

Tarantola, A., *Inverse problem theory and methods for model parameter estimation*, SIAM, Philadelphia, USA, 2005.

Tarantola, A., and B. Valette, Inverse problems: Quest for information, *J. Geophys., 50*, 159, 1982.

Tian, Y., K. Sigloch & G. Nolet, Multiple-frequency SH-wave tomography of the western US upper mantle, *Geophys. J. Int., 178(3)*, 1384-1402, doi:10.1111/j.1365-246X.2009.04225.x, 2009.

Trampert, J., Global seismic tomography: the inverse problem and beyond, *Inverse Problems, 14*, 371, doi: 10.1088/0266-5611/14/3/002, 1998.

Trampert, J. & J. H. Woodhouse, Assessment of global phase velocities, *Geophys. J. Int., 144*, 165, 2001.

Trampert, J. & J. H. Van Heijst, Global Azimuthal Anisotropy in the Transition Zone, *Science, 296*, 1297, doi:10.1126/science.1070264, 2002.

Trampert, J., Deschamps, F., Resovsky, J. & D. Yuen, 2004. Chemical heterogeneities throughout the lower mantle, *Science, 306*, 853, 2004.

Trampert, J. & R. D. van der Hilst, Towards a quantitative interpretation of global seismic tomography, in *Earth's Deep Mantle: Structure, Composition and Evolution*, Geophys. Monogr. Ser., vol. 160, eds. R. D. van der Hilst et al., AGU, Washington, D. C, 47, 2005.

Van der Lee, S. & G. Nolet, Seismic image of the subducted trailing fragments of the Farallon plate, *Nature, 386*, 266, 1997a.

Van der Lee, S. & G. Nolet, Upper-mantle S-velocity structure of North America, *J. Geophys. Res., 102*, 22815, 1997b.

Van der Lee, S. & A. Frederiksen, Surface wave tomography applied to the North American upper mantle, *Seismic Earth: Array Analysis of Broadband Seismograms*, Geophys. Monogr. Ser., vol. 157, eds. A. Levander and G. Nolet, AGU, Washington, D. C., 67, 2005.

van Gerven, L., F. Deschamps, and R.D. van der Hilst, Geophysical evidence for chemical variations in the Australian continental mantle, *Geophys. Res. Lett., 31*, L17607, doi:10.1029/2004GL020307, 2004.

Visser, K., Trampert, J. & B. L. N. Kennett, Global anisotropic phase-velocity maps for higher mode Love and Rayleigh waves, *Geophys. J. Int., 172*, 1016, 2008a.

Visser, K., Trampert, J., S. Lebedev & B. L. N. Kennett, Probability of radial anisotropy in the deep mantle, *Earth Planet. Sci. Lett.*, doi:10.1016/j.epsl.2008.03.041, 2008b.

Wessel, P. & W. H. F. Smith, New, improved version of Generic Mapping Tools released, *EOS trans. AGU, 79*, 579, 1998.

Wookey, J., J.-M. Kendall & G. Barruol, Mid-mantle deformation inferred from seismic anisotropy. *Nature, 415*, 777, 2002.

Xu, W., C. Lithgow-Bertelloni, L. Stixrude & J. Ritsema, The effect of bulk composition and temperature on mantle seismic structure, *Earth Planet. Sci. Lett.*, doi:10.1016/j.epsl.2008.08.012, 2008.

Yoshida, M., and T. Nakakuki, Effects on the long-wavelength geoid anomaly of lateral viscosity variations caused by stiff subducting slabs, weak plate margins and lower mantle rheology, *Phys. Earth Planet. Inter., 172*, 278, 2009.

Yuan, H. & B. Romanowicz, Lithospheric layering in the North American craton, *Nature, 466*, doi:10.1038/nature09332, 2010.

Yuan, H., B. Romanowicz, K. Fisher & D. Abt, 3-D shear wave radially and azimuthally anisotropic velocity model of the North American upper mantle, *Geophys. J. Int., 184*, doi: 10.1111/j.1365-246X.2010.04901.x, 2010.

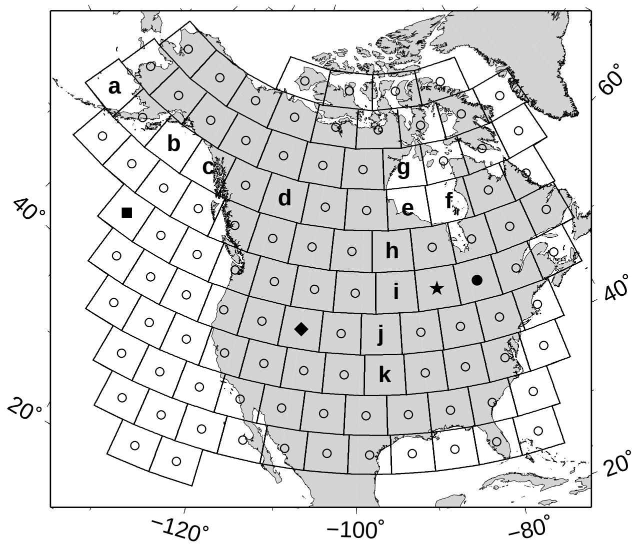


Figure 1. Parameterization of the model. Dots at the center of each pixel denote the locations at which properties are defined laterally. Grid spacing is 5°. Radially the model is parameterized in terms of layers (see section 4.1 for further discussion). Symbols (diamond, circle, square and star) indicate the location for which 1D marginal *pdf*s are shown in figures 3-6. Letters a-f refer to the locations for which radial shear-wave velocity profiles are displayed in figure 9, while letters e,g-k refer to figure 10.

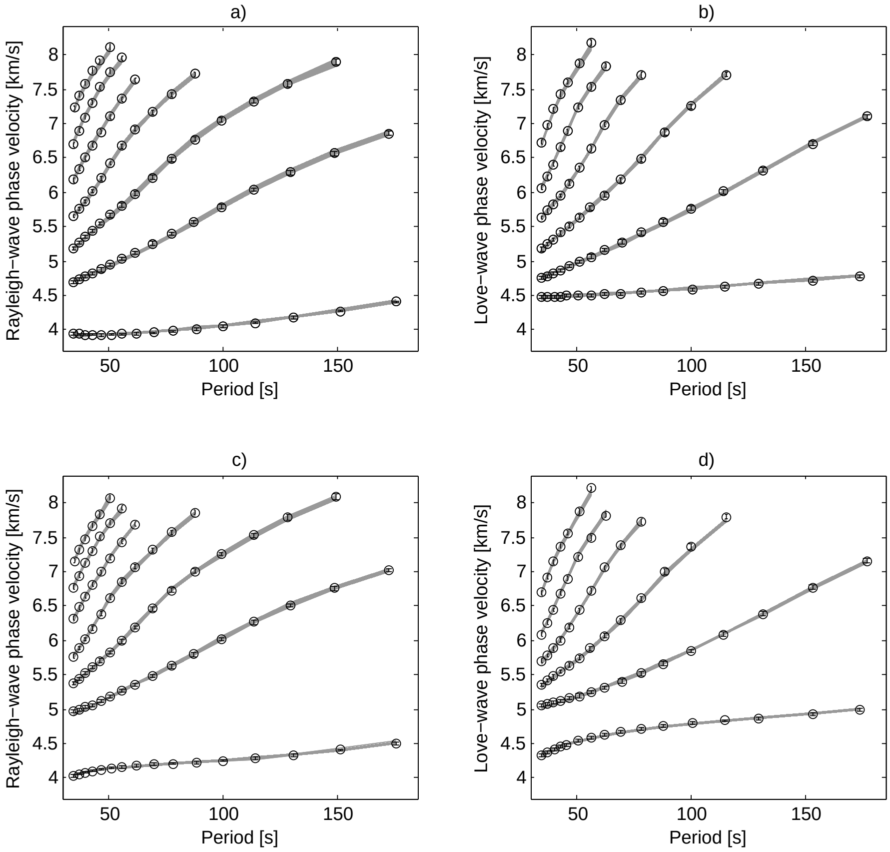


Figure 2. Datafit. Comparison of calculated (gray lines) and observed Rayleigh and Love-wave phase-velocities (circles), including uncertainties (error bars) at two different locations, which are shown in figure 1 (a,b - filled square and c,d - filled circle).

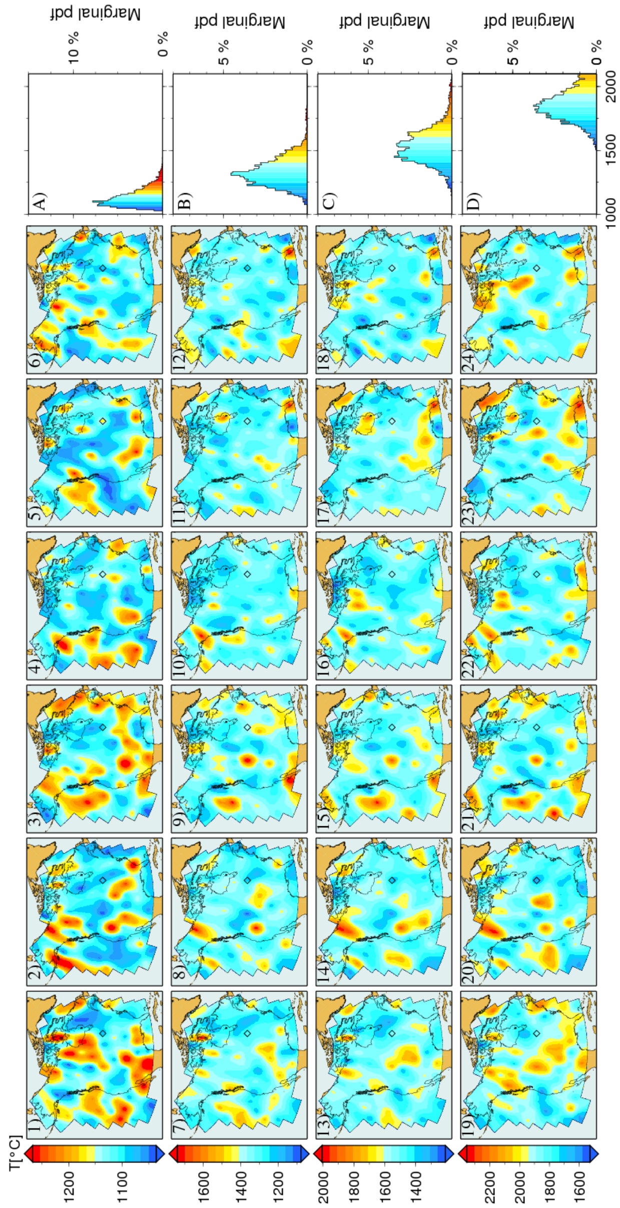


Figure 3. Prior thermal movie. The first six maps of each panel show six thermal models that are picked randomly from the prior distribution at a particular depth: first panel (plots labeled 1-6, A) - 100 km, second panel (plots labeled 7-12, B) - 300 km, third panel (plots labeled 13-18, C) - 500 km and fourth panel (plots labeled 19-24, D) - 1000 km. Plots labeled A-D indicate the prior marginal mantle temperature distribution at these depths for the pixel identified by the diamond (see also figure 1). For all maps shown here and in the following, data were gridded using continuous curvature splines (Smith & Wessel, 1990) in tension method (with tension set to 0.15) as implemented in GMT.

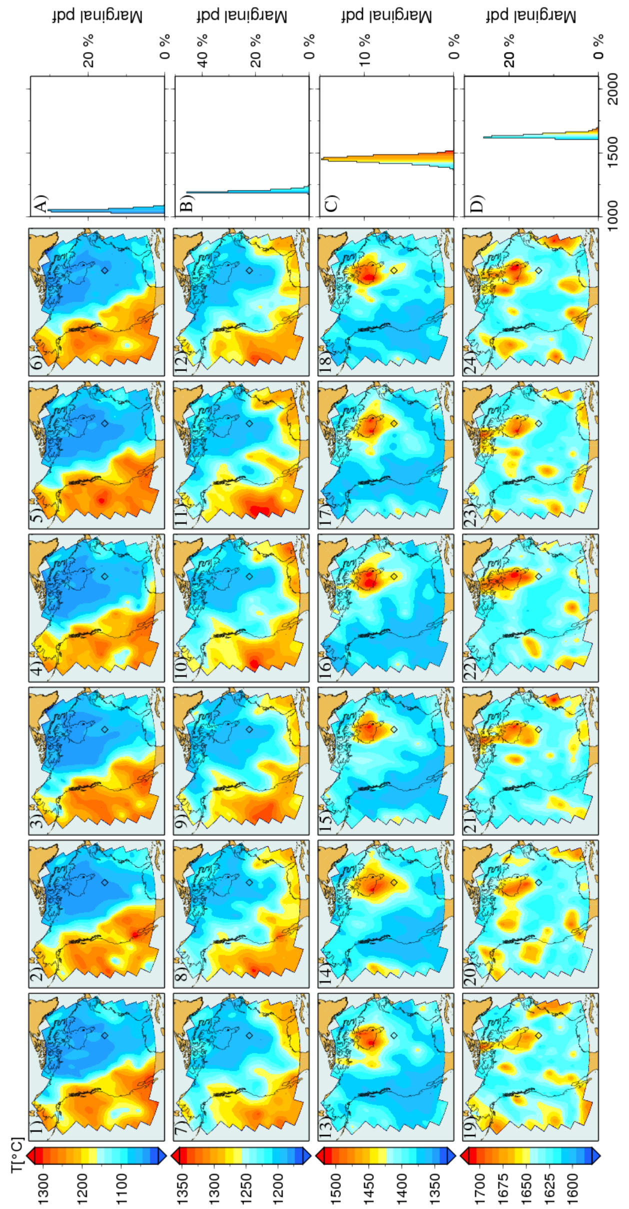


Figure 4. Posterior thermal movie. The first six maps of each panel show six thermal models that are picked randomly from the posterior distribution at the same depths as in figure 2, i.e. 100 km (1-6), 300 km (7-12), 500 km (13-18) and 1000 km (19-24) depth. Plots labeled A-D indicate the posterior marginal mantle temperature distribution at these depths for the pixel identified by the diamond (see also figure 1). Note that prior and posterior colourbars do not overlap in temperature range for a given depth, but that the temperature axis for the 1D marginal prior and posterior distributions have the same range.

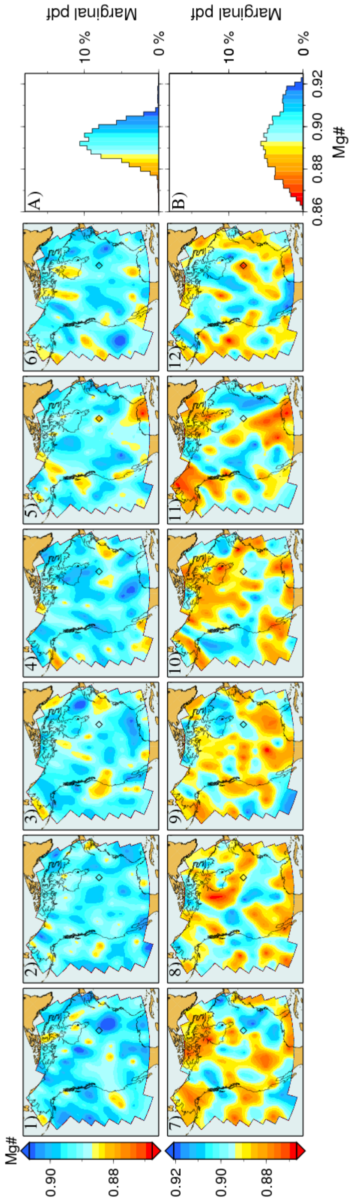


Figure 5. Prior compositional movie. The first six maps of each panel show six compositional (in terms of Mg#) models that are picked randomly from the prior distribution at depths of 100 km (1-6) and 1000 km (7-12). Plots labeled A and B indicate prior marginal distribution of upper and lower mantle composition (Mg#), respectively, at the above depths at the location of the pixel identified by the diamond (see also figure 1).

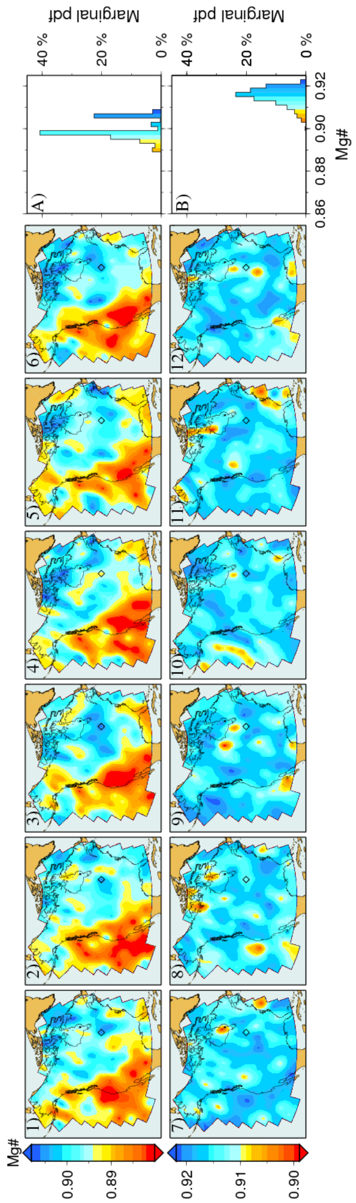


Figure 6. Posterior compositional movie. The first six maps of each panel show six compositional (in terms of Mg#) models that are picked randomly from the posterior distribution at depths of 100 km (1-6) and 1000 km (7-12). Plots labeled A and B indicate posterior marginal distribution of upper and lower mantle composition (Mg#), respectively, at the above depths at the location of the pixel identified by the diamond (see also figure 1). Note that prior and posterior colourbars do not overlap in Mg# range for a given depth, but that the Mg#-axis for the 1D marginal prior and posterior distributions have the same range.

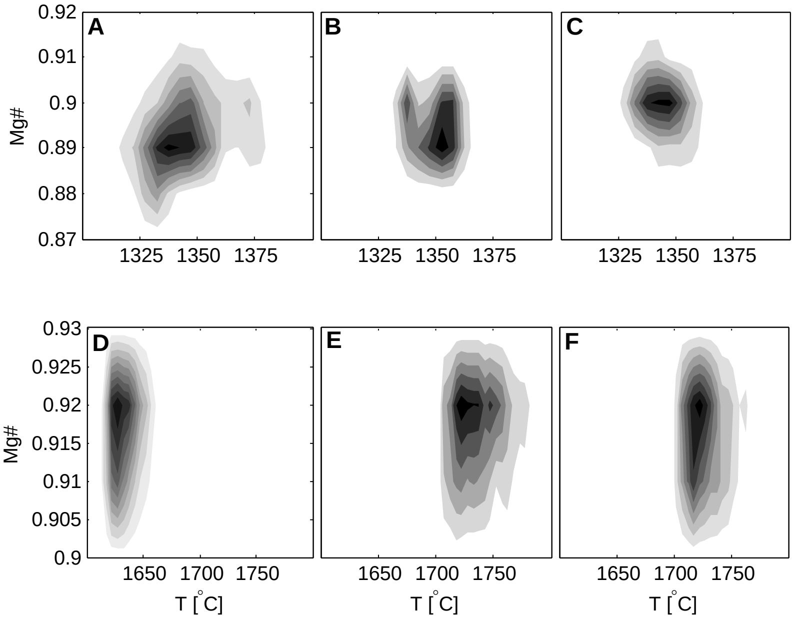


Figure 7. Two-dimensional (2D) marginal posterior probability density functions showing correlation between temperature and composition (here Mg#) for three different tectonic settings in the upper and lower mantle: Oceanic lithosphere (A,D), young continental lithosphere (B,E) and old continental lithosphere (C,F). Location of pixels are indicated in figure 1 by a filled square (A,D), a filled diamond (B,E) and a filled circle (C,F). The 2D marginals are envisioned as contours directly relating their probability of

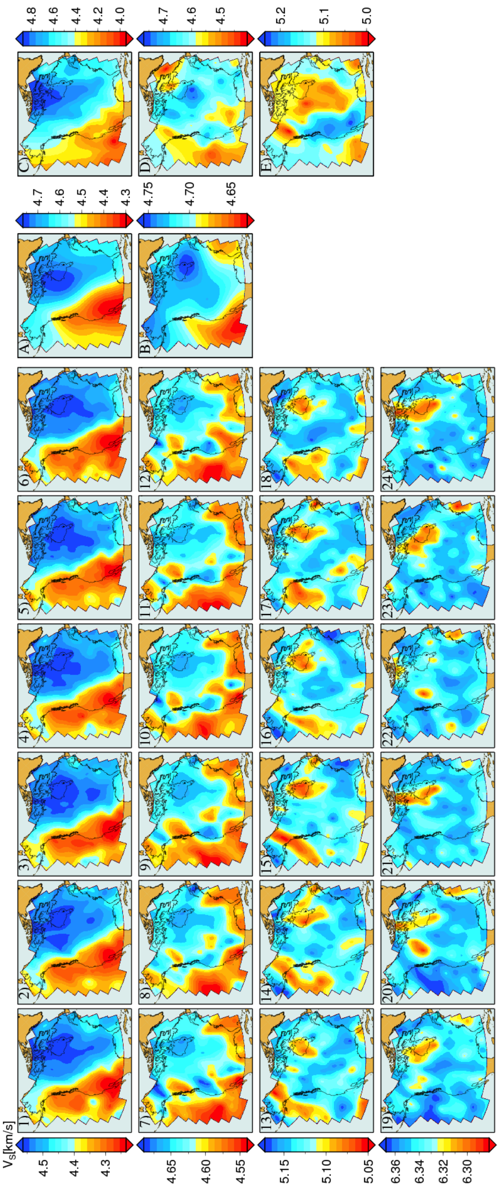
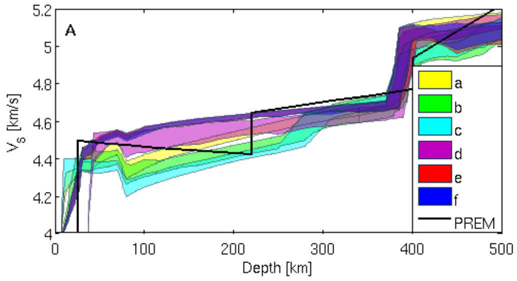


Figure 8. Posterior isotropic shear-wave velocity movie. In each panel the six maps represent six shear-wave velocity models that are picked randomly from the posterior distribution at depths of 100 km (1-6), 300 km (7-12), 500 km (13-18) and 1000 km (19-24), respectively. Note that prior and posterior colourbars do not bracket similar V_S ranges at individual depths. For comparison other regional shear-wave velocity models have been included. Plots A (at 100 km depth) and B (at 300 km depth) show the model of *Nettles & Dziewonski* [2008]; plots E (100 km), F (300 km) and G (500 km) depict the model of *Yuan et al.* [2010]. Note differences in absolute shear-wave velocities (colourbars) between present and previous models.



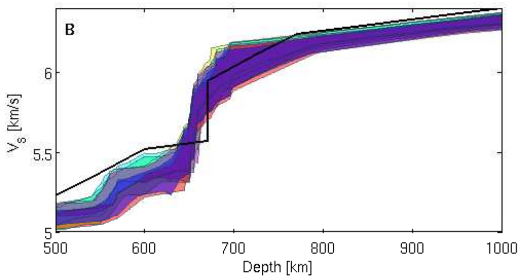


Figure 9. Selected shear-wave velocity models beneath different tectonic settings in the upper mantle and transition zone (A) and lower transition zone and mantle (B) : Oceanic (a-c), young continent (d) and old stable continent (e,f). Profiles encompass all sampled models. Geographic location of letters are indicated in figure 1.

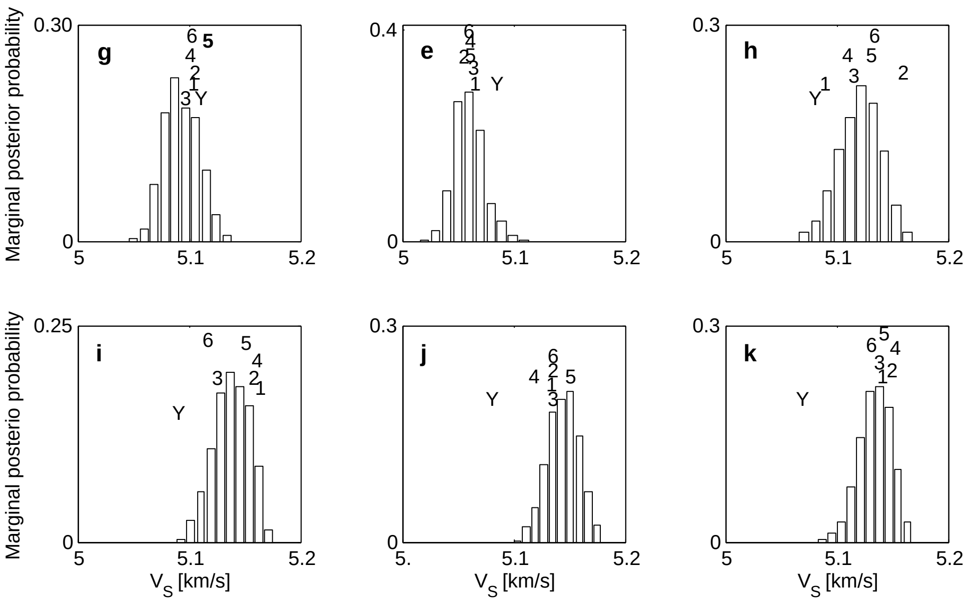


Figure 10. Marginal posterior probability distributions of sampled isotropic shear-wave velocities at 500 km depth beneath the North American continent. Location of pixels are shown in figure 1, with letters identifying the specific pixel. Numbers 1-6 above the distributions indicate the S-wave velocity for each of the six posterior models shown in figure 9 at 500 km depth (plots 13-18), while 'Y' is the shear-wave velocity at the same depth for the model of *Yuan et al.* [2010].

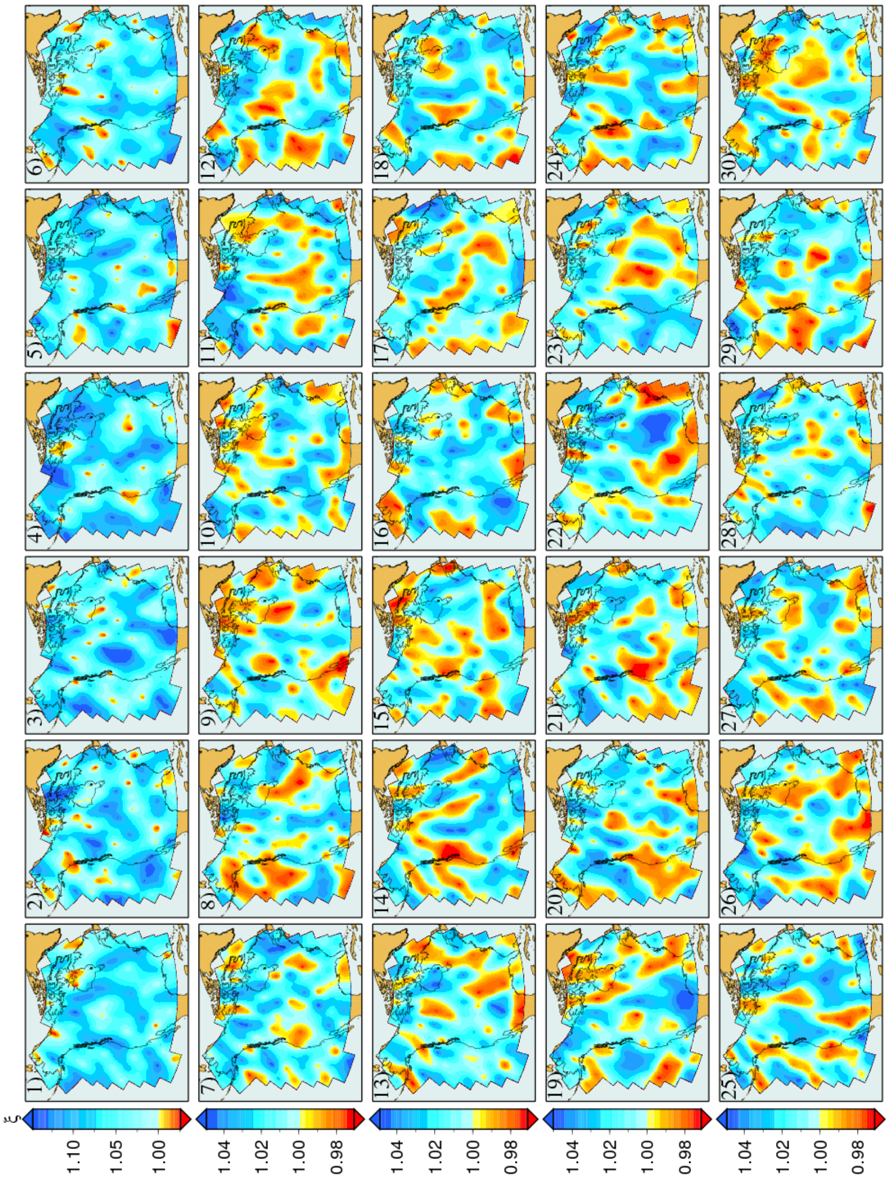


Figure 11. Prior shear-wave anisotropy movie. In each panel the six maps represent six different shear-wave anisotropy models that are picked randomly from the prior distribution at depths of 100 km (1-6), 300 km (7-12), 500 km (13-18) and 1000 km (19-24), respectively.

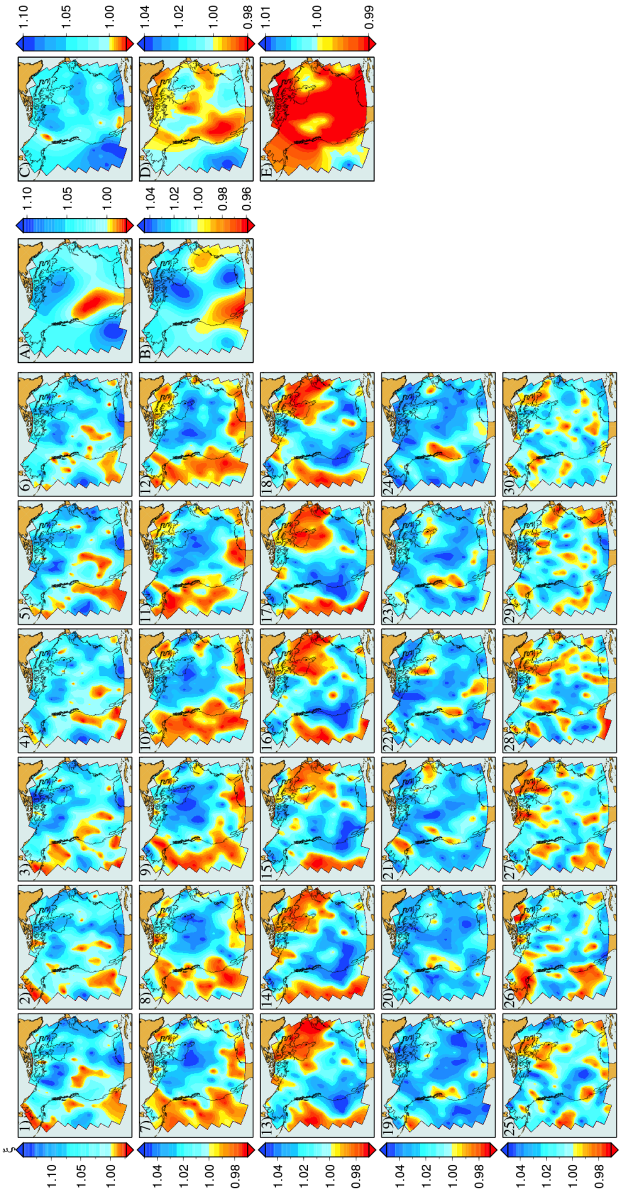


Figure 12. Posterior shear-wave anisotropy movie. In each panel the six maps represent six different shear-wave anisotropy models that are picked randomly from the posterior distribution at depths of 100 km (1-6), 300 km (7-12), 500 km (13-18), 1000 km (19-24) and 1800 km (25-30), respectively. Note that prior and posterior colourbars do not bracket similar ξ ranges. For comparison other regional anisotropic tomography models have been included. Plots A (at 100 km depth) and B (at 300 km depth) show the model of *Nettles & Dziewonski* [2008]; plots E (100 km), F (300 km) and G (500 km) depict the model of *Yuan et al.* [2010]. Note differences in colourbars, i.e. absolute shear-wave anisotropy, between present and previous models.

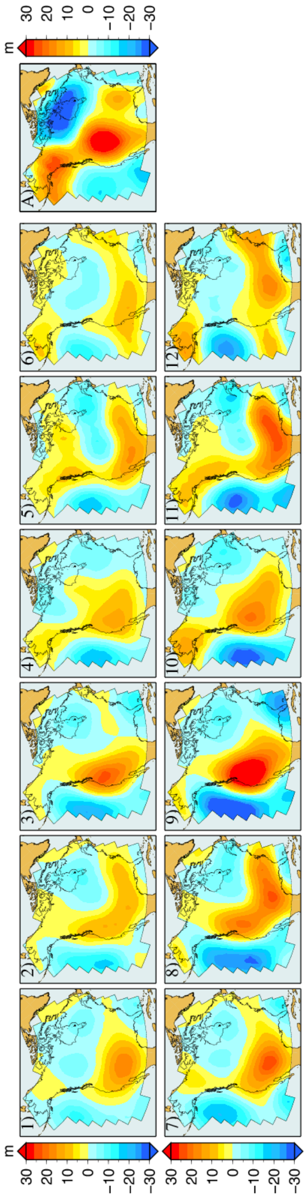
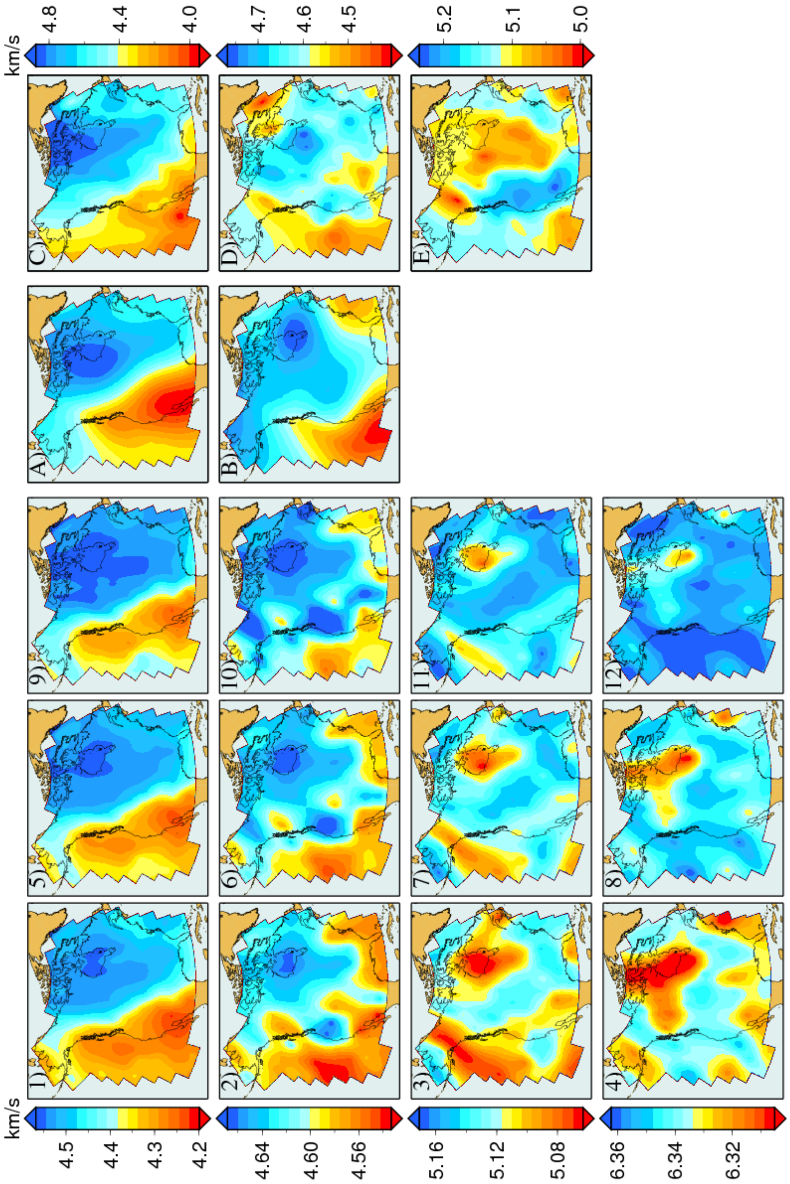
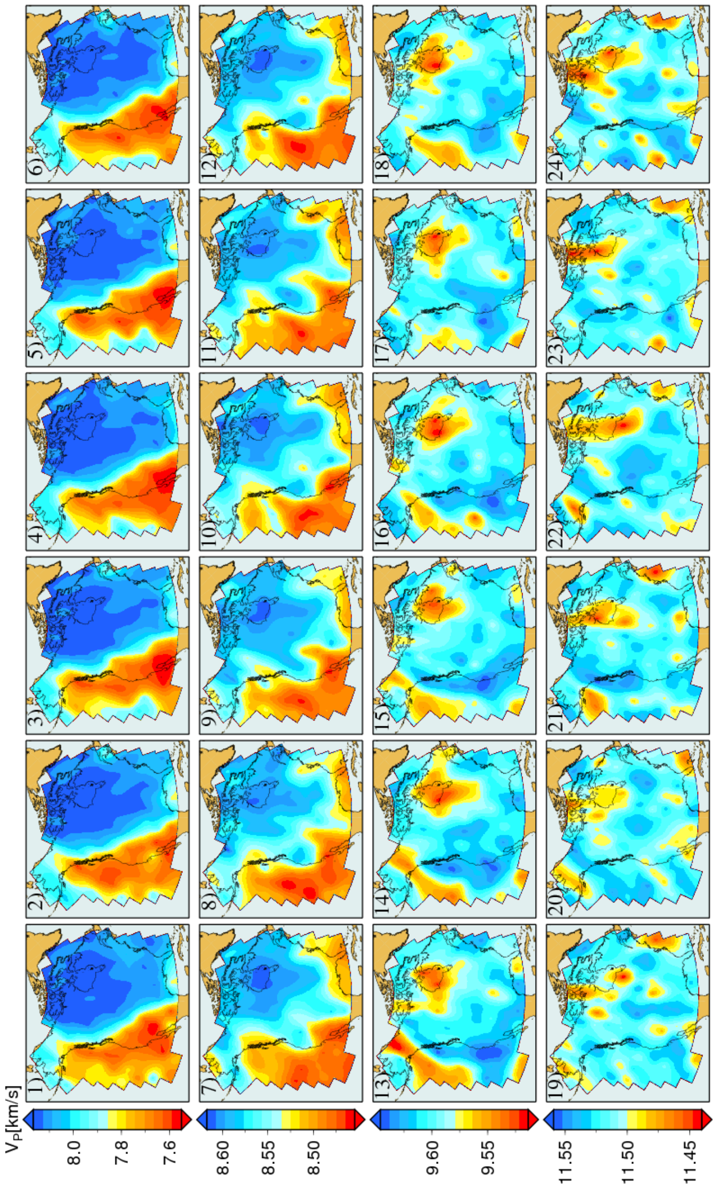


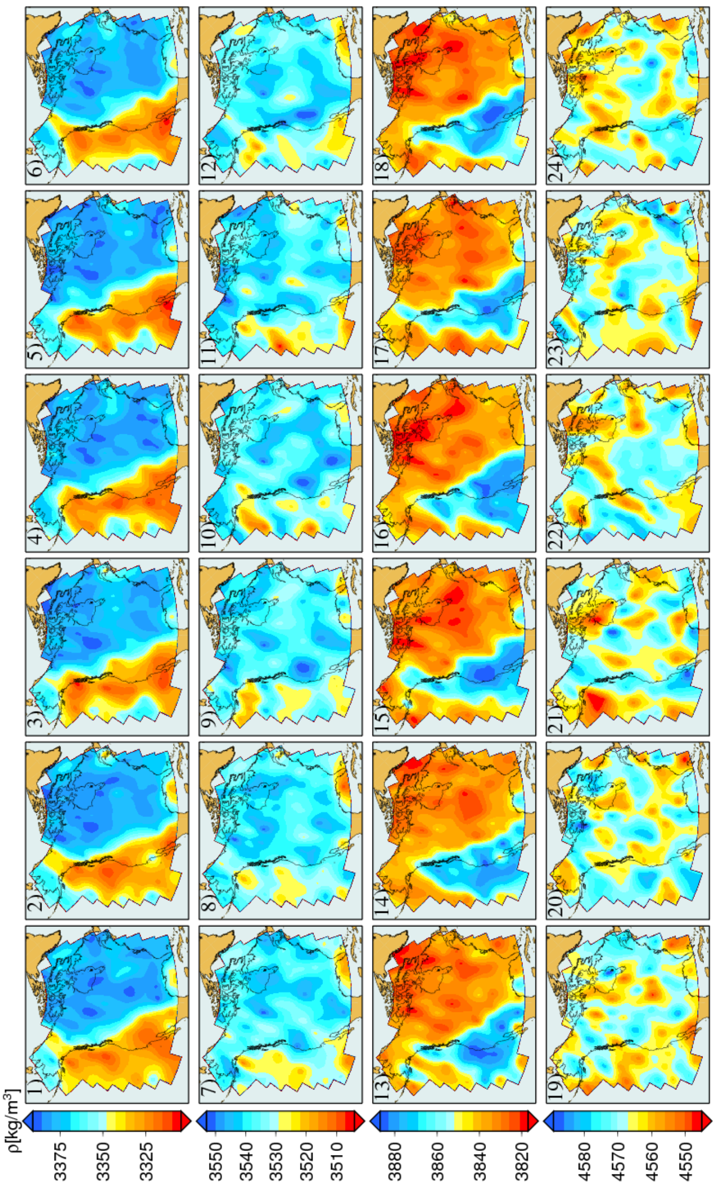
Figure 13. Reconstructed geoid anomalies for the six posterior density models shown in figure S3 (online supporting material) using a continental average viscosity profile (plots 1-6) and regionally averaged viscosity profiles (plots 7-12). Viscosity profiles are shown in online supporting material (figure S4). Only harmonic degrees 6 to 20 are used. For further details see main text. For comparison, plot A shows the observed geoid GGM02 of *Tapley et al.* [2005] for the same harmonic degrees.



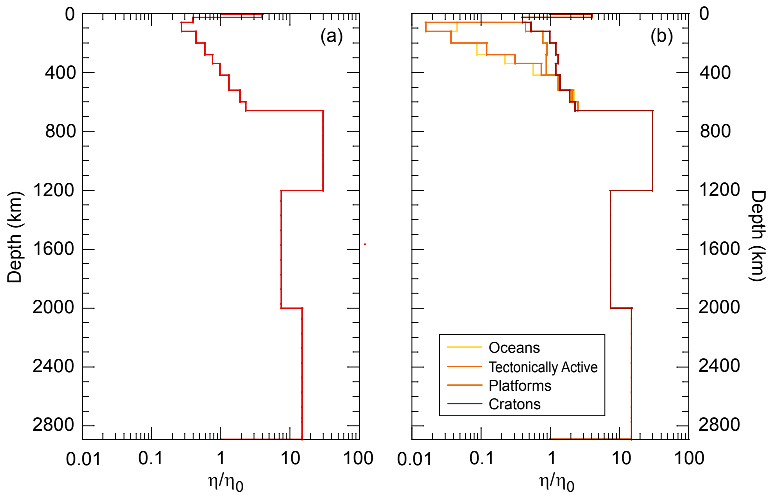
Online supporting material - Figure S1. Mean isotropic shear-wave velocities (plots 5-8) and mean±standard deviation (plots 1-4 and 9-12) computed directly from the posterior probability distributions. Panel 1 is at a depth of 100 km, panel 2 at 300 km depth, panel 3 at 500 km depth and panel 4 at a depth of 1000 km, respectively. For comparison other regional shear-wave velocity models have been included. Plots A (at 100 km depth) and B (at 300 km depth) show the model of *Nettles & Dziewonski* [2008]; plots E (100 km), F (300 km) and G (500 km) depict the model of *Yuan et al.* [2010]. Note differences in absolute shear-wave velocities (colourbars) between present and previous models.



Online supporting material - Figure S2. Posterior isotropic P-wave velocity movie. In each panel the six maps represent six P-wave velocity models that are picked randomly from the posterior distribution at depths of 100 km (1-6), 300 km (7-12), 500 km (13-18) and 1000 km (19-24), respectively.



Online supporting material - Figure S3. Posterior density model movie. In each panel the six maps represent six density models that are picked randomly from the posterior distribution at depths of 100 km (1-6), 300 km (7-12), 500 km (13-18) and 1000 km (19-24), respectively.



Online supporting material - Figure S4. The two viscosity models considered in this study. In (a), the upper mantle viscosity is averaged over the entire study region (continental average), whereas in (b) the models represent averages within four tectonically coherent regions (regional average, see legend).