Fully nonlinear inversion of fundamental mode surface waves for a global crustal model

U. Meier,[1] A. Curtis,[2,3] and J. Trampert[1]

Received 13 June 2007; accepted 24 July 2007; published 18 August 2007.

[1] We use neural networks to find 1-dimensional marginal probability density functions (pdfs) of global crustal parameters. The information content of the full posterior and prior pdfs can quantify the extent to which a parameter is constrained by the data. We inverted fundamental mode Love and Rayleigh wave phase and group velocity maps for pdfs of crustal thickness and independently of vertically averaged crustal shear wave velocity. Using surface wave data with periods $T > 35$ s for phase velocities and $T > 18$ s for group velocities, Moho depth and vertically averaged shear wave velocity of continental crust are well constrained, but vertically averaged shear wave velocity of oceanic crust is not resolvable. The latter is a priori constrained by CRUST2.0. We show that the resulting model allows to compute global crustal corrections for surface wave tomography for periods $T > 50$ s for phase velocities and $T > 60$ s for group velocities. **Citation:** Meier, U., A. Curtis, and J. Trampert (2007), Fully nonlinear inversion of fundamental mode surface waves for a global crustal model, *Geophys. Res. Lett.*, *34*, L16304, doi:10.1029/2007GL030989.

1. Introduction

[2] Nonlinear inverse problems are generally solved using iterated linearized or Monte Carlo, sampling based inversion techniques [e.g., *Mosegaard and Tarantola*, 1995; *Sambridge*, 1999a, 1999b]. *Devilee et al.* [1999] and *Meier et al.* [2007] instead focused on a neural network approach to solve the nonlinear problem of inverting surface wave velocity data for crustal thickness. The latter paper showed how the full posterior (post-inversion) pdf of Moho depth can be found using a Mixture Density Network (MDN). In this paper we demonstrate that the same methodology can be applied to other parameters of general seismological interest. We focus on quantifying the extent to which a parameter is constrained by the data, by evaluating the information content of the full posterior and prior pdfs. Finally we present a global crustal velocity model constrained by surface wave data with corresponding uncertainty statistics which, when combined with the crustal thickness model from *Meier et al.* [2007] represents a global seismological crustal model. Although we evaluate the model on a $2^\circ \times 2^\circ$ grid for convenient comparison with other crustal models such as CRUST2.0 [*Bassin et al.*, 2000] and CUB2 [*Shapiro and Ritzwoller*, 2002], the lateral resolution of our model is that of the combined resolution of the input phase and group velocity maps, ranging between 500 and 1000 km [*Ritzwoller et al.*, 2002; *Trampert and Woodhouse*, 2003].

[3] Crustal structure varies greatly over small length scales and has a first order effect on shear wave velocity and surface waves. In global surface wave tomography the crustal contributions need to be removed in order to access mantle structure. Many different approaches to compute crustal corrections can be found in the literature [e.g., *Woodhouse and Dziewonski*, 1984; *Montagner and Jobert*, 1988; *Ekström and Dziewonski*, 1998; *Boschi and Ekström*, 2002; *Beghein and Trampert*, 2003]. Existing crustal models are based on refraction and reflection seismics as well as receiver function studies. As a consequence, resolution is high in regions with good data coverage but in regions with poor or no data coverage crustal structure is largely extrapolated. Hence, resolution and uncertainty vary greatly as a function of location, and differently to uncertainty in surface wave data. Ideally we should have a global crustal model with a resolution similar to that of the data used in surface wave tomography. We present such a model, and demonstrate that it allows to compute crustal corrections for surface wave phase velocities with periods $T > 50$ s and group velocities with periods $T > 60$ s.

2. Data and Method

[4] The data used by *Meier et al.* [2007] consisted of azimuthally averaged global phase [*Trampert and Woodhouse*, 2003] and group [*Ritzwoller et al.*, 2002] velocity maps, from which dispersion curves were constructed on a $2^\circ \times 2^\circ$ grid globally between 35 and 145 s for phase velocities, 18 and 145 s for Rayleigh group velocities, and 25 and 145 s for Love group velocities. Here we invert the same data set for crustal shear wave velocity using a Mixture Density Network (MDN) that gives a full Bayesian pdf as a solution. A detailed description of the network training method and the prior constraints on each model parameter is given by *Meier et al.* [2007]. They found that as long as the model of velocity with depth is over-parameterized with respect to the resolving power in the data, different parameterizations do not alter the obtained solutions of crustal thickness. We confirm the equivalent result: an over-parameterized model does not introduce any implicit prior information on average shear wave velocity of the crust. Thus we ensure that all significant prior information is explicitly defined by the bounds of variations on model parameters describing velocity structure with depth in the neural network's training data set. We allow the training models to have three crustal layers, and for continental

[1] Department of Earth Sciences, Utrecht University, Utrecht, Netherlands.
[2] Grant Institute of Earth Sciences, School of GeoScience, University of Edinburgh, Edinburgh, UK.
[3] Edinburgh Collaborative of Subsurface Science and Engineering, Edinburgh, UK.

0094-8276/07/2007GL030989$05.00

crust an additional sedimentary layer with variable thickness is added on top. There are 15 mantle layers and variable depths of discontinuities [*Meier et al.*, 2007]. The training set consists of 500'000 continental and 500'000 oceanic models, selected randomly from the prior pdf defined by *Meier et al.* [2007], and corresponding synthetic dispersion curves for each such model.

[5] Surface waves in the period range considered cannot resolve all three crustal layers and the sedimentary layer, which was already noticed by *Shapiro and Ritzwoller* [2002] who inverted a similar data set. Instead of inverting for shear wave velocity in each crustal layer individually, we train the network to invert for the vertically averaged crustal shear wave velocity $\overline{V_s}$, including the sedimentary layer within the average. The MDN is trained to take synthetic dispersion curves for the three layered models as the network input, and to output the Bayesian posterior pdf of $\overline{V_s}$ represented in the corresponding models. It is important to note that $\overline{V_s}$ should not be interpreted literally as the velocity of a 1-layer crustal model. Instead, we show that under certain conditions, the dispersion curves are mathematically equivalent.

[6] Within each of the three crustal layers we impose hard prior bounds on V_s, V_p and ρ, and linearly scale their values by fixing their relative variations. The scaling relations for the crustally averaged properties are then obtained by fitting a line to the variation of depth averaged $\overline{V_s}$, $\overline{V_p}$ and $\overline{\rho}$ in the training data set. For continental crust the following relations were obtained: $\overline{V_p} = 1.5399\overline{V_s} + 840$ m/s, $\overline{\rho} = 0.2277\ \overline{V_s} + 2016$ kg/m^3. As justified below, oceanic $\overline{V_s}$ is fixed a priori to CRUST2.0 and the following scaling relations derived from CRUST2.0 have to be used: $\overline{V_p} = 1.5865\ \overline{V_s} + 844$ m/s, $\overline{\rho} = 0.2547\ \overline{V_s} + 1979$ kg/m^3. It is important to note that these relations reflect the prior information about the average crustal structure that is imposed via hard bounds on V_s, V_p and ρ within each of the individual crustal layers and the sedimentary layer in the training set. In Table 1 we check the validity of our scaling relation by converting the average continental crustal P-wave velocity $\overline{V_p} = 6450 \pm 230$ m/s as given by *Christensen and Mooney* [1995] using: (1) our scaling relation; (2) the scaling relation derived from CRUST2.0 ($\overline{V_p} = 1.6288\ \overline{V_s} + 512$ m/s); (3) a standard ratio of $V_p/V_s = 1.74$; and (4) Brocher's regression fit [*Brocher*, 2005]. From Table 1 we conclude that converting $\overline{V_p}$ to $\overline{V_s}$ with the different scaling relations gives very similar and consistent results within the standard deviations. It is important to note, however, that in converting our $\overline{V_s}$ model to a model of $\overline{V_p}$ and $\overline{\rho}$ our scaling relations have to be used.

3. Information Content of the Data

[7] Using the neural network approach it is straightforward to invert for any combination of parameters which might be constrained by the data (e.g., $\overline{V_s}$, velocity at a fixed depth, vertically integrated travel time to Moho, velocity contrast across the Moho [*Devilee et al.*, 1999]). In the following we demonstrate how to determine the extent to which a specific parameter is constrained by the data in nonlinear problems.

[8] A two stage procedure is common to all Bayesian inversion techniques [*Tarantola and Valette*, 1982]: (1) Define a prior pdf, which represents all the information available from sources independent of the current data (in this case represented by bounds of variations of all model parameters). (2) Combine this prior information with the information contained in the current data set, resulting in the posterior pdf (network training and its application to the data set).

[9] If a parameter is not constrained by the data, no information is gained and the resulting a posteriori state of information equals the a priori state of information. By comparing the information content contained in the prior and posterior pdfs, we can quantify how well a certain parameter is constrained by the current data set. For this purpose we evaluate the information content I of the posterior and prior pdfs [*Tarantola and Valette*, 1982]:

$$I_{prior} = \int_{-\infty}^{+\infty} p_{prior}(m) \log\left(\frac{p_{prior}(m)}{\mu(m)}\right) dm$$
$$I_{post} = \int_{-\infty}^{+\infty} p_{post}(m) \log\left(\frac{p_{post}(m)}{\mu(m)}\right) dm \quad (1)$$

where $\mu(m)$ is the homogeneous pdf ($\mu = 1/V_s$ for velocities and $\mu = constant$ for Moho depth [*Tarantola and Valette*, 1982]). Information is either measured in bits if the logarithmic base is 2 or in nats if the natural logarithm is used. We define the information gained from the data, as $I_{gain} = I_{post} - I_{prior}$, which is equally applicable to linear or nonlinear problems. We evaluate I_{gain} for the 1-dimensional marginal pdf of the parameter of interest.

[10] The prior pdf of Moho depth is uniform between 10–110 km for continental crust and between 0–50 km for oceanic crust respectively. The varying thickness of the sedimentary layer over continents produces a prior pdf of continental $\overline{V_s}$ which is skewed towards lower values. A figure of the continental and the oceanic prior pdf is provided in the auxiliary material.[1]

[11] In Figure 1 histograms of the information gain of Moho depth (Figure 1, top) and $\overline{V_s}$ (Figure 1, middle) are shown over continents (Figure 1, left) and oceans (Figure 1, right) respectively. Note that globally the information gain for Moho depth is between 1.5–3 nats and generally bigger over continents than over oceans (for comparison, if we have two 1-dimensional Gaussians p_1, p_2 with $\mu_1 = \mu_2$ and $\sigma_1 = 2\sigma_2$, the information gain moving from p_1 to p_2 is $I_{gain} = 0.6931$ nats). This indicates that Moho depth is a well constrained parameter. Generally the information gain for $\overline{V_s}$ (Figure 1, middle) is smaller compared to the information gain for Moho depth (Figure 1, top). We further notice that the information gain for continental $\overline{V_s}$ is significant (Figure 1, middle left), whereas the information gain of oceanic $\overline{V_s}$ is negligible (Figure 1, middle right).

[12] There are two possible explanations for the latter observation: (1) oceanic $\overline{V_s}$ is not well constrained by the surface wave data; (2) the oceanic prior pdf of $\overline{V_s}$ is unreasonably narrow. The oceanic prior $\overline{V_s}$ spans only 3320–3440 m/s (see auxiliary material), because there is strong evidence that the average oceanic crust is seismologically very homogeneous and well characterized. To further investigate if oceanic $\overline{V_s}$ can be constrained by the current data, we generated additional oceanic models, with

[1] Auxiliary material data sets are available at ftp://ftp.agu.org/apend/gl/2007gl030989. Other auxiliary material files are in the HTML.

Table 1. Comparison of Different Scaling Relations Between V_s and V_p, Including One Standard Deviation Uncertainty

	P-Wave, [m/s]	S-Wave, [m/s]
Our scaling relation	6450 ± 230	3643 ± 148
Scaling relation derived from CRUST2.0	6450 ± 230	3646 ± 131
Standard scaling relation	6450 ± 230	3707 ± 132
Brocher's regression fit	6450 ± 230	3751 ± 101

an artificial sedimentary layer of variable thickness on top, resulting in a broader prior pdf of $\overline{V_s}$. We then solve the inverse problem for $\overline{V_s}$ using this prior (a table with the explicit prior constraints on the model parameters is given in the auxiliary material). The information gain of the posterior pdf with respect to the broader prior pdf is shown in Figure 1 (bottom right). Because of the broader prior, the information gain is larger than above, but is still small (84% < 0.6931 nats) compared to that on the continent, indicating that the resulting posterior pdfs do not differ much from the prior pdfs. This demonstrates that oceanic $\overline{V_s}$ can not be determined with fundamental mode surface waves in the period range considered. We conclude that given the period range of our data: (1) Moho depth is well constrained globally; (2) $\overline{V_s}$ of the continental crust is constrained but to a lesser extent than Moho depth; (3) oceanic $\overline{V_s}$ is not resolvable, and hence will be constrained a priori to CRUST2.0 in this study.

4. Results

[13] The global model of crustal thickness is discussed in detail by *Meier et al.* [2007] and shown in the auxiliary material. Figure 2 shows the mean of the posterior pdf of $\overline{V_s}$ with corresponding standard deviations. Although this model was not constructed with the purpose of direct interpretation, it is comforting to note a correlation with some known features. The most striking features are the

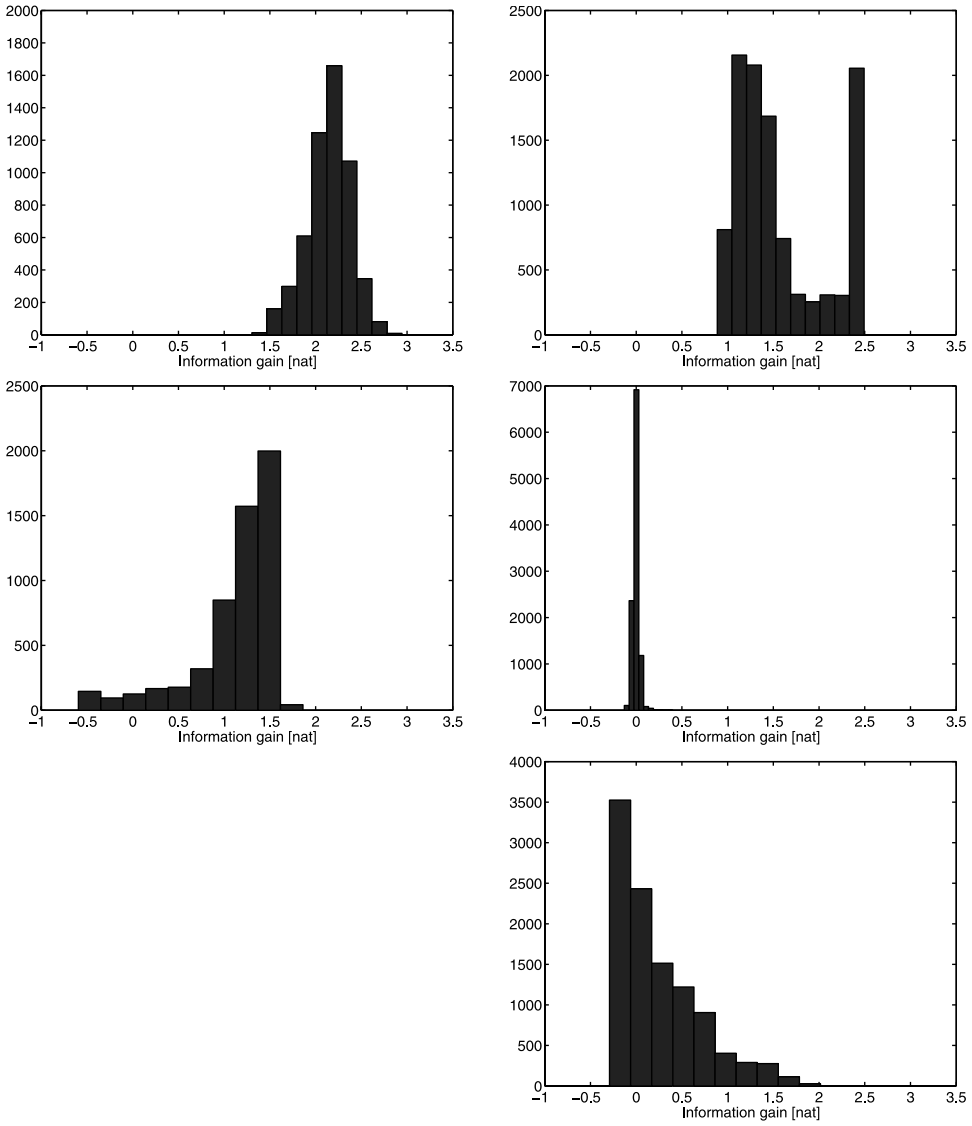

Figure 1. Histograms of the information gain for (top) Moho depth and (middle) $\overline{V_s}$ over (left) continents and (right) oceans. (bottom right) The information gain for $\overline{V_s}$ using the new, broad oceanic prior pdf is also shown.

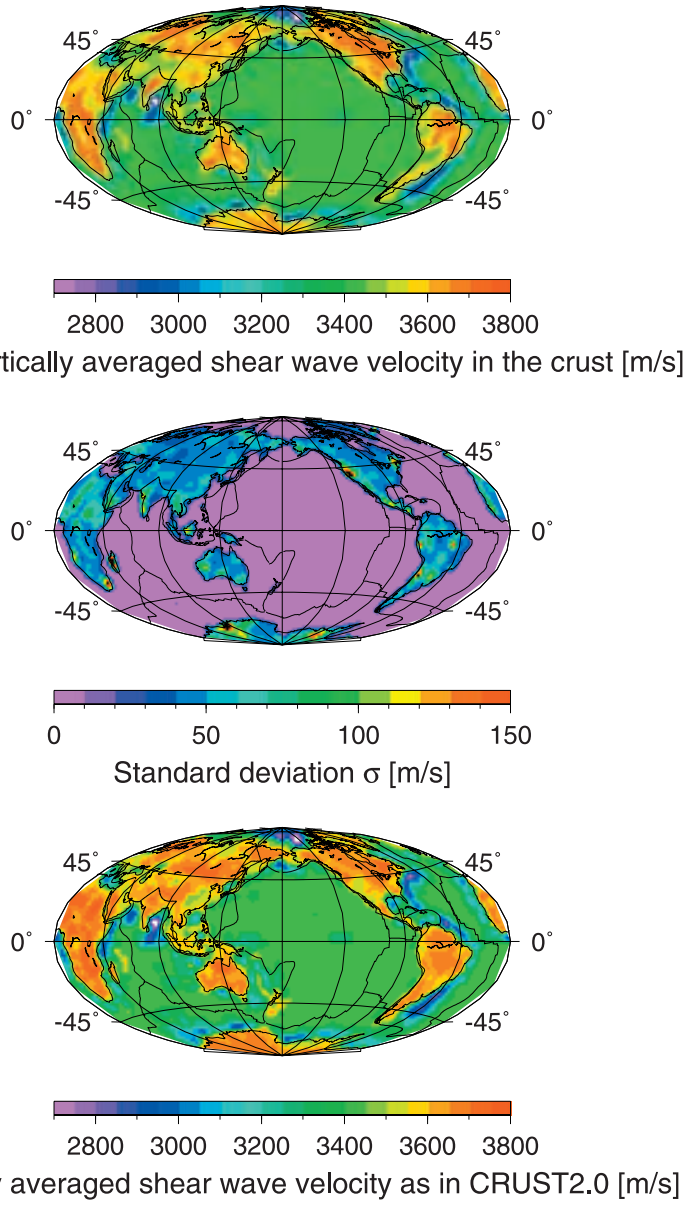

Figure 2. Global map of vertically averaged crustal shear wave velocity extracted from the output of a MDN network. (top) Mean value [m/s]. (middle) Standard deviation σ [m/s]. (bottom) For comparison, vertically averaged crustal shear wave velocity as in CRUST2.0.

high values of $\overline{V_s}$, up to 3800 m/s, in old continental regions such as the back arc of the Rocky Mountains, the cratons of Africa, Australia and the northern part of Eurasia. The high velocity anomaly over India is coincident with the Lava intrusions of the Deccan Trap. The low velocity anomalies in Pakistan and Bangladesh correspond to the huge sedimentary deltas of the Hindus and Ganges river systems respectively. *Mitra et al.* [2006] found similar correlations with geological structure in low period group velocity maps in this region. The main difference between our model and CRUST2.0 are in Africa and the Himalayas where we get lower values. Over most of Eurasia we obtain slightly lower values too, whereas in North America agreement is very good. Comparing the two models one has to keep in mind different lateral resolutions: our model represents averages over areas ranging from 500 to 1000 km whereas CRUST2.0 is based on local estimates which are extrapolated to regions with poor or no data coverage.

[14] The global average continental S-wave velocity is $\overline{V_s} = 3587 \pm 94$ m/s and corresponds to $\overline{V_p} = 6364 \pm 145$ m/s using our scaling relation. This compares well to global average continental P-wave velocity $\overline{V_p} = 6450 \pm 230$ m/s as given by *Christensen and Mooney* [1995] and $\overline{V_p} = 6430 \pm 123$ m/s as in CRUST2.0.

[15] A probabilistic global crustal model consisting of the crustal thickness marginal pdfs from *Meier et al.* [2007] and the marginal pdfs of $\overline{V_s}$ found above is constructed. $\overline{V_p}$ and $\overline{\rho}$ are scaled to $\overline{V_s}$ by the linear relations for continental and oceanic crust defined in section 2.

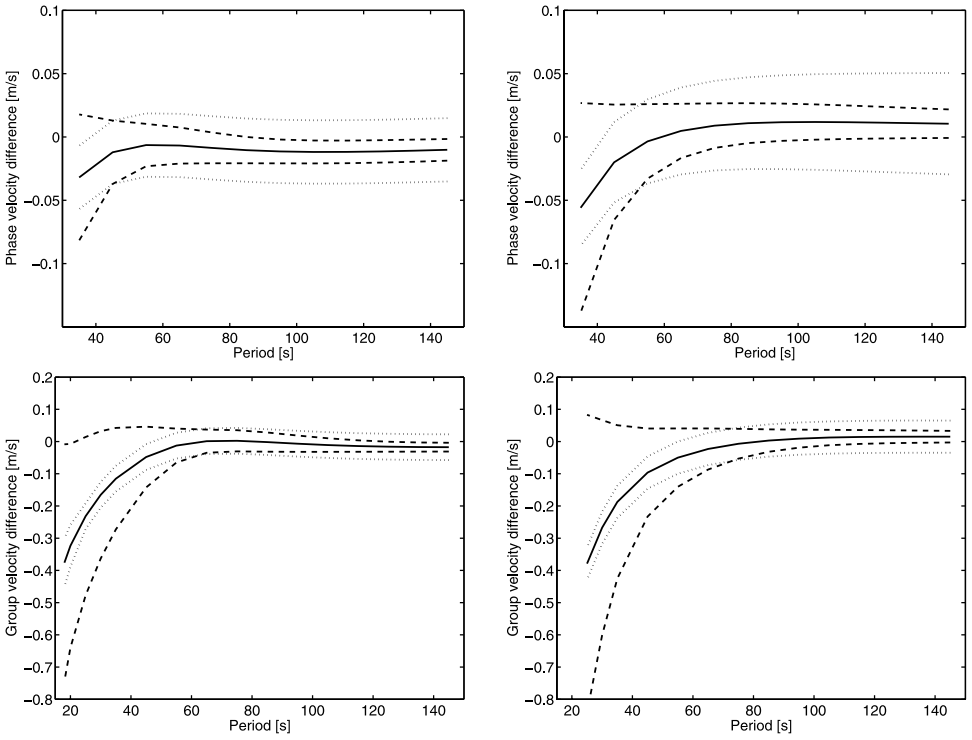

Figure 3. (top) Phase and (bottom) group velocity differences of (left) Rayleigh and (right) Love waves between a 3-layer and a 1-layer crust. Mean differences (solid) with corresponding standard deviations (dashed) of 10000 realizations from the prior pdf; the assumed measurement errors in the data (dotted) are also plotted around the mean difference.

[16] This model can in principle be used to compute crustal corrections if the dispersion response of the 3-layer original crustal model does not differ from the corresponding 1-layer crustal model with vertically averaged properties. We generated 10'000 random realizations from the prior model pdf and computed the corresponding synthetic dispersion curves for the 3-layer and the equivalent 1-layer crustal models. From these synthetic dispersion curves we compute the average difference between the 3-layer and the 1-layer crust, and the corresponding standard deviations as a function of period.

[17] In Figure 3 the average difference (solid) ± one standard deviation (dashed) of phase (Figure 3, top) and group (Figure 3, bottom) velocities of Rayleigh (Figure 3, left) and Love (Figure 3, right) waves are plotted against period. Note that with increasing period the average difference as well as its uncertainty decreases. The difference is more significant for group (Figure 3, bottom) than for phase (Figure 3, top) velocities and to a lesser extent for Love (Figure 3, right) than for Rayleigh (Figure 3, left) waves. We further note that the standard deviation of the difference between a 3-layer and a 1-layer crust becomes smaller than the standard deviation of measurement errors (dotted) given by *Meier et al.* [2007] for phase velocities (Figure 3, top) for periods $T > 50$ s and for group velocities (Figure 3, bottom) for periods $T > 60$ s, confirming that group velocities are more sensitive to crustal structure. Note that in this analysis the Moho depth of models is allowed to vary between 0–110 km; limiting the variation to 0–80 km, for example, would further decrease the difference between a 3-layer and a 1-layer crust (i.e. shifting all significant differences towards lower periods).

5. Concluding Remarks

[18] *Meier et al.* [2007] pointed out various advantages of a MDN over sampling based inversion techniques for solving nonlinear inverse problems. We extended their study and demonstrated that a 1-dimensional posterior marginal pdf for any desired model parameter can be found using the same training set, and hence without the need to re-sample the model space. This posterior pdf embodies all information about a parameter given the data and the prior information. In fully nonlinear problems it is not straightforward to see whether the data provide additional information to the prior. In linear problems, this can be measured by the diagonal of the resolution operator [*Tarantola*, 2005, equation 3.63]. We generalized this concept by introducing the information gain and demonstrated that the data resolve Moho depth everywhere, and resolve velocities beneath continents. We would get the best understanding about the posterior uncertainties by analysing each pdf individually (Mosegaard and Tarantola's movie philosophy [*Mosegaard and Tarantola*, 1995]). However analysing 16200 pdfs by eye is impractical and having checked that the mean, the median and the maximum likelihood point of the pdfs do not differ greatly, we decided to present the solution by the mean of the posterior pdf and its standard deviation. We note that these are not standard deviations estimated from local, linear approximations about the maximum likelihood model. We thus have a global crustal model with the

corresponding fully nonlinearly estimated uncertainties constrained by surface wave data. The model is available in the auxiliary material.

[19] We additionally demonstrated that given the variation of crustal thickness, a single crustal layer is equivalent to a multi-layered crust in terms of their corresponding dispersion responses for periods $T > 50$ s (phase velocities) and $T > 60$ s (group velocities) respectively. This means that the presented global crustal model can be used as a reference model to compute crustal corrections in those period ranges. Since we provide uncertainties, it is possible to quantify the error in the final inversion result due to inaccurate knowledge of our crustal reference model.

[20] **Acknowledgments.** We would like to thank M.H. Ritzwoller for providing the group velocity maps and A. Tarantola for discussions on information content. U. Meier is grateful for the support of the HPC-Europa programme, funded under the European Commission's Research Infrastructures activity of the Structuring the European Research Area programme, contract RII3-CT-2003-506079. Computational resources for this work were provided by the Dutch National Science Foundation under grant NWO:VICI865.03.007 and the Netherlands Research Center for Integrated Solid Earth Science (ISES 3.2.5 High End Scientific Computation Resources).

References

Bassin, C., G. Laske, and G. Masters (2000), The current limits of resolution for surface wave tomography in North America, *Eos Trans. AGU*, *81*(48), Fall Meet. Suppl., Abstract F897.

Beghein, C., and J. Trampert (2003), Probability density functions for radial anisotropy: Implications for the upper 1200 km of the mantle, *Earth Planet. Sci. Lett.*, *217*, 151–162.

Boschi, L., and G. Ekström (2002), New images of the Earth's upper mantle from measurements of surface wave phase velocity anomalies, *J. Geophys. Res.*, *107*(B4), 2059, doi:10.1029/2000JB000059.

Brocher, T. M. (2005), Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. Seismol. Soc. Am.*, *95*(6), 2081–2092.

Christensen, N. I., and W. D. Mooney (1995), Seismic velocity structure and composition of the continental crust: A global view, *J. Geophys. Res.*, *100*(B7), 9761–9788.

Devilee, R., and A. Curtis (1999), An efficient, probabilistic neural network approach to solving inverse problems: Inverting surface wave velocities for Eurasian crustal thickness, *J. Geophys. Res.*, *104*(B12), 28,841–28,857.

Ekström, G., and A. M. Dziewonski (1998), The unique anisotropy of the Pacific upper mantle, *Nature*, *394*, 168–172.

Meier, U., A. Curtis, and J. Trampert (2007), Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.*, *169*, 706–722.

Mitra, S., K. Priestley, V. Gaur, S. Rai, and J. Haines (2006), Variation of Rayleigh wave group velocity dispersion and seismic heterogeneity of the Indian crust and uppermost mantle, *Geophys. J. Int.*, *164*, 88–98.

Montagner, J.-P., and N. Jobert (1988), Vectorial tomography: 2. Application to the Indian Ocean, *Geophys. J.*, *94*, 309–344.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*(B7), 12,431–12,447.

Ritzwoller, M. H., N. M. Shapiro, M. P. Barmin, and A. L. Levshin (2002), Global surface wave diffraction tomography, *J. Geophys. Res.*, *107*(B12), 2335, doi:10.1029/2002JB001777.

Sambridge, M. (1999a), Geophysical inversion with a neighborhood algorithm: 1. Searching a parameter space, *Geophys. J. Int.*, *138*, 479–494.

Sambridge, M. (1999b), Geophysical inversion with a neighborhood algorithm: 2. Appraising the ensemble, *Geophys. J. Int.*, *138*, 727–746.

Shapiro, N. M., and M. H. Ritzwoller (2002), Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.*, *151*, 88–105.

Tarantola, A. (2005), *Inverse Problem Theory*, Siam, Philadelphia, Pa.

Tarantola, A., and B. Valette (1982), Inverse problems = Quest for information, *J. Geophys.*, *50*, 159–170.

Trampert, J., and J. H. Woodhouse (2003), Global anisotropic phase velocity maps for fundamental mode surface waves between 40 and 150 s, *Geophys. J. Int.*, *154*, 154–165.

Woodhouse, J. H., and A. M. Dziewonski (1984), Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, *89*(B7), 5953–5986.

A. Curtis, Grant Institute of Earth Sciences, School of GeoScience, University of Edinburgh, West Mains Road, Edinburgh EH9 3JW, UK. (andrew.curtis@ed.ac.uk)

U. Meier and J. Trampert, Department of Earth Sciences, Utrecht University, Budapestlaan 4, NL-3584 CD Utrecht, Netherlands. (meierue@geo.uu.nl; jeannot@geo.uu.nl)