Geophys. J. Int. (2006) **164,** 394–400 doi: 10.1111/j.1365-246X.2006.02864.x

FAST TRACK PAPER

Surface wave tomography: finite-frequency effects lost in the null space

Jeannot Trampert[1] and Jesper Spetzler[2]

[1]*Department of Earth Sciences, Utrecht University,* PO Box 80021, TA 3058 *Utrecht, the Netherlands. E-mail: jeannot@geo.uu.nl*
[2]*Department of Geotechnology, TU Delft, Mijnbouwstraat* 120, 2600 *GA Delft, the Netherlands*

Accepted 2005 November 8. Received 2005 September 29; in original form 2005 April 22

SUMMARY

We compared surface wave tomography models obtained using finite-frequency kernels and ray theory. We systematically changed regularization in both cases and plotted data misfit against the number of independent parameters in the solution. Our tests show that models from finite-frequency kernels and ray-theoretical kernels are statistically similar. This means that any model obtained using one forward theory can be obtained using the other one by appropriately changing the damping constant. It is clear that finite-frequency theory is a better forward theory to represent the wavefield, but the associated inverse problem is not less ill posed. Indeed, current data coverage is such that the solution is dominated by the chosen regularization. This prevents us from achieving a resolution of the order of the Fresnel zone, or beyond, and noticing the benefits of a better forward theory.

Key words: finite-frequency kernels, inverse theory, model comparison, phase velocity, ray theory, surface wave tomography.

1 INTRODUCTION

Surface waves, fundamental and overtone modes, have been central to understanding the 3-D earth structure ever since tomography started. With the development of automatic measuring techniques and the online availability of large quantities of digital seismograms in well-kept databases, data coverage has grown exponentially in recent years. Accordingly, the degree of detail in the resulting maps has increased. The theoretical foundations have, for simplicity mainly, continued to rely on geometrical ray theory, known to be a high-frequency approximation (Woodhouse 1974). Wang & Dahlen (1995) derived empirically that ray theory breaks down if the scale length of heterogeneity is smaller than the width of the first Fresnel zone. We have clearly reached this limit (Spetzler *et al.* 2001) and we should move beyond ray theory if smaller-scale structure is to be mapped more reliably. A convenient way to include the finite-frequency content of surface waves is to use scattering theory based on the first Born or Rytov approximation (Snieder 1986). The integral along the ray is then replaced by the integral over an influence zone, the shape of which depends on the approximations made (Snieder & Nolet 1987; Yomogida & Aki 1987; Friederich *et al.* 1993; Meier *et al.* 1997; Clévédé *et al.* 2000; Ritzwoller *et al.* 2002; Spetzler *et al.* 2002; Yoshizawa & Kennett 2002, 2005; Zhou *et al.* 2004; de Hoop & van der Hilst 2005). It has been shown that scattering theory is superior to ray theory (Hung *et al.* 2001; Spetzler *et al.* 2005) because it models the finite zone of sensitivity of a travelling wave and accounts for first-order diffraction effects. It is less clear if better physics is a guarantee for better models, a possibility already mentioned by de Hoop & van der Hilst (2005). Indeed, Ritzwoller *et al.* (2002), Yoshizawa & Kennett (2004) and Zhou *et al.* (2005) report improved tomographic models, while Sieminski *et al.* (2004) and Levshin *et al.* (2005) claim that ray theory can produce similar models, at least with reasonable ray coverage and in the presence of noise. This apparent discrepancy is most likely the result of comparing models somewhat arbitrarily. We propose to turn to statistics, which offers several approaches to deal with the difficult problem of model comparisons. To focus the discussion, we require that better models should verify at least several of the following criteria:

(i) to be derived using a better forward theory,
(ii) to achieve a better data fit,
(iii) to be less sensitive to data errors,
(iv) to achieve a better resolution and
(v) to be less dependent on arbitrary choices of regularization.

2 METHOD AND DATA

The exact nature of surface wave dispersion depends on the details of the 3-D earth and the seismic source. In the framework of single scattering, this relation can be written as (Zhou *et al.* 2004):

$$\delta \ln \bar{c}(\omega) = \int_V \left[K^{\alpha}_{3\mathrm{D}}(\mathbf{x}, \omega)\delta \ln \alpha(\mathbf{x}) + K^{\beta}_{3\mathrm{D}}(\mathbf{x}, \omega)\delta \ln \beta(\mathbf{x}) + K^{\rho}_{3\mathrm{D}}(\mathbf{x}, \omega)\delta \ln \rho(\mathbf{x}) \right] d^3\mathbf{x}, \quad (1)$$

where $\delta \ln \bar{c}$ is the average relative phase velocity perturbation at frequency ω measured between a source and a receiver, $K^{\alpha,\beta,\rho}_{3\mathrm{D}}$ is a 3-D sensitivity kernel for relative P-velocity perturbation ($\delta \ln \alpha$),

relative S-velocity perturbation ($\delta \ln \beta$) and relative density perturbations ($\delta \ln \rho$), respectively, and the volumetric integration is over all points $\mathbf{x}$ in the 3-D earth. The shape of $K_{3D}^{\alpha,\beta,\rho}$ depends on mode coupling, the source radiation, directional scattering of the waves and the frequency content of the measurement (Zhou *et al.* 2004). If the kernels are calculated in a spherically symmetric reference model (typical of a first iteration in a non-linear inverse process), forward scattering prevails (Snieder 1988). Further neglecting mode coupling, the 3-D kernels reduce to a 2-D kernel (Zhou *et al.* 2004), and we may write

$$\delta \ln \bar{c}(\omega) = \int_S K_{2D}^{c}(\hat{\mathbf{s}}, \omega)\delta \ln c(\omega)\, d^2\hat{\mathbf{s}}, \tag{2}$$

where $\delta \ln c$ is the local phase velocity perturbation at point $\hat{\mathbf{s}}$ on the unit sphere. Expressions for K_{2D}^{c} can be found in Spetzler *et al.* (2002). The local phase velocity perturbation is of course linearly related to a local 1-D (spherically symmetric) structure by the Fréchet derivatives $K_F^{\alpha,\beta,\rho}$ (e.g. Dahlen & Tromp 1998), which may be inserted in the previous equation to yield

$$\delta \ln \bar{c}(\omega) = \int_S \int_{\text{radius}} K_{2D}^{c}(\hat{\mathbf{s}}, \omega)\big[K_F^{\alpha}(r, \omega)\delta \ln \alpha(\mathbf{x}) + K_F^{\beta}(r, \omega)\delta \ln \beta(\mathbf{x}) + K_F^{\rho}(r, \omega)\delta \ln \rho(\mathbf{x})\big]\, dr\, d^2\hat{\mathbf{s}}. \tag{3}$$

In the high-frequency limit, $K_{2D}^{c} = 1/\Delta$ on the ray path l, where Δ is the epicentral distance. Eq. (3) then reduces to the commonly used expression for surface wave rays:

$$\delta \ln \bar{c}(\omega) = 1/\Delta \int_{\text{ray}} \int_{\text{radius}} \big[K_F^{\alpha}(r, \omega)\delta \ln \alpha(\mathbf{x}) + K_F^{\beta}(r, \omega)\delta \ln \beta(\mathbf{x}) + K_F^{\rho}(r, \omega)\delta \ln \rho(\mathbf{x})\big]\, dr\, dl. \tag{4}$$

In the following, we will compare tomographic models obtained using finite-frequency kernels (eq. 3) and ray theory (eq. 4). The usual approximations are made, neglecting density perturbations and scaling $\delta \ln \alpha$ to $\delta \ln \beta$ using relations from Ritsema & van Heijst (2002). $\delta \ln \beta$ is expanded laterally into spherical harmonics up to degree and order 20 and each coefficient is parametrized on a cubic spline basis (18 knots) along the radius. The parametrization is equivalent to that of model S20RTS (Ritsema *et al.* 1999) and comprises some 8000 unknowns. We used a total of 1.5 million phase velocity measurements (table 1) from fundamental mode Rayleigh waves between 40 and 150 s (Trampert & Woodhouse 1995) and the first five Rayleigh wave overtones between 40 and 130 s (van Heijst & Woodhouse 1999). These data formally provide sensitivity to a depth of at least 1000 km. Here it is important to note that we neglected mode coupling. While this is approximately correct for fundamental modes, coupling is important for overtones, which in all rigour requires 2-D vertical kernels (Li & Tanimoto 1993; Li & Romanowicz 1995) rather than the 1-D kernels K_F we are using throughout this study. As pointed out by Mégnin & Romanowicz (1999), the difference is important for the resolution in the mid-mantle. It is not our purpose to present detailed earth models, but to compare different forward theories. Our conclusions should, therefore, not be affected, since both sets of models are built using the same approximation. All data are corrected for crustal effects using Crust5.1 (Mooney *et al.* 1998). In each $5^\circ x 5^\circ$ cell, the local phase velocity is calculated exactly in a local 1-D model. The relative crustal correction is evaluated from the phase velocities in models consisting of Crust5.1 superimposed on PREM and of PREM alone. The corrections to the measurements are made using eq. (2) where K_{2D}^{c} represents the finite-frequency or ray-theoretical kernel, respectively.

Eqs (3) or (4) result in a linear inverse problem for the coefficient vector $\mathbf{m}$ given the average phase velocity measurements gathered in a data vector $\mathbf{d}$. We minimize the cost function

$$C = (\mathbf{d} - \mathbf{Gm})^t \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{Gm}) + \mathbf{m}^t \mathbf{Dm}, \tag{5}$$

where the superscript (t) denotes matrix transpose. Eq. (5) expresses the competing requirements of data misfit and some chosen quality of the model (e.g. size, roughness). The estimated model is then found by applying a linear operator $\mathbf{L}$ to the data vector:

$$\mathbf{m} = \mathbf{Ld} = \left(\mathbf{G}^t \mathbf{C}_d^{-1}\mathbf{G} + \mathbf{D}\right)^{-1} \mathbf{G}^t \mathbf{C}_d^{-1}\mathbf{d}, \tag{6}$$

where $\mathbf{G}$ is the matrix built from the kernels projected onto the horizontal and vertical basis functions and $\mathbf{C}_d$ is the diagonal data variance matrix, estimated by cluster analyses for similar paths where Gaussian statistics are assumed. It is very difficult to assess whether phase velocity data really follow Gaussian statistics, but again this uncertainty should not affect our conclusions since we are comparing models based on the same assumptions. $\mathbf{D}$ is a regularization matrix of which we considered three end members used individually or as a mixture in various studies. In the case of minimum norm damping, $\mathbf{D} = \lambda \mathbf{I}$, where λ is a constant and $\mathbf{I}$ the identity matrix. We also separately imposed horizontal and vertical smoothing, in which case $\mathbf{D}$ is λ times an operator, which either describes the horizontal gradient on the sphere, or the radial derivative. The quality of the estimated model is characterized by the resolution (e.g. Tarantola 1987)

$$\mathbf{R} = \mathbf{LG} = \left(\mathbf{G}^t \mathbf{C}_d^{-1}\mathbf{G} + \mathbf{D}\right)^{-1} \mathbf{G}^t \mathbf{C}_d^{-1}\mathbf{G} \tag{7}$$

and by the posterior model covariance (e.g. Matsu'ura & Hirata 1982; Tarantola 1987)

$$\mathbf{C}_m = \left(\mathbf{G}^t \mathbf{C}_d^{-1}\mathbf{G} + \mathbf{D}\right)^{-1} = (\mathbf{I} - \mathbf{R})\mathbf{D}^{-1}(\mathbf{I} - \mathbf{R})^t + \mathbf{LC}_d\mathbf{L}^t. \tag{8}$$

Reducing λ (within the limits of stability) in eq. (6) will decrease the data misfit, increase the chosen model quality (eq. 5) and increase the resolution (eq. 7). The number of independent parameters from which $\mathbf{m}$ is built, which is equal to the trace of the resolution matrix (Tarantola 1987), is thus also increased. In general, more complex models will always fit the data better. The question is thus whether a better fit is balanced by the increase in model parameters or if the data really prefer one model over another, in which case we qualify the improved fit as significant. Intuitively, if two models give the same misfit, we tend to choose the simplest (e.g. smallest number of parameters, smallest size, smoothest) model. We refer to this as Occam's razor or the principle of simplest explanation. There are a number of statistical measures, which can be used to determine whether a better data fit is significant. Bayesian inference embodies Occam's razor implicitly (e.g MacKay 2003) and models can be ranked using the Bayesian evidence. In our case, $\mathbf{D}$ is not a covariance operator and hence derivative damping is not a proper prior (normalized probability density function) and the interpretation of the Bayesian evidence is meaningless. Other popular criteria for model comparisons are the Akaike information criterion (AIC) (Akaike 1973) and the Bayesian information criterion (BIC) (Schwarz 1978). AIC and BIC are both based on penalized model likelihoods and, therefore, depend on the correct characterization of data errors. For an increasing number of data points, they will both either favour increasingly complex or increasingly simple models depending on the value of $\det \mathbf{C}_d$. A more conventional statistical answer to this question is given by the F-test (e.g. Bevington & Robinson 1992), which is based on the reduced χ^2 defined by

$$\chi^2 = \frac{1}{N - M}(\mathbf{d} - \mathbf{Gm})^t \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{Gm}), \tag{9}$$

Table 1. Number of Rayleigh wave phase velocity measurements used to construct the models.

Branch	Number of picked frequencies on the branch	Number of measurements on the branch
Fundamental mode	10	755 150
First overtone	15	208 507
Second overtone	12	172 116
Third overtone	8	141 261
Fourth overtone	7	112 196
Fifth overtone	4	59 153
total	56	1 448 383

Table 2. Minimum difference in χ^2 needed to conclude with 99 per cent confidence that 2 models are different. The difference depends on χ^2 itself and the number of free parameters $N - M$.

	$N - M = 10^3$	$N - M = 10^4$	$N - M = 10^5$	$N - M = 10^6$
$\chi^2 = 2.0$	0.156	0.051	0.016	0.005
$\chi^2 = 1.5$	0.117	0.038	0.012	0.004
$\chi^2 = 1.0$	0.078	0.025	0.08	0.003

where N is the total number of data points (Table 1) and $M =$ trace(**R**) is the number of independent parameters in **m**. The ratio of χ^2 from two different models is F distributed. This allows one to calculate the probability whether two χ^2 are significantly different or due to random fluctuations in the data. In Table 2, we show minimum differences in χ^2 to have a 99 per cent probability that this difference is due to a better model. There thus remains a 1 per cent chance that the difference is due to random fluctuations in the data. While this indicative table shows that with an increasing number of data, increasingly smaller changes in χ^2 become significant, the F-test depends less on the exact knowledge of data errors than Bayesian tests because the key number is a ratio of χ^2. Because solutions from scattering theory or ray theory use different **G** matrices, the resulting model will be different even if the same regularization is used. The important quantities to compare are the misfit, of course, and the number of independent parameters in **m**. We performed multiple inversions, varying systematically λ and plotted χ^2 as a function of independent parameters M from which the solution is built. Visual inspection of the curves immediately tells us which model is preferred by the data. Plotting misfit as a function of model norm or smoothness (Yoshizawa & Kennett 2004; Zhou *et al.* 2005) should give the same results, but we chose M because it is a more fundamental parameter of the solution.

3 RESULTS

Fig. 1 shows that models from finite frequency and ray theory are statistically similar, except perhaps for some marginal cases using purely vertical smoothing. The important point, demonstrated by these results, is that one cannot construct a finite-frequency model, which cannot also be obtained from ray theory by changing λ appropriately. This is a confirmation of the conclusions of Sieminski *et al.* (2004) and Levshin *et al.* (2005). Why then have different authors come to opposite conclusions? Ritzwoller *et al.* (2002) compared models using the same regularization. Such models can be identified by close by circles on the trade-off curves. They show that for equal λ, the finite-frequency model is built from less parameters than the corresponding ray-theory model. We further found that the

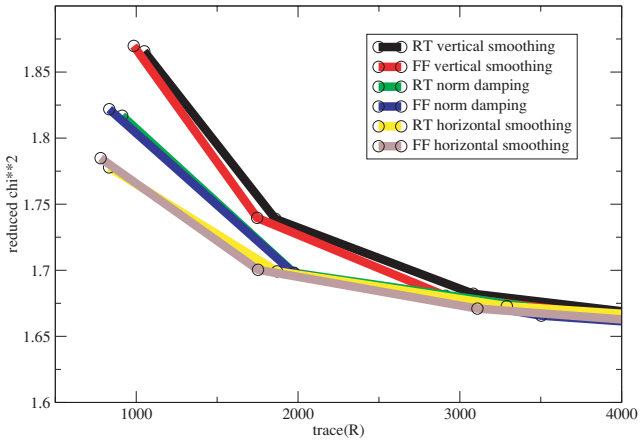

Figure 1. Trade-off curves for inversions using finite frequency and ray theory. Different regularization operators are specified on the figure. The width of the lines indicates the 99 per cent confidence level calculated from the F-distribution (Table 2.). If the difference between two curves at a given $M =$ trace(R) is greater than this width, it is statistically significant. Neighbouring circles indicate equal values of λ.

finite-frequency model norm is always bigger than the ray-theory model norm for the same damping. This explains the observations of Ritzwoller *et al.* (2002). Their results are amplified compared to those shown here, because they used truncated and simplified finite-frequency kernels (Levshin *et al.* 2005). Yoshizawa & Kennett (2004) and Zhou *et al.* (2005) compared different models plotting misfit versus model norm or smoothness curves. Without a proper assessment of the significance of their observed differences, the question of which model is to be preferred remains unanswered. Some readers might argue that significance tests hinge on a particular definition of misfit (eq. 9) or a statistical test. As a more intuitive approach, we simply plot several model vectors against each other (Fig. 2). This reiterates the point that finite-frequency and ray-theoretical models with comparable regularization are similar.

The main difference in the models does not come from the forward theory but from the applied regularization. For high λ, around 1000 independent model parameters, the data clearly favour horizontal smoothing over norm damping and vertical smoothing. Only from 3000 independent model parameters onwards are the solutions independent of the nature of damping (minimum norm versus smoothness). The model norm will of course increase further as λ is reduced. How much we can reduce λ depends on the eigenvalue spectrum of the problem at hand. It is common practice in tomography to show 'nice' (smooth) models, while 'blobby' models are generally suspected to be unstable. Fig. 3 shows models built from 1000 independent parameters using horizontal and vertical smoothing, respectively. We recognize some well-documented features such as strong heterogeneity near the surface with a sharp decline in amplitude below 200 km. Continent–ocean differences are well mapped out, but the depth extent of the structures (e.g. thickness of continents, depth of mid-ocean ridges, mass transfer through the 660 km discontinuity) crucially depends upon regularization. In terms of horizontal versus vertical streakiness, the norm-damped model lies in between those shown in Fig. 3. This model, built with approximately 1000 independent parameters and $\chi^2 = 1.825$, is almost identical to model S20RTS (Ritsema et al. 1999) in the top 1000 km, indicating that S20RTS is heavily norm damped. Because Fig. 1 leads to the inference that the data favour horizontal smoothing, a rash interpretation could be that ridges and continents

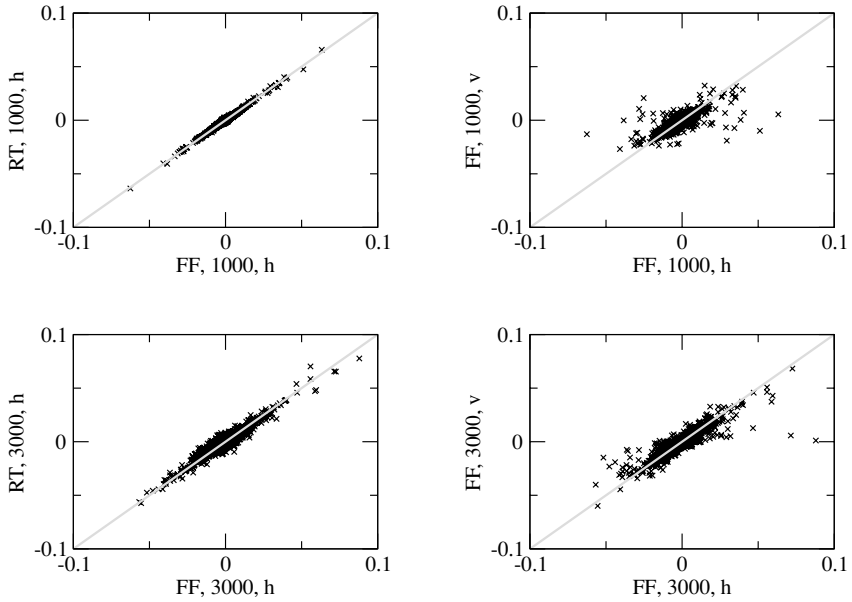

Figure 2. Scatter plots for different models. Plotted are the model coefficient vectors rather than shear velocity at a particular depth. FF are finite-frequency models and RT are the ray-theoretical ones. The number in the legend represents the independent parameters in the solution and h and v stand for vertical and horizontal smoothing, respectively. The only scatter plot with a clear deviation (horizontal trend) from the grey line (slope 1) is the top right one, as already inferred from Fig. 1. The bottom right plot shows some outliers, but retains a clear trend along the grey line.

are shallow features and that we should expect little vertical mass transfer.

In practice, things are unfortunately not as easy. Why should we select models built from 1000 independent parameters out of a possible 8000? Fig. 1 shows that χ^2 does not change significantly beyond 3000 independent model parameters. Applying Occams's razor, we should clearly consider models up to this point, because only beyond $M = 3000$ do models become more complicated without any change in χ^2. A look at the eigenvalue spectra (Fig. 4) shows that they are slowly decaying, but that the first 3000 eigenvalues remain bigger than 1. There is thus no prospect of instability if we use the first 3000 eigenvalues in the solution. In fact this is what is recommended by the analyses of Matsu'ura & Hirata (1982), in agreement with what we would choose using Occam's razor in Fig. 1. Models built from 3000 independent parameters are shown in Fig. 5. The models are now very similar despite different regularization strategies. The rms amplitude of the anomalies has doubled. The fast decay of the amplitude below 200 km has disappeared, suggesting the mantle might contain much more smaller-scale structure with higher amplitude than previously assumed. Such interpretations clearly need further investigation and are beyond the scope of this work, but demonstrate that a subjective choice of regularization has tremendous consequences for our understanding of the Earth's interior, while the forward theory has not. We have only shown vertical slices of models obtained from finite-frequency theory, the slices for ray-theoretical models being similar as demonstrated in Figs 1 and 2. This shows that the dependence of the solution on arbitrary choices of regularization is similar for both theories and one of the main sources of uncertainty in existing models. Recently, we acquired the capability to calculate exact seismograms in 3-D earth models using Beowulf clusters (Komatitsch *et al.* 2002). This will allow us to address this questions more systematically by comparing seismograms from our 3-D models against observed seismograms.

The question of data error propagation is more difficult and has two contributions as can explicitly be seen on the right-hand side of eq. (8), which splits the total uncertainty into a term due to imperfect resolution and a term due to data error propagation. At 1000 independent parameters, on average for any regularization strategy, the total rms uncertainty is 0.2 and 75 per cent comes from imperfect resolution. This is huge compared to rms amplitude of the models, which is around 0.003, and indicates that the valley in the cost function (eq. 5) is locally rather flat. This does not come as a surprise since we found that the solution strongly depends on the regularization (norm, smoothing, etc.). If we just look at data error propagation, $\mathrm{rms}\,(\mathrm{diag}(\mathbf{L}\mathbf{C}_d\mathbf{L}^t)) = 0.05$, which is still big and shows that if we want to keep data error propagation under control with damping, we really have to look at very strongly damped models. This puts us in the awkward situation that the solution depends too heavily on the adopted regularization philosophy. At 3000 independent parameters, the total rms uncertainty has risen to 0.8, but the contributions from imperfect resolution and data error propagation are now equally split. In severely ill posed inverse problems, there can be multiple solutions, even for a linear problem, which casts doubts on derivative techniques. Furthermore, $\mathbf{D}$ is not a proper covariance operator. We, therefore, believe that the interpretation of $\mathbf{C}_m$, or parts of it, is essentially meaningless, and is maybe an excuse why error analysis is notoriously absent in seismic tomography. A promising direction for physically meaningful error analyses is a full model space search by forward modelling (e.g. Trampert *et al.* 2004). This avoids the pitfalls of instabilities in the inverse problem, but comes at a huge computational cost. Returning to our comparisons between finite-frequency and ray-theoretical models, a comparison is more meaningful because both posterior covariances have the same limitations. We find that finite-frequency models systematically show a slightly higher (10 per cent) uncertainty for the full $\mathbf{C}_m$ or $\mathbf{L}\mathbf{C}_d\mathbf{L}^t$ alone. This can also be seen in the eigenvalue spectra (Fig. 4.) which decay slightly faster for finite-frequency theory giving a higher uncertainty (Matsu'ura & Hirata 1982). Overall though, this difference has a minor effect, as can be seen in the left column of Fig. 2. Both sets of models clearly show similar trends,

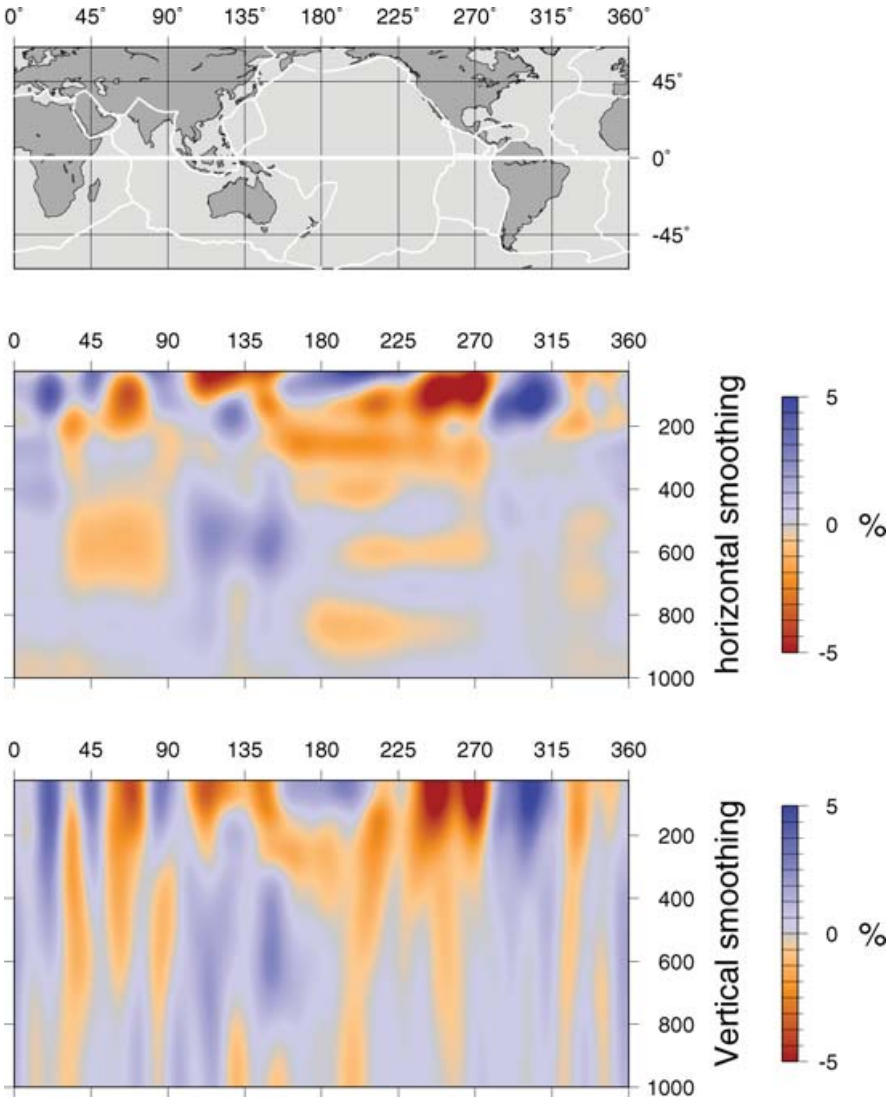

Figure 3. Vertical slice along the equator through the top 1000 km of two models built from 1000 independent parameters using finite-frequency theory and vertical (bottom panel) and horizontal (middle panel) smoothing. Note the fundamental difference in appearance between the two models.

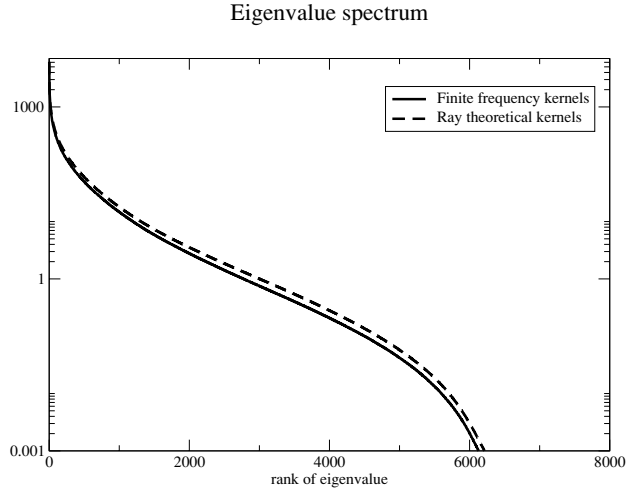

Figure 4. Eigenvalue spectrum of the normalized problem using finite frequency and ray theory. Matsu'ura & Hirata (1982) recommend to normalize the problem to unit data and model covariances and cut off the eigenvalues below 1 to guarantee a solution with optimal uncertainty.

with the scattering in the plots probably due to different model uncertainties.

4 CONCLUDING REMARKS

We have shown that surface wave tomography models obtained from finite-frequency and ray-theoretical kernels are statistically alike, meaning that for every finite-frequency model, we can obtain a similar model using ray theory if we change λ appropriately. To be more specific, how do they perform for each individual criterion mentioned in the introduction? We stated that better models should verify at least several of the following criteria:

(i) to be derived using a better forward theory,
(ii) to achieve a better data fit,
(iii) to be less sensitive to data errors,
(iv) to achieve a better resolution and
(v) to be less dependent on arbitrary choices of regularization.

It is clear that finite-frequency kernels are more accurate [criterion (i)] than ray theory in describing wavefield characteristics in

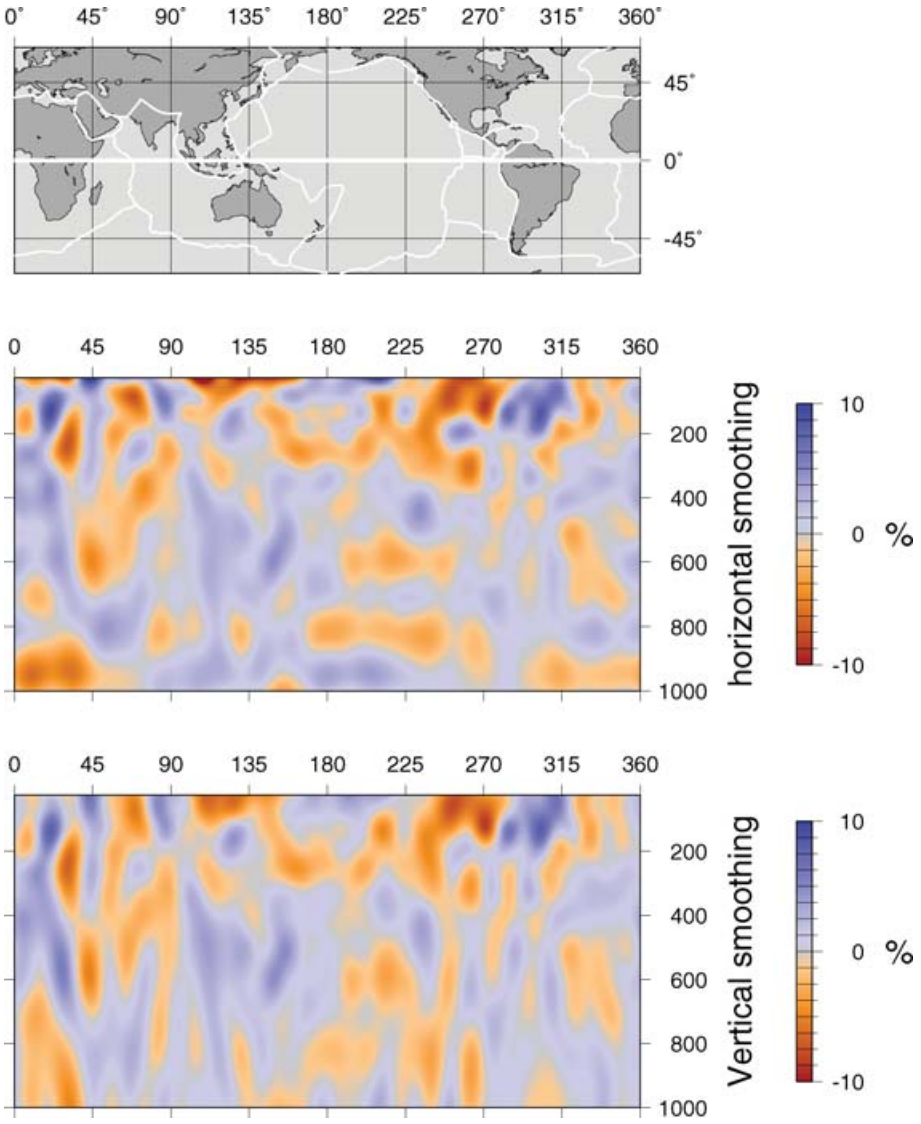

Figure 5. Vertical slice along the equator through the top 1000 km of 2 models built from 3000 independent parameters using vertical (bottom panel) and horizontal (middle panel) smoothing. Note the similar appearance of the two models as predicted by our trade-off curves.

complicated structures (Hung *et al.* 2001; Spetzler *et al.* 2005). Our main result is that changing the damping constant appropriately, the achieved data fit [criterion (ii)] and resolution [criterion (iv)] is similar using either forward theory. Another important finding is that regularization dominates significant aspects of our recovered models and finite-frequency and ray-theoretical models are similarly affected [criterion (v)]. This means that the null space (due to uneven or insufficient data coverage) is still too large to attempt to recover structures of the size of Fresnel zones. To fix ideas, the Fresnel zone width for a fundamental mode surface wave at 100 s propagating over a distance of 90 degrees is about 2000 km. Data error propagation is slightly worse for finite-frequency kernels (criterion iii), but given the large influence of the regularization on the solution, this is a minor problem.

These conclusions, strictly speaking, are only valid for surface wave tomography. We speculate, however, that the situation is not much different for body waves. The reason is that rays, or banana–doughnuts, are spanning the 3-D sphere and the inhomogeneity of the coverage is worse than in the case of surface waves. This means that the null space is at least equal, if not larger for body wave tomography. An equivalent model to that of Montelli *et al.* (2004) can probably be obtained from ray theory alone. In fact, van der Hilst & de Hoop (2005) showed that their model is not as different as initially thought.

Does this mean that finite-frequency kernels are unnecessary? Of course not. It is clear that finite-frequency kernels are more accurate than ray theory and they should be used if data with large frequency variations are combined (de Hoop & van der Hilst 2005) or explicit antispectral leakage operators are employed (Spetzler & Trampert 2003). To increase the resolution of the current models, we have to remove the ill posedness in the inverse problem, which in turn spares us the need of regularization. The only option is to increase or homogenize the data coverage. While we keep trying to persuade the funding agencies that ocean bottom arrays are worth investing in, another possibility has recently emerged. With the ability to calculate accurate seismograms in 3-D earth models (Komatitsch *et al.* 2002), Tromp *et al.* (2005) recently showed how we can use this new tool to increase data coverage. So far, we only use very little information from the complete seismogram (first arrival times of a few

main waves). It becomes now possible to associate each measurable wiggle in the seismogram with its Fréchet derivative, each one sampling the Earth differently. These Fréchet derivatives are nothing else than finite-frequency kernels or generalizations thereof. While they might not fulfil their promise right now, finite-frequency kernels should open the doors to a new generation of tomographic models in the years to come.

ACKNOWLEDGMENTS

We would like to thank Hendrik Jan van Heijst for sharing his overtone data. Roel Snieder and Sergei Lebedev provided thoughtful comments on a preliminary text, which are gratefully incorporated in this manuscript. Andy Jackson helped with numerous discussions on inverse theory. His suggestions and those of Jeroen Tromp improved the text. Barbara Romanowicz, Andrew Curtis and an anonymous reviewer provided constructive comments during the review process.

REFERENCES

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in *Proceedings of the Second International Symposium on Information Theory,* pp. 267–281, eds Petrov, B.N. & Csaki, F., eds, Akadamiai Kiado, Budapest.

Bevington P.R. & Robinson, D.K., 1992. *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, Boston, Massachusetts.

Clévédé, E., Mégnin, C., Romanowicz, B. & Lognonné, P., 2000. Seismic waveform modeling and surface wave tomography, *Phys. Earth planet. Inter.,* **119,** 37–56.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton University Press, Princeton, New Jersey.

Friederich, W., Wielandt, E. & Stange, S., 1993. Multiple forward scattering of surface waves: comparison with an exact solution and Born single-scattering methods, *Geophys. J. Int.,* **112,** 264–275.

de Hoop, M.V. & van der Hilst, R.D., 2005. On sensitivity kernels for 'wave-equation' transmission tomography, *Geophys. J. Int.,* **160,** 621–633.

Hung, S., Dahlen, F. & Nolet, G., 2001. Wavefront healing: a banana-doughnut perspective, *Geophys. J. Int.,* **146,** 289–312.

Komatitsch, D., Ritsema, J. & Tromp, J., 2002. The spectral element method, beowulf computing and global seismology, *Science,* **298,** 1737–1742.

Levshin, A.L., Barmin, M.P., Ritzwoller, M.H. & Trampert, J., 2005. Minor-arc and major-arc global surface wave diffraction tomography, *Phys. Earth planet. Int.,* **149,** 205–223.

Li, X.-D. & Tanimoto, T., 1993. Waveforms of long-period body waves in a slightly aspherical Earth model, *Geophys. J. Int.,* **112,** 92–102.

Li, X.-D. & Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. Int.,* **121,** 695–709.

MacKay, D.J.C., 2003. *Information theory, inference and learning algorithms*, Cambridge University Press, Cambridge, UK.

Matsu'ura, M. & Hirata, N., 1982. General least-squares solutions to quasi-linear inverse problems with a priori information, *J. Phys. Earth,* **30,** 451–468.

Mégnin, C. & Romanowicz, B., 1999. The effect of theoretical formalism and data selection scheme on mantle models derived from waveform tomography, *Geophys. J. Int.,* **138,** 366–380.

Meier, T., Lebedev, S., Nolet, G. & Dahlen, F.A., 1997. Diffraction tomography using multimode surface waves, *J. geophys. Res.,* **102,** 8255–8267.

Mooney, W., Laske, G. & Masters, G., 1998. Crust5.1: a global crustal model at 5degx5deg, *J. geophys. Res.,* **103,** 727–747.

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., Engdahl, E.R. & Hung, S.-H., 2004. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science,* **363,** 338–343.

Ritsema, J. & van Heijst, H.-J., 2002. Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle form P, PP, PPP and PKPab traveltimes, *Geophys. J. Int.,* **149,** 482–489.

Ritsema, J., van Heijst, H.-J. & Woodhouse, J.H., 1999. Complex shear velocity structure imaged beneath Africa and Iceland, *Science,* **286,** 1925–1928.

Ritzwoller, M.H., Shapiro, N.M., Barmin, M.P. & Levshin, A.L., 2002. Global surface wvae diffraction tomography, *J. geophys. Res,* **107,** B12, 2335, doi:10.1029/2002JB001777.

Schwarz, G., 1978. Estimating the dimension of a model, *Annals of Statistics,* **6,** 461–464.

Sieminski, A., Lévêque, J.-J. & Debayle, E., 2004. Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophys. Res. Lett.,* **31,** L24614, doi:10.1029/2004GL021402.

Snieder, R., 1986. 3-D linearized scattering of surface waves and a formalism for surface wave holography, *Geophys. J. R. astr. Soc.,* **84,** 581–605.

Snieder, R., 1988. Large-scale waveform inversions os surface waves for lateral hetrogeneity 1. Theory and numerical examples, *J. geophys. Res.,* **93,** 12 055–12 065.

Snieder, R. & Nolet, G., 1987. Linearized scattering of surface waves on a spherical Earth, *J. Geophys.,* **61,** 55–63.

Spetzler, J., Trampert, J. & Snieder, R., 2001. Are we exceeding the limits of the great circle approximation in surface wave tomography, *Geophys. Res. Lett.,* **28,** 2341–2344.

Spetzler, J., Trampert, J. & Snieder, R., 2002. The effects of scattering in surface wave tomography, *Geophys. J. Int.,* **149,** 755–767.

Spetzler, J. & Trampert, J., 2003. Implementing spectral leakage corrections in global surface wave tomography, *Geophys. J. Int.,* **155,** 532–538.

Spetzler, J., Jocker, J., Smeulders, D. & Trampert, J., 2005. Validation of first-order diffraction theory for the traveltimes and amplitudes of propagating waves, *Geophysics*, under revision.

Tarantola, A., 1987. *Inverse Problem Theory*, Elsevier, Amsterdam, Netherlands.

Trampert, J. & Woodhouse, J.H., 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.,* **122,** 675–690.

Trampert, J., Deschamps, F., Resovsky, J. & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science,* **306,** 853–856.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160,** 195–216.

Van der Hilst, R.D. & De Hoop, M.V., 2005. Banana-doughnut kernels and mantle tomography, *Geophys. J. Int.*, **163,** 956–961.

van Heijst, H.-J. & Woodhouse, J.H., 1999. Global high-resolution phase velocity distributions of overtone and fundamental-mode surface waves determined by mode branch stripping *Geophys. J. Int.,* **137,** 601–620.

Wang, Z. & Dahlen, F.A., 1995. Validity of surface-wave ray theory on a laterally heterogeneous Earth, *Geophys. J. Int.,* **123,** 757–773.

Woodhouse, J.H., 1974. Surface waves in a laterally varying layered structure, *Geophys. J. R. astr. Soc.,* **37,** 461–490.

Yomogida, K. & Aki, K., 1987. Amplitude and phase data inversions for phase velocity anomalies in the Pacific Ocean basin, *Geophys. J. R. astr. Soc.,* **88,** 161–204.

Yoshizawa, K. & Kennett, B.L.N., 2002. Determination of the influence zone for surface wave paths, *Geophys. J. Int.,* **149,** 440–453.

Yoshizawa, K. & Kennett, B.L.N., 2004. Multimode surface wave tomography for the Australian region using a three-step approach incorporating finite frequency effects, *J. Geophys. Res.,* **109,** B02310, doi:10.1029/2002JB002254.

Yoshizawa, K. & Kennett, B.L.N., 2005. Sensitivity kernels for finite-frequency surface waves, *Geophys. J. Int.,* **162,** 910–926.

Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.,* **158,** 142–168.

Zhou, Y., Dahlen, F.A., Nolet, G. & Laske, G., 2005. Finite-frequency effects in global surface-wave tomography, *Geophys. J. Int.*, **163,** 1087–1111.