Geophys. J. Int. (2003) ***,** 000–000

Reply to comment by F.A. Dahlen and G. Nolet on: "On sensitivity kernels for 'wave-equation' transmission tomography"

Robert D. van der Hilst[1] and Maarten V. de Hoop[1,2]

[1]*Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Rm 54-522, Cambridge, MA 02139, USA.*

[2]*Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA.*

SUMMARY

We thank Dahlen and Nolet (2005) – hereinafter DN05 – for the comments on our paper (De Hoop and Van der Hilst (2005); hereinafter HH05) and for giving us the opportunity to elaborate our points. In some cases we agree with DN05, but we respectfully continue to differ in opinion on fundamental aspects of the finite frequency sensitivity kernels known as 'banana doughnut' kernels (Dahlen et al., 2000) – hereinafter BDKs, as per their original nomenclature – and on the significance of their effect on the tomographic models presented so far. DN05 dismiss HH05's analysis as a "tale ... full of sound and fury, signifying nothing". We are not sure of the meaning of this in scientific dialog, but the quote is interesting *, and we invite the reader to ponder its merit in view of the observations made below. HH05

*For those less familiar with Shakespeare's play "Macbeth", the full quote is: 'It is a tale told by an idiot, full of sound and fury, signifying nothing'. This speech was made by Macbeth as he prepares for battle and ridicules the warnings of impending doom. Moments later he is told that Birnam Wood has started to move and is advancing on his troops. 'Liar and slave!' he cries at the messenger who brings him this news. But the same day, all the prophecies are fulfilled, and Macbeth is defeated and killed.

do not comment on the models by Montelli *et al.* (2004a; 2004b), hereinafter M04a,b, but in response to DN05, and for readers interested in the implications for the images, we take this opportunity to discuss some salient issues before returning to the more technical matters. We refer to ray theory (that is, the high frequency approximation and the use of 'ray paths') as RT and finite frequency theory (with BDKs) as FFT and the Princeton models, which were gracefully made available to us by Raffaella Montelli, as PRI-RT/PRI-FFT.

Key words: transmission tomography – finite-frequency sensitivity kernels – plumes

1 ARE THE EFFECTS OF BDKS ON THE TOMOGRAPHIC IMAGES SIGNIFICANT?

HH05 remark that the effects of BDKs on global tomographic images are small compared to effects of the damping (regularization) that is needed to mitigate effects of uneven data coverage and quality. DN05 take issue with this and claim that (i) "[BDKs] enable improved imaging of small-scale wavespeed anomalies", such as 'plumes', and (quoting M04a) that (ii) "depending on the depth and size of the anomaly, the amplitudes of the velocity perturbations in [PRI-FFT] are $30-50\%$ larger than in [PRI-RT]". To demonstrate that these claims of image improvement are overstated we compare the RT and FFT models published by M04a and M04b, respectively, and model MIT-P05 by Van der Hilst, Li, and Kárason (in preparation). The latter is based on a data set that is comparable to that used by M04b, with the pertinent exception that we did not use the long period P data from Bolton and Masters (2001).

1.1 Resolution of slab structure

We first assess the effect of FFT on the imaging of mantle structures that have not been disputed. Figure 1 depicts slabs of subducted lithosphere below South America according to MIT-P05 (left), PRI-FFT (middle), and PRI-RT (right). It appears that the RT and FFT images are virtually the same (despite the incorporation of short period data from Engdahl *et al.* (1998) in the latter), but both lack the upper mantle structures revealed in MIT-P05. M04b state that "slabs are large features that cause little or no wavefront healing". This may be so, but BDKs are calculated in a smoothly varying background medium and have no knowledge about either the level of wavefront healing in the data to which they are applied or the scales and types of heterogeneity (such as, plumes, slabs) through which they pass. Besides, if the large

FIGURE 1 GOES HERE.

Figure 1.

size of slabs is used to explain that there is no 'wavefront healing', and, by implication, that the data have not 'lost' information about these structures, then why are they not resolved better?

Figure 1 suggests that (i) the short period data used in PRI-FFT were given much lower weight than the long period data, that (ii) the long period data do not contain information about length scales pertinent to slab structure in the upper mantle (but, then, how about 'plumes'?), or – most likely – that (iii) the specific model parameterization and regularization (damping) used by M04a,b does not allow better resolution of slabs. The latter can be fixed, but it is – in our view – not satisfactory that much confidence is placed in the images of elusive 'plumes' while rather incontroversial structures such as subducted slabs (which, for several reasons, should be easier to image) are not well constrained.

1.2 How about plumes?

To stay closer to the topic of Montelli and co-workers, figure 2 illustrates the effects of the use of BDKs for 'plume' structures in the upper and in the deep mantle. Since small is in the eye of the beholder we invite the readers to judge for themselves the significance of the differences.

For comparison, figure 3 depicts the lateral variation in P-wavespeed according to MIT-P05. Apart from differences in amplitude that arise from different choices of parameterization, damping, and data fit criterion (see caption to figure 1), there are two observations that are relevant for the discussion here. First, compared to figures 2a,b, plume-like structures in the Indian Ocean are conspicuously absent in figure 3a. This does not necessarily mean that they do not exist. Owing to the differential nature of the

FIGURE 2 GOES HERE.

Figure 2.

FIGURE 3 GOES HERE.

Figure 3.

data and the behavior of the sensitivity kernels of the long period $PP - P$ data [†] the sensitivity to shallow structure beneath source and receiver is strongly reduced, which degrades our ability to resolve these upper mantle structures. This difference between the models is most likely due to the use of the long-period P data by M04. Second, in support of M04b's interpretation in terms of deep mantle plumes, MIT-P05 also reveals a low velocity anomaly beneath Hawaii (figure 3b). In MIT-P05 this anomaly is located to the west of Hawaii and agrees well with the westernmost part of the split anomaly in the RT and FFT models (figures 2c,d).

[†]We also use 3D kernels to combine long period $PP - P$ (and $P_{diff} - PKP$, but that is not of interest here) with the short period P and pP data (Engdahl et al., 1998) – see Van der Hilst and Kárason (2001) and Kárason (2002). They found that 'fat rays' and BDK-like kernels produced similar images (Kárason, 2002).

FIGURE 4 GOES HERE.

Figure 4.

The similarity of the images according to PRI-RT (figures 2a,c) and PRI-FFT (figures 2b,d) suggests that the plume-like structures do not appear (or are better imaged) because of FFT. By itself, this does not invalidate M04b's interpretations. Yet, there is reason for caution. Figure 4 suggests that, for this geographical region, (i) the plume signatures show up only (always?) below the ocean island stations that contribute to the set of long period P data and that (ii) at 450 km depth the image of these 'plumes' is virtually the same as at 150 km depth (2a,b). These observations suggest that for these upper mantle 'plumes' preferred sampling along steep rays (or BDKs) in poorly sampled regions is an issue. This is a classical resolution problem, but the perception of continuity in radial direction may be enhanced by M04's use of an isotropic gradient minimization, which may also explain the large thickness of some continental keels (e.g., beneath South Africa).

1.3 What about the magnitude of wavespeed perturbations?

In the examples given above, both the spatial pattern and the amplitudes of the RT and FFT wavespeed variations are remarkably similar. How can this be reconciled with DN05's claim of a 30-50% increase in amplitude resulting from the use of BDKs? The answer may be surprisingly simple.

We understand that M04a – and DN05 – infer the amplitude increase from histograms of the ratio of wavespeed variations in PRI-FFT and PRI-RT ($\delta c_{\rm FF}/\delta c_{\rm RT}$) "for those anomalies that are significantly different from zero ($|\delta c/c| > 0.2\%$)". Indeed, M04a use histograms like the one reproduced in figure 5a to suggest that the average ratio is significantly larger than 1: they report 1.5 for the lowermost mantle and 1.3 near Earth's surface (hence an increase of 50% and 30%, respectively). Of course, taking a bite out of one side of a (statistical) distribution biases the average in the opposite direction. In fact, in a

similar fashion one could plot the inverse ratio to argue that RT inversions produce the larger amplitudes! Furthermore, the magnitude of the bias depends critically on the value of the minimum cut-off and the treatment of (very large) outliers. Clearly, the average is an incorrect (and misleading) measure for such distributions. In fact, the mode of $\sim$1 suggests that $\delta c_{\rm FF} \sim \delta c_{\rm RT}$ in all examples shown in M04a.

To shed further light on this, we performed a simple analysis of the FFT and RT models [‡]. Including all values produces normal distributions with a mean ratio of $\sim$1 (Figure 5b). Excluding the small values does indeed increase the average $\delta c_{\rm FF}/\delta c_{\rm RT}$ (by about 15%, but the precise value is not important) (5c). Ironically, this is also true for the ratio $\delta c_{\rm RT}/\delta c_{\rm FF}$ (5d), but we do not suggest that RT amplitudes exceed the corresponding FFT ones. Evaluation of the relative behavior of $\delta c_{\rm RT}$ and $\delta c_{\rm FF}$ with scatter plots (figure 5e) indicates that (i) the correlation between $\delta c_{\rm FFT}$ and $\delta c_{\rm RT}$ is not significantly different from one, that (ii) the large anomalies are well correlated (and are associated with relatively small spatial features because regridding at 0.5 by 0.5 deg would otherwise have produced many more points), and that (iii) large ratios only occur close to zero and are thus associated with anomalies that could reasonably be regarded as in the noise. The latter becomes evident when we plot the (RMS) magnitude of wavespeed perturbations as a function of the ratio between FFT and RT anomalies (figure 5f): for all depths the strongest anomalies produce ratios close to one whereas the negative and large positive ratios are associated with much weaker anomalies. We find that in the top $\sim$2,000 km the FFT and RT anomalies are statistically similar; the discrepancy is larger in the lowermost mantle but does not exceed 10% (that is, 0.1 of perturbations of the order of 0.5%).

Unless M04a used other evidence, the amplitude increase resulting from the use of FFT is (much) smaller than they claim. And we argue that the differences that do occur are insignificant. For instance, M04a use *Occam's razor* to select 'simple' models, but another perception of 'simple' may yield different amplitudes. Indeed, by adjusting the damping M04a can produce RT models that are (also in their own view) similar to the FFT results, but they chose to compare models with the same χ^2. In the absence of precise knowledge about data error the χ^2-criterion is difficult to implement objectively, however. Uncertainties about data error are substantial; for example, M04a remove an off-set in the long period P travel times of up to $\sim$5 s without – by their own admission – knowing the origin of it. With intrinsic uncertainties this large the χ^2–criterion is no less subjective than regularization (damping).

[‡] The FFT model used here – after M04b – may slightly differ from the FFT model that was used by M04a in the comparisons with their RT results. Moreover, we used regridded models. This may give minor differences compared to results presented by M04a, but that is inconsequential for the points we make here.

FIGURE 5 GOES HERE.

Figure 5.

1.4 The Banana Doughnut paradox

The above observations support – and reinforce – HH05's remark that differences induced by BDKs are not significant in view of image uncertainty due to uneven data coverage, variable data quality, and the effects of practical considerations or choice of different inversion strategies. These include parameterization, damping, normalizaton of sensitivity within projected kernels, choice of data error when a χ^2-criterion is used for model selection, and the weighting of different data sets. We note that similar conclusions can be reached for surface wave tomography (Trampert & Spetzler, 2005).

The similarity of the RT and FFT results (illustrated in figures 1, 2, 4, or Fig. 9 of M04a, for that matter) leads to a paradox: If, on the one hand, the RT inversion maps anomalies on unperturbed rays and if, on the other hand, BDKs have negligible sensitivity to anomalies on those rays, then why (barring fortuitous source-receiver distributions) does the FFT inversion reconstruct – within error – heterogeneity at the same location and with practically the same spatial dimensions and amplitudes as the RT inversion? In fact, in most parts of the model, the projection of BDKs onto basis functions (blocks, splines, voronoi cells, ...) would not preserve the doughnut hole but produce a 'fat ray' kernel with negative sensitivity on the ray (see HH05, their Fig. 5a); the effective kernels are then rather similar to the kernels used in the RT inversions. Furthermore, the sensitivity on the ray is not negligible for scales represented in the data. We elaborate and generalize this statement in the next section.

FIGURE 6 GOES HERE.

Figure 6.

2 MULTI-SCALE ASPECTS

We concur with Dahlen *et al*. that finite-frequency waves sense structure (also) off the geometrical ray and that scattering must be considered to image heterogeneity with length scales that are small compared to the seismic wavelengths. Fundamentally, however, data (only) resolve structure at spatial scales that are represented in them. The ultimate consequences of this are important and call for a different approach toward finite frequency imaging than defended by DN05.

In the linearized framework of wave equation tomography, with the Green's functions calculated in a smooth background, small structures will be 'invisible' if the data lack the relevant frequencies (through diffraction), and this is true for anomalies on the ray and away from it. For structure on the ray this is consistent with BDKs for long-period data. However, small anomalies placed within the high-sensitivity ring around the 'doughnut hole' should not be visible either if the data lack the relevant (high) frequencies. The subtle implication for tomography is that, upon back-projection along long-period BDKs, the kernel intersections may yield anomalies that are smaller in size than the scales represented in the data. This small-scale structure is not (and cannot be) resolved by the data used. (NB in the classical Radon concept this paradoxical situation does not occur because the intersection of crossing lines is a point and infinite frequency data will be sensitive to that). Furthermore, the sensitivity does not vanish on the unperturbed ray for scales represented in the data. In a space-scale sense this is consistent with the 'fat man' notion of Hung *et al.* (2001) that if you place a large enough anomaly *on the ray* the data would be sensitive to it.

These issues (and the banana-doughnut paradox) signal the need for a basis – or frame – that properly matches the multi-scale aspects of finite frequency wave behavior. HH05 introduced an explicit

multi-resolution analysis based on *curvelets,* which provide a frame with the correct properties for finite-frequency wave equation tomography. They showed (see figure 6) that: (1) data resolve variations in structure primarily in directions perpendicular to the ray; (2) finite bandwidth data resolve only certain scales (that is, the number of annuli, or scales, that contribute to the sensitivity is limited); (3) sensitivity *on the ray* is not zero for scales represented in data, and that (4) the higher the frequency, the smaller the length scale of the structures that can be resolved. Figure 6 demonstrates that the resolution of structure at a certain scale is the same whether it is located on the unperturbed ray (i.e., in the 'doughnut hole') or away from it (e.g., on the BDK ring of maximum sensitivity). In other words, waves with a certain frequency content are (in)sensitive to certain scales whether they occur on the ray or away from it. A detailed analysis of the multi-resolution concept for wave equation tomography is beyond the scope of this "Reply" and will be presented elsewhere (Douma, De Hoop, and Van der Hilst, in preparation).

3 MEASUREMENT

We agree with Dahlen *et al.* (2000) that the type of measurement should dictate which approach to (approximating) propagating waves should be used [§]. We also agree that – in principle – one should not use ray theory to back project 'finite-frequency travel times' measured by time domain cross correlation. HH05 does not, however, state that the cross-correlation criterion approach to transmission tomography and the associated optimization scheme, as used by Dahlen *et al.* (2000), are incorrect. The cross-correlation/optimum criterion is, indeed, a 'stable' criterion to compare the 'distance' between two wavefields. But it is just a criterion, and it is only tied to the notion of travel time proper in the asymptotic broad-band limit of two delta waves (possibly convolved with a source or station signature). This is analyzed in detail in Hörmann and De Hoop (2002). One can give any name to the maximum of the cross correlation, such as 'finite-frequency travel time shift', but frequency dependent travel-time shifts have little meaning outside the context of an explicit multi-resolution, time-frequency analysis.

[§] In this regard Dahlen *et al.*'s comparisons of RT and FFT through resolution tests with synthetic data are somewhat misleading. For a proper discussion of the relative merits of FFT over RT one should use data consistent with RT in the RT inversions and data consistent with FFT in the case of BDK inversions. M04a test the RT inversions on synthetic delay times that are consistent with the FFT cross correlation. These data do not contain the scales assumed in RT. Vice versa, FFT inversion of data produced by RT may well yield results that are (slightly) inferior to results of the RT inversion.

4 THEORETICAL CONSIDERATIONS

The basic theoretical observations of HH05 are that (i) the kernels induced by the cross-correlation criterion are oscillatory (no surprise), the precise pattern of their zero crossings is determined by the background medium but does not matter, and a zero crossing needs not coincide with the unperturbed ray paths, and that (ii) one should not view the kernels point-wise (Strichartz, 2003) but in a multi-resolution framework (see section 2).

4.1 Ray kernel as an infinite bandwidth realization of finite frequency (wave) kernel

We assume that (i) the velocity model is smooth and that (ii) the sources can be represented by point sources and, hence, that the wavefield (that is, the displacement u at receiver r) can be written as a time convolution of a source signature $W_m(t)$ with a Green's function G_r (that is, $u_r = W_m \overset{(t)}{*} G_r$). We do not claim that our analysis applies to frequencies arbitrarily close to zero.

The key ingredient in the kernel development is the Green's function in the unperturbed model. The Green's function can be represented by oscillatory integrals (OIs) – see Duistermaat (1996) and Dencker (1982) for the scalar wave and elastic wave case, respectively. The leading-order amplitude in the OI yields a Maslov (that is, asymptotic) representation of the Green's function. This is warranted under a 'far-field' approximation (NB. the analysis and arguments presented in HH05 hold for general OIs, but to keep the exposition transparent HH05 used only the simplest form and left out the higher order contributions – HH05 eq. 1 then reduces to HH05 eq. 2.)

Starting from the Green's functions representation of the kernel, HH05 show that the kernel tends to the travel time tomography kernel as the frequency (bandwidth) becomes infinite. Conversely, the band-limited Maslov representations of the Green's functions in the kernel capture its leading finite-frequency features. In this sense, 'travel time' and 'wave-equation' tomography do indeed share the same kernel. This is consistent with the observation that asymptotically (that is, 'delta waves') the maximum of the cross correlation yields the travel time shift. Indeed, in the multi-resolution setting mentioned in section 2 the behavior of the kernel can be described correctly – using the multi-resolution representation of the Green's functions – all the way from the narrow-band situation to the infinite bandwidth limit.

DN05 seem confused about the notion of distribution and regularization. Separating out source and receiver signatures, the exact kernel would be a distribution built from Green's functions, which are (Lagrangian) distributions themselves. The approximate, finite bandwidth kernel is a regularization of this distribution (through the instrument response and source signature); this regularization is smooth

but not a distribution itself. The full distribution described by its regularization should not be considered point-wise, however (Strichartz, 2003).

Finally, the singular support of the kernel coincides with the unperturbed ray. To analyze what happens on this ray (for example, upon regularization with a source signature) it suffices to focus on the most singular part of the kernel, which can be derived from the Maslov representations of the Green's functions. This notion allows the development of fast finite frequency methods, even for complex media.

4.2 Caustics and kernel complexity

Any statement relating to the behavior of a kernel of a (locally linearized) inversion procedure should be generic. To tie the behavior of a kernel to a particular, simple medium is in our opinion meaningless. The breakthrough needed in (finite frequency) tomography is not the linearization with approximate kernels calculated for quasi-homogeneous media but the proper accounting for broad-band wave interaction with heterogeneity at a wide range of length scales. DN05 and HH05 concur that caustics affect the locations of the zero crossings in the point-wise evaluation of the kernels. What should be realized, however, is that caustics form readily and ubiquitously due to smooth heterogeneities in the medium (see White; cf. HH05), and can thus be expected to occur throughout Earth's mantle (both between the source and scatter point and between the receiver and the scatterer). The fact that most are not observable at the surface does not mean that they can be ignored in the theories and kernel evaluations under consideration. Accounting only for the PP caustic is, therefore, not sufficient. In this context we argue that the use of BDKs will merely give a false sense of improvement when imaging complex media, such as the upper mantle, volcano interiors, the shallow subsurface, etc., because caustics will be galore and the actual 'kernels' may not even remotely resemble BDK features.

4.3 Errors in source signature

We agree that this will be a relatively small effect (but we did not ignore the error in origin time: this is captured by the difference between W_d and W_m). We also agree that, if the (error in) source signature is known a deconvolution can remove this effect altogether. If the parameter γ in $u_r[\gamma]$ would appear both in W_m and G_r, then HH05 eq. (12) would change (using HH05 eq. (9)) into a form coinciding with the leading two terms of DN05 eq. (2) with $\varepsilon = \delta W_m \overset{(t)}{*} G_r$ or $(W_m - W_d) * G_r$. With $W_m(t) = W_d(t - \delta t_0)$ the second term does indeed reduce to the error in origin time. But the linearization is not obvious because we do not know how to compare quantitatively the error in shape of the signature in $\delta W_m \overset{(t)}{*} G_r$, in

general, with the error in G_r due to an error in velocity model in $W_m \overset{(t)}{*} \delta G_r$. If there is no error in origin time but an error in shape, the second term will not vanish. In principle, one could try a 'joint inversion' for signature (source) and velocities (medium), but that was not the point we wanted to make.

5 DISCUSSION AND OUTLOOK

The use of finite frequency sensitivity kernels in (travel-time) tomography is not new (see, for instance, Luo and Schuster (1991) and Woodward(1992) for shalow sub-surface imaging and Li and Romanowicz (1995) for applications to global tomography), but the theoretical work of Dahlen *et al.* and the application by M04b have brought it to the attention of a broad community. Even though DN05 responded with significant fury, HH05 concur that new approaches are needed for the proper interpretation of finite frequency data. We point out, however, that the evaluation - in simple media - and interpretation of the so called banana doughnut kernels have serious limitations for broad-band signal analysis and tomography. Based on the material presented here (and our experience with tomography), we also stand by our view that the differences produced by BDKs are small (compared to image uncertainty due to, for instance, uneven data coverage and quality, model parameterization, damping) and, by implication, that M04a,b – and DN05 – overstate their significance. That does not mean that the tomographic models are incorrect, but the examples given here may inspire a re-assessment of the structures labeled as plumes, unclouded by unwarranted claims about BDKs. We hope – and expect – that such an exercise reinforces much of the interpretations by Montelli et al. (2004b), but some revision may be necessary.

We fully agree with DN05 that the use of Green's functions in transmission tomography has distinct advantages over travel times: if done properly, (i) it allows correct joint interpretation of data measured in different ways and with different sensitivities to Earth's structure, (ii) it enables wave propagation analysis and kernel evaluation in models with more realistic, less smooth heterogeneity, (iii) it helps mitigate effects of uneven data coverage (by exploiting the additional scale-orientation information – see section 2 – one can obtain comparable spatial resolution with fewer source-receiver pairs), and, thus, (iv) it improves local (multi-)parameter estimation. The ultimate objective is the detection, extraction, and interpretation of more information from the broad band data and to map correctly the scales represented in them (implied by frequency range) to the spatial scales in the physics (or the 'model'). We argue that the use of individual measurements of ballistic waves (i.e., phase arrivals) combined with back projection along a linearized kernel (rays, BDKs, whatever) will not meet this challenge and that, instead, a rigorous

broad-band, multi-scale analysis is needed. HH05 and de Hoop and Van der Hilst (under revision for GJI) introduce approaches for wave equation transmission and reflection tomography based on multi resolution and adjoint state methods, and also full computational approaches (e.g.,Tromp *et al.* (2005)) hold significant promise.

ACKNOWLEDGMENTS

We thank Raffaella Montelli and Guust Nolet (Princeton University) for making their models available and for the open exchange of information pertaining to their imaging. Discussions with Bradford Hager (MIT) and Richard O'Connell (Harvard) helped us improve the manuscript. We thank Chang Li (MIT) for preparing material for figures 1 to 5, and Huub Douma (CSM) for preparing figures 6. Our research is supported by the NSF grant EAR-0409816 and a Dutch National Science Foundation grant (NWO:VICI 865.03.007) for Innovative Research.

REFERENCES

Bolton, H. & Masters, G., 2001. Travel times of P and S from the global digital seismic networks: Implications for the relative variation of P and S velocity in the mantle, *J. Geophys. Res.*, **106 (B7)**, 13527–13540.

Dahlen, F. & Nolet, G., 2005. Comment on "on sensitivity kernels for wave-equation transmission tomography" by M.V. de Hoop and R.D. van der Hilst, *Geophys. J. Int.*, p. in press.

Dahlen, F., Hung, S.-H., & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes–I. Theory, *Geophys. J. Int.*, **141**, 157–174.

De Hoop, M. & Van der Hilst, R., 2005. On sensitivity kernels for 'wave-equation' tomography, *Geophys. J. Int.*, **160**, DOI: 10.1111/j.1365–246X.2004.02509.

Dencker, N., 1982. On the propagation of polarization sets for systems of real principal type, *J. Funct. Anal.*, **46**, 351–372.

Duistermaat, J., 1996. *Fourier Integral Operators*, Birkhäuser, Boston.

Engdahl, E., Van der Hilst, R., & Buland, R., 1998. Global teleseismic earthquake relocation from improved travel times and procedures for depth determination, *Bulletin Seismological Society of America*, **88**, 722–743.

Hörmann, G. & De Hoop, M., 2002. Detection of wave front set perturbations via correlation: Foundation for wave-equation tomography, *Appl. Anal.*, **81**, 1443–1465.

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

Kárason, H., 2002. *Constraints on mantle convection from seismic tomography and flow modeling*, Ph.D. thesis, Massachusetts Institute of Technology.

Kárason, H. & Van der Hilst, R., 2001. Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (pkp, pdiff), *J. Geoph. Res.*, **106**, 6569–6588.

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

Luo, Y. & Schuster, G., 1991. Wave-equation travel time inversion, *Geoph.*, **56**, 645–653.

Montelli, R., Nolet, G., Dahlen, F., Masters, G., Engdahl, E., & Hung, S.-H., 2004a. Global P and PP travel time tomography: rays vs. waves, *Geophys. J. Int.*, **158**, 637–654.

Montelli, R., Nolet, G., Dahlen, F., Masters, G., Engdahl, E., & Hung, S.-H., 2004b. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, **303**, 338–343.

Strichartz, R., 2003. *A guide to distribution theory and Fourier transforms*, World Scientific Publishing, Singapore.

Trampert, J. & Spetzler, J., 2005. Surface wave tomography: finite frequency effects lost in the null space, *Geophys. J. Int.*, p. under review.

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

Woodward, M., 1992. Wave equation tomography, *Geoph.*, **57**, 15–26.

Figure captions

FIGURE 1. Mantle structure beneath subduction zones in South America according to – from left to right – MIT-P05, a recent model by Van der Hilst et al. (in preparation); the Princeton model based on finite frequency theory (PRI-FFT), and a Princeton model based on ray theory (PRI-RT). The color scale is the same for all models, but the amplitude of the anomalies in the models by M04a,b is up to 50% larger than in MIT-P05 so that some color saturation occurs. Based on the similarity of PRI-RT and PRI-FFT we attribute the differences between MIT-P05 and PRI-FFT not to the use of BDKs but to the use of a different parameterization, regularization, and data misfit criterion. (NB. True amplitude imaging remains a challenge, and test inversions with synthetic data suggest that amplitudes are underestimated by a significant fraction.)

FIGURE 2. To illustrate of effect of 'banana doughnut kernels' on the images of plume-like structures we show the lateral variation in P-wave speed at 150 km depth beneath the beneath the Indian Ocean (K is Kerguelen) and at 1350 km depth beneath Hawaii; we display results of inversions by Montelli et al. (2004a,b) with classical ray theory – on the left – and with finite frequency theory (that is, 'banana doughnut kernels') – on the right. The color scales are the same. We leave it to the reader to assess the differences between the RT and FFT results.

FIGURE 3. Lateral variation in P-wavespeed according to MIT-P05 for the regions and depths used in figure 2.

FIGURE 4. P-wave speed at 450 km depth beneath the Indian Ocean according to the RT (left) and FFT inversions (middle) due to M04a and M04b, respectively. Figure on the right depicts geographical

distribution of sources (red dots), receivers (blue dots), and the PP bounce points (green dots). The PP bounce point distribution is sparse in much of the Indian Ocean.

FIGURE 5. Comparison of wavespeed perturbations δc inferred from FFT and RT inversions. We use δc_{FF} (δc_{RT}) to denote the relative variation in wavespeed according to the FFT (RT) models. (a) histogram of $\delta c_{FF}/\delta c_{RT}$ for depth $\sim$1450 km and for models with $\chi^2/N = 1.0$ (after Fig. 13 of M04a). According to M04a the mean of this distribution is around 1.7. The low count near zero and the skewness of the main lobe result from omitting low values of $|\delta c/c|$ from the analysis. From PRI-FFT (M04b) and PRI-RT (M04a) we produced (here shown for $\sim$1650 km depth) histograms of (b) $\delta c_{\rm FF}/\delta c_{\rm RT}$ for all values; (c) $\delta c_{\rm FF}/\delta c_{\rm RT}$ for $|\delta c/c|$ larger than 0.2%; and (d) $\delta c_{\rm RT}/\delta c_{\rm FF}$, also for $|\delta c/c|$ larger than 0.2%. Note that the mean shifts to values larger than 1 because of the omission of the weaker anomalies. (e) Scatter plot of δc_{FF} vs. δc_{RT} (same depth) (coorelation coefficient $R = 0.92$; slope (from bi-variant regression) is 0.96). The dashed line (slope of 1.5) marks the expectation from DN05's claim of a $\sim$50% amplitude increase. (f) (Normalized) Root mean square of wavespeed amplitude as function of wavespeed ratio for six different depths. At each depth, the anomalies with the largest wavespeed values have ratios $\sim$1, whereas the negative or large positive wavespeed ratios involve the weak and – perhaps – poorly resolved anomalies.

FIGURE 6. This figure illustrates the scale and directional content of a kernel (see HH05 for a more detailed discussion). Left: space domain representation; right: Fourier domain representation. The space domain depicts the contributions in space to the (normalized) sensitivity kernel, which is projected onto a curvelet frame. In the Fourier domain representation the largest length scales (i.e., smallest wave numbers k) plot toward the center, and the direction in which the scale annuli are filled mark the direction of maximum resolution. In this example, the kernel is of 'banana doughnut' type and the (background) medium is, thus, homogeneous. The top row (same as Fig. 8d of HH05) illustrates contributions (right) to

the kernel from curvelets located on the unperturbed ray half-way between source and receiver (indicated with a cross in the space domain panel). The Fourier transform shows that data with the specific frequency content used here do sense structure on the unpertubed ray but only for certain scales, and (primarily) in directions orthogonal to the ray. In the bottom row the spatial location of curvelets considered is moved away from the ray to the first 'optimum' of the regularized kernel (lower left). The Fourier domain representation, and hence the 'curvelet content', is practically the same as for the point on the ray. This demonstrates that data with a certain frequency content are sensitive to structure at certain length scales (wave numbers) and that this sensitivity is the same for structure on the ray and away from it.

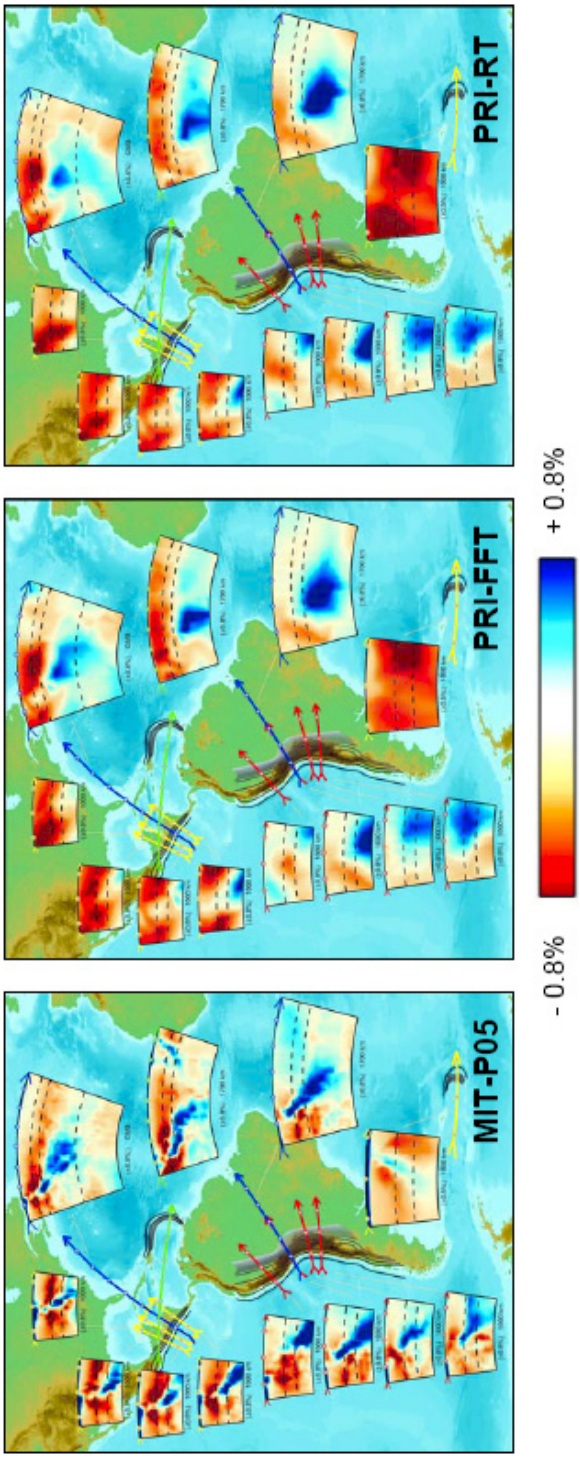

Figure 1.

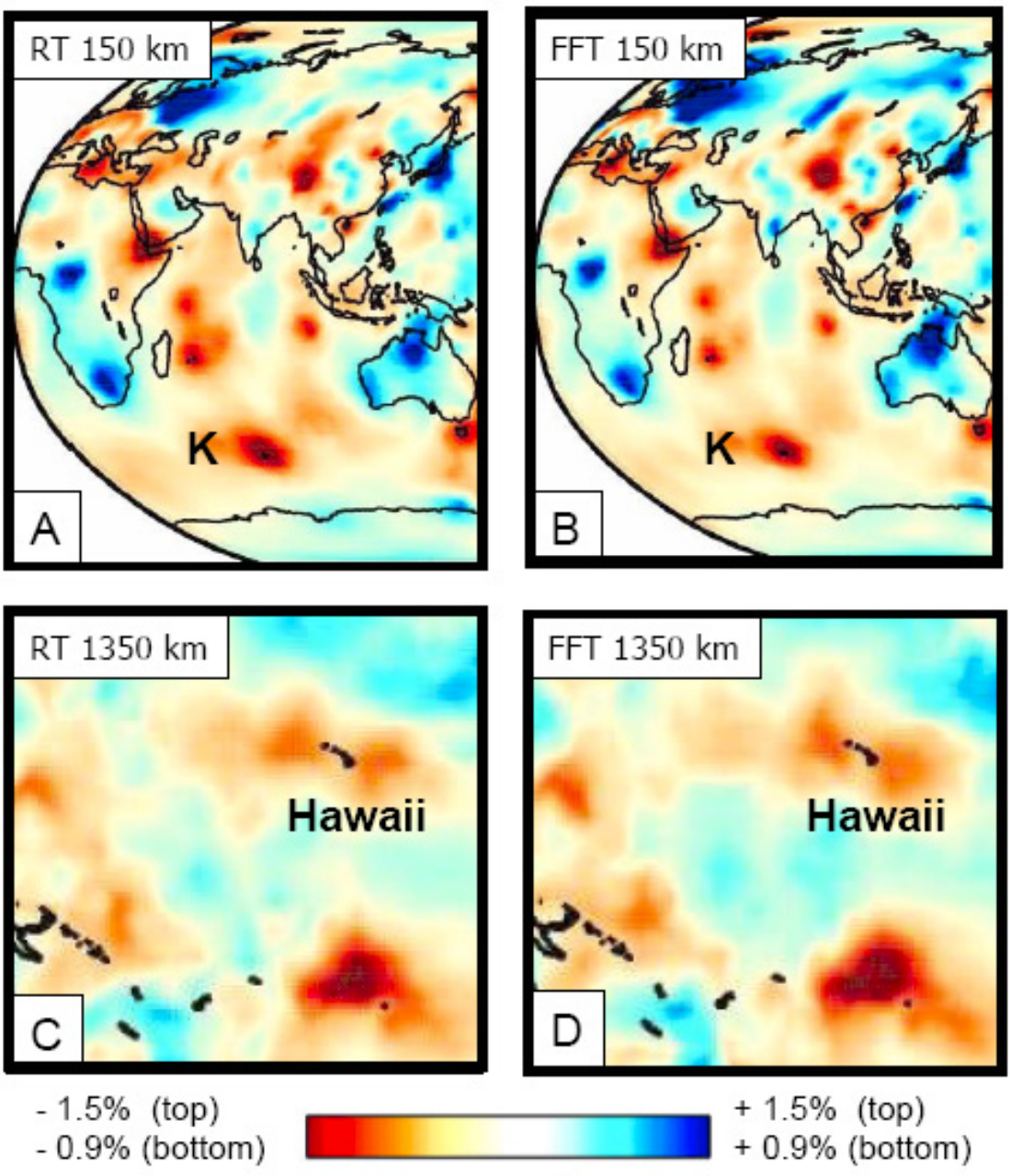
RT 150 km
K
A
FFT 150 km
K
B
RT 1350 km
Hawaii
C
FFT 1350 km
Hawaii
D
- 1.5% (top)
- 0.9% (bottom)
+ 1.5% (top)
+ 0.9% (bottom)

Figure 2.

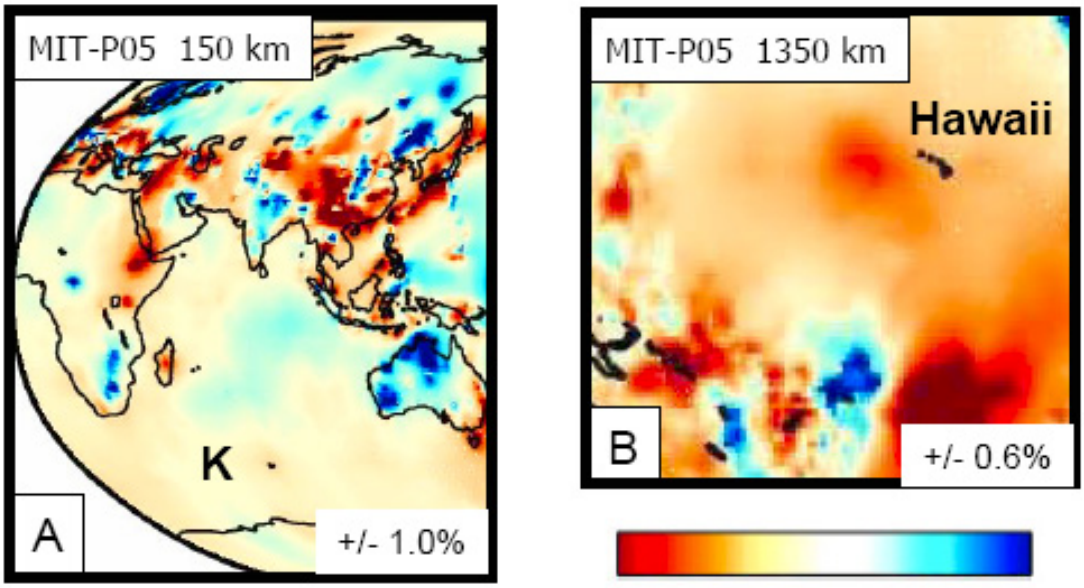

Figure 3.

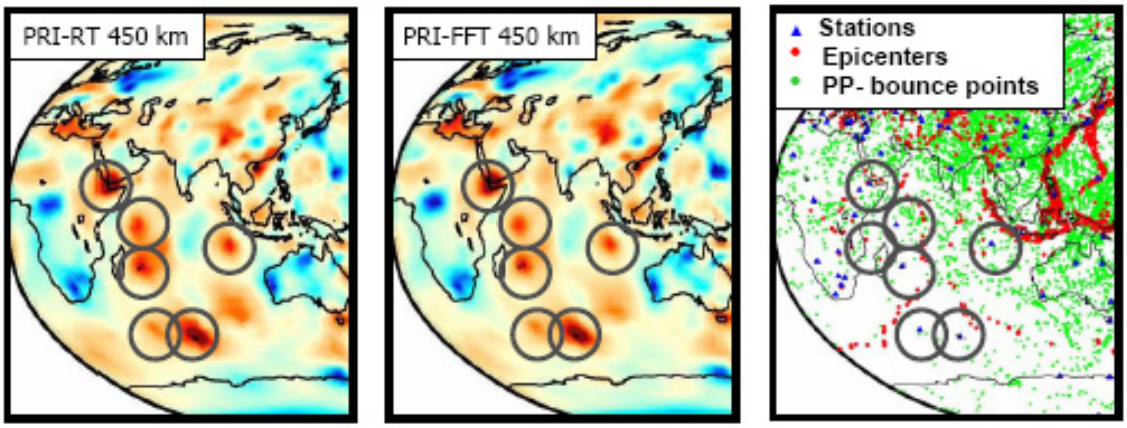
PRI-RT 450 km
PRI-FFT 450 km
Stations
Epicenters
PP- bounce points

Figure 4.

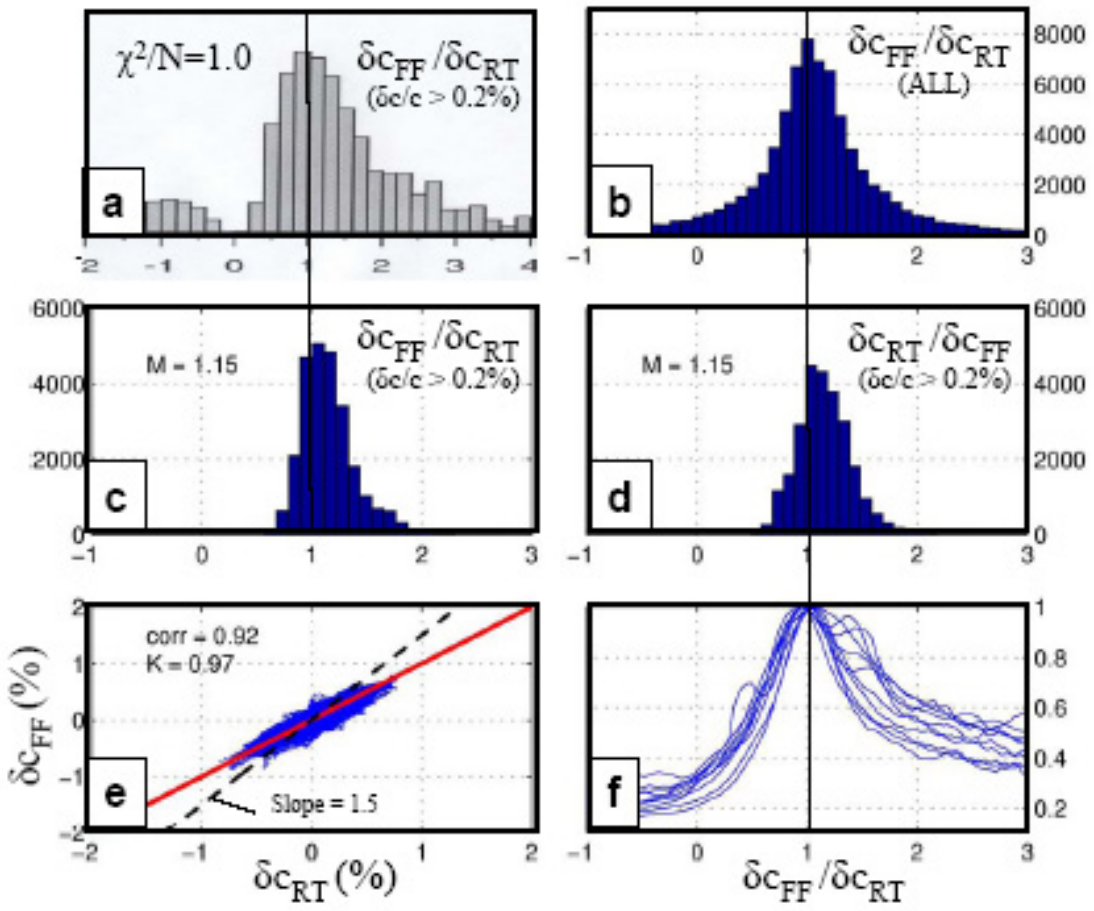
χ²/N=1.0
δc_FF /δc_RT
(δc/c > 0.2%)
a
δc_FF /δc_RT
(ALL)
b
M = 1.15
δc_FF /δc_RT
(δc/c > 0.2%)
c
M = 1.15
δc_RT /δc_FF
(δc/c > 0.2%)
d
corr = 0.92
K = 0.97
Slope = 1.5
e
f
δc_FF (%)
δc_RT (%)
δc_FF /δc_RT

Figure 5.

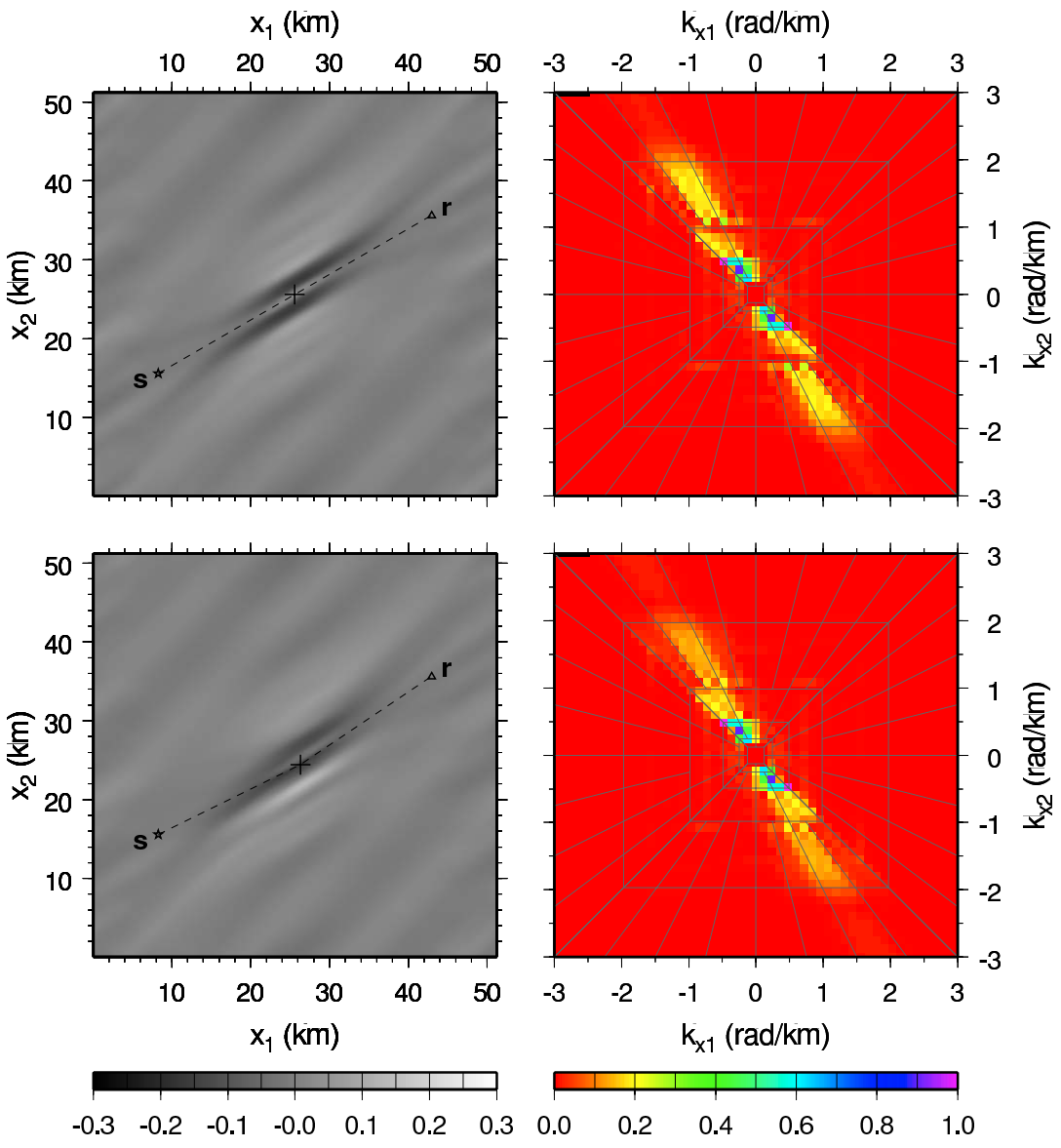
x1 (km)
kx1 (rad/km)
x2 (km)
kx2 (rad/km)
s
r
-0.3 -0.2 -0.1 -0.0 0.1 0.2 0.3
0.0 0.2 0.4 0.6 0.8 1.0

Figure 6.