

# Automated multimode inversion of surface and *S* waveforms

Sergei Lebedev,<sup>1</sup> Guust Nolet,<sup>2</sup> Thomas Meier<sup>3</sup> and Rob D. van der Hilst<sup>4</sup>

<sup>1</sup>Department of Earth Sciences, Utrecht University, Budapestlaan 4, Utrecht, 3584 CD, the Netherlands. E-mail: sergei@geo.uu.nl

<sup>2</sup>Department of Geosciences, Princeton University, Princeton, NJ 08544, USA

<sup>3</sup>Department of Geology, Mineralogy and Geophysics, Ruhr University Bochum, Bochum 44780, Germany

<sup>4</sup>Massachusetts Institute of Technology, EAPS, Cambridge, MA 02139, USA

Accepted 2005 June 8. Received 2005 May 16; in original form 2004 November 23

## SUMMARY

Inversion of the surface, *S*, and multiple-*S* waveforms is an effective means of constraining the structure of the upper mantle, including the transition zone. Exploiting the resolving power of the enormous volume of presently available data requires efficiency of data processing and waveform modelling. An established method for rapid generation of synthetic seismograms is the summation of surface-wave modes under an assumption that the effect of seismic-wave scattering is negligible. This assumption is valid at best for portions of a broad-band signal, the portions being generally different for different seismograms. Interactive selection of such parts of the signal is impractical for large data sets of tens or hundreds of thousand seismograms. Here we present a fully automated waveform inversion technique with selection of signal uncontaminated by scattered waves implemented as its integral element. Waveforms are inverted using non-linear optimization and the results are put in the form of orthogonalized linear constraints on average elastic structure along the source–station paths. Structural information from waves of different amplitudes and different types is balanced by means of time- and frequency-dependent weighting, also used to suppress the propagation of errors in the data. The equations obtained from processing thousands of seismograms can be inverted together for high-resolution upper-mantle models. The technique has been applied to a large Western Pacific data set. Analysis of the results suggests that it has been effective and, in particular, confirms that ‘chance’ fits of scattered waves or noise do not pass the automated selection procedure. Results of the processing also provide an empirical mapping of the field of the JWKB-approximation validity for modelling the propagation of surface waves. While there is no sharp boundary of the JWKB regime, the probability of the approximation validity decreases with increasing distance and frequency.

**Key words:** inverse problem, scattering, synthetic seismograms, tomography, upper mantle, waveform analysis.

## 1 INTRODUCTION

Waveform inversion techniques are designed to constrain Earth structure using the abundant information contained in seismic-wave forms. Assuming a model of the Earth, one can compute synthetic seismograms—or solve the forward problem. One can also search for perturbations to the model such that the synthetic waveforms match the observed ones—that is, ‘fit’ the waveforms. Constraining the perturbations amounts to constraining Earth structure—or solving the non-linear inverse problem.

In early waveform-inversion applications (e.g. Dziewonski & Steim 1982; Woodhouse & Dziewonski 1984) a small number of Earth-model parameters (up to  $O(10^2)$ ) were adjusted so as to fit simultaneously a set of long-period seismograms [up to  $O(10^3)$ ]. The volume of available broad-band data has since increased dramatically, and so has its resolving power. In order to enable handling

the much larger number of parameters needed for high-resolution imaging, waveform analysis schemes have been broken into steps (e.g. Cara & Lévêque 1987; Nolet 1990; Gee & Jordan 1992), with an initial extraction of linear equations from individual seismograms followed by solving the resulting linear system for an Earth model. The equations may constrain data functionals specially designed for each recording or standard functionals such as phase velocities of multiple modes (e.g. van Heijst & Woodhouse 1997; Yoshizawa & Kennett 2002; Beucler *et al.* 2003; in this paper we focus on multimode analysis only).

Increasing sophistication of waveform inversion techniques and the related broadening of the ranges of analysed frequencies and source–station distances have helped to constrain detailed upper-mantle models (e.g. Zielhuis & Nolet 1994; Gaherty *et al.* 1999; Debayle & Kennett 2000; Simons *et al.* 2002; Yoshizawa & Kennett 2004). In particular, waveform analysis enables mapping

of transition-zone structure (e.g. Lebedev *et al.* 2002; Ritsema *et al.* 2004) using the information carried by triplicated  $S$  and multiple- $S$  waves, regardless of the fact that individual arrivals within their wave trains are usually unidentifiable. Despite the progress, however, we are not yet close to exploiting fully the resolving power of the enormous volume of available broad-band data. Fundamental improvements in modelling are being introduced (e.g. Megnin & Romanowicz 2000; Friederich 2003) but even methods based on conventional approaches—such as the computationally efficient JWKB (Jeffreys–Wentzel–Kramers–Brillouin) approximation for modelling the propagation of multimode surface waves (Dahlen & Tromp 1998)—can still be improved in accuracy and efficiency and still hold a potential for an advance in resolution of large- to global-scale imaging. In this paper, we present the Automated Multimode Inversion (AMI) technique aimed to enable

- (1) processing very large amounts of data, by making use of full automation;
- (2) extracting the maximum of information from a broad-band seismogram while ensuring that relevant theoretical approximations hold; and
- (3) balancing this information when relating it to Earth structure (in particular by weighting waves of different amplitudes and different types) so as to constrain unbiased Earth models.

Previously, van Heijst & Woodhouse (1997) introduced a high degree of automation into measurements of phase velocities of multimode surface waves. The phase-velocity maps obtained from the measurements can be used to constrain 3-D mantle models (e.g. Ritsema *et al.* 2004), with a convenient option of using accurate, local Fréchet derivatives of phase velocities with respect to elastic parameters at every point. The main difficulty with this approach is that maps for different modes and different frequencies may have different lateral resolution, both for fundamental physical reasons and because measurements for some modes and frequencies are more abundant than for others; phase-velocity-map constraints are thus easiest to combine at relatively long wavelengths ( $O(10^3)$  km).

Debayle (1999) has automated the procedure developed by Cara & Lévêque (1987) and inverted waveforms directly for constraints on average elastic-structure perturbations along source–station paths. Although this approach involves measurements of observables other than phase velocities, one remaining problem is the correlation between measurement errors. Ignoring this may result in an inconsistency of linear equations, the consequence being a decrease in the resolution of the imaging. In order to avoid such correlations, AMI produces linear constraints that are orthogonalized (Nolet 1990). The linear equations are provided with uncorrelated uncertainties that are directly related to seismic waveforms; the uncertainties can be used to weight the equations.

Two most important features of AMI set it apart from other techniques. The first is the case-by-case selection of time–frequency portions of seismograms that can be modelled with the JWKB approximation. The second is the elaborate, time- and frequency-dependent waveform weighting. A simple least-squares, data-synthetic misfit function would depend on both the phase and the amplitude of the signal. For modelling the amplitudes of seismic waves, the limits of the JWKB-approximation validity are expected to be much tighter than for modelling the phases (e.g. Wang & Dahlen 1995). Errors from sources other than the breakdown of the approximation are also larger for synthetic amplitudes than they are for the phases, as we discuss below. For these reasons, we wish to extract structural constraints from phase information only. We

also wish to prevent the most energetic waves from dominating the misfit function; we would like to weight the contributions of different wavegroups according to the amount of structural information they contain rather than according to their relative amplitude. Nolet (1990) equalized the energy of the synthetic with that of the data (so as to remove the effect of the absolute amplitude) and also applied envelope weighting in the time domain (so as to compensate for relative amplitude difference between different wavegroups). Here we introduce a weighting scheme that tunes the relative amplitudes both in the time and the frequency domains and enables extracting phase information fully from all parts of the signal. For example, an energetic intermediate-period fundamental mode will have the same weight as a smaller-amplitude, long-period fundamental mode, and  $S$  waves will be given the same (or larger if desired) weight.

AMI has been applied to a Southeast Asia–Western Pacific data set of over 4000 seismograms; the resulting tomographic model has been presented by Lebedev & Nolet (2003). Currently, the method is used for processing hundreds of thousands of seismograms from around the globe with the goal of improving the resolution of global upper-mantle imaging. It is the full automation of the processing that enables utilization of such a large multimode-waveform data set; it will also make AMI well suited for analysis of the data from the rapidly expanding regional broad-band networks, such as the USArray in the US. Here we describe the set-up of AMI in detail and examine the results of its first application with a focus on the accuracy and effectiveness of the procedure.

## 2 PARTITIONED WAVEFORM INVERSION

AMI uses the partitioned-waveform-inversion scheme formulated by Nolet (1990). Adopting the JWKB approximation, we compute synthetic seismograms as

$$s(\omega) = \sum_m A_m(\omega) \exp \left[ i\omega\Delta / (C_m^0(\omega) + \overline{\delta C_m(\omega)}) \right], \quad (1)$$

with phase-velocity perturbations expressed as functions of path-averaged  $S$ -velocity perturbations:

$$\overline{\delta C_m(\omega)} = \int_0^a \frac{\partial C_m^0(\omega)}{\partial \beta(r)} \overline{\delta \beta(r)} dr. \quad (2)$$

The sum is over modes  $m$ , the source–receiver distance is  $\Delta$ , and the radius of the Earth is  $a$ . The complex amplitudes of the modes  $A_m(\omega)$  depend on the source mechanism (taken from, e.g. Dziewonski *et al.* 1994) and the Earth model assumed for the source region, as well as on geometrical spreading and attenuation (the calculations involved are detailed in, e.g. Aki & Richards 1980; Dahlen & Tromp 1998). The reference phase velocities  $C_m^0(\omega)$  and their Fréchet derivatives  $\partial C_m^0(\omega)/\partial \beta(r)$  are calculated (Masters, Reference Earth Model web page; Takeuchi & Saito 1972; Nolet 1990) for a 1-D model with an estimated average structure of the mantle and crust (including surface elevation or thickness of the water layer) along the source–station path. The Fréchet derivatives relate the path-averaged perturbation in  $S$  velocity  $\overline{\delta \beta(r)}$  to perturbations in phase velocities  $\overline{\delta C_m(\omega)}$  and through them to perturbations in synthetic waveforms. The effect of variations of other parameters such as  $P$  velocity can be modelled in the same fashion through appropriate derivatives.

We fit the Fourier-transformed synthetic  $s(t)$  to the data  $d(t)$  in the time domain by minimizing a least-squares misfit function of the form

$$F(\overline{\delta \beta(r)}) = \int |\mathbf{R} d(t) - \mathbf{R} s(t, \overline{\delta \beta(r)})|^2 dt, \quad (3)$$

where  $\mathbf{R}$  is a filtering, windowing and weighting operator. Waveform fitting is performed by means of non-linear optimization (Nolet *et al.* 1986).

We parametrize  $\overline{\delta\beta(r)}$  using a set of 1-D basis functions  $h_i(r)$ :

$$\overline{\delta\beta(r)} = \sum_{i=1}^M \gamma_i h_i(r). \quad (4)$$

The basis functions span the crust, upper mantle, and the upper part of the lower mantle. The number  $M$  of the inversion parameters  $\gamma_i$  is 15–20 for  $S$ -velocity perturbations  $\overline{\delta\beta(r)}$ ; if independent parameters are used for  $P$ -velocity perturbations as well, a sufficient  $M$  is 20–30.

Both the best-fitting values of the parameters  $\gamma$  and the properties of  $F(\gamma)$  in the vicinity of its minimum depend—and represent information—on the structure of the Earth. Diagonalization of the Hessian matrix  $H_{ij} = \partial^2 F / \partial \gamma_i \partial \gamma_j$  puts this information in the form of independent linear equations with uncorrelated uncertainties

$$\int_{\oplus} G_i^{\beta}(\mathbf{r}) \delta\beta(\mathbf{r}) d^3\mathbf{r} = \eta_i \pm \Delta\eta_i, \quad (5)$$

where  $i = 1, \dots, M$ . New parameters  $\eta_i$  correspond to new basis functions  $g_i(r)$ :

$$\overline{\delta\beta(r)} = \sum_{i=1}^M \eta_i g_i(r), \quad (6)$$

with each  $g_i(r)$  a linear combination of the  $M$  original basis functions  $h_i(r)$  (for details see Nolet 1990).  $G_i^{\beta}(\mathbf{r})$  define the  $S$ -velocity sensitivity volumes within the 3-D Earth  $\oplus$ . At fixed latitude and longitude, the dependence of  $G_i^{\beta}(\mathbf{r})$  on radius is proportional to that of  $g_i(r)$ . For parameters of different types, additional terms will be added on the left-hand side of (5), for example,  $\int_{\oplus} G_i^{\alpha}(\mathbf{r}) \delta\alpha(\mathbf{r}) d^3\mathbf{r}$  in the case of  $P$  velocity. Eqs (5) obtained from different seismograms of the data set are weighted with their uncertainties  $\Delta\eta_i$  and solved together for a 3-D tomographic model.

### 3 AUTOMATED MULTIMODE INVERSION (AMI)

Due to the occurrence of small-scale heterogeneity in the Earth, scattered waves are recorded on any broad-band seismogram. They are unaccounted for by (1) and interfere with the signal predicted by (1). If the scattered contribution is substantial, the data can no longer be expected to be matched by a JWKB synthetic. Phase shifts in the recorded signal are no longer related to Earth structure through eqs (1) and (2). The JWKB approximation is not valid and—if used—will bias the inferred structural constraints.

The proportion of scattered energy on a seismic record varies from one point in time to another and from one frequency to another. Some seismograms are dominated by scattered waves throughout their surface-wave part. For others, there exist portions of the signal in the time–frequency plane with negligibly small contribution from scattering. We wish to select these portions and use only them for constraining Earth models. Our main assumption in achieving this is simple: if synthetics computed with (1) and (2) can reproduce the data almost exactly (in a number of time–frequency windows satisfying a set of criteria, as defined below) then the scattered-wave contribution is negligible, the JWKB approximation is valid, and (1) and (2) accurately relate the phase information in the recorded signal to Earth structure.

The quality of the data-synthetic fit is an obvious indicator of a success of a waveform inversion and has been employed as such in numerous studies. Nolet (1990), for example, interactively selected one or two broad-band, time–frequency windows for each seismogram, one for the fundamental mode and possibly another for the

higher modes (for the  $S$ - to multiple- $S$  part of the seismogram). The widths of the windows were modified and the waveform fits within them were recomputed and inspected visually; the procedure was repeated until the fits were judged acceptable. The waveforms in broad-band windows, however, are complex, the fits are usually far from exact, and the judgement on their quality is inevitably subjective.

Here we develop quality-of-fit criteria that can be effective and consistent in a fully automated processing. By filtering the seismogram through a series of time–frequency windows, we replace the complex broad-band waveform with a series of much simpler ones. Thanks to the simplicity of the waveforms, the computed value of the misfit becomes an accurate and sufficient measure of similarity between the actual and synthetic signals. For each wave train (one or more in each of the time–frequency windows) we can obtain an almost perfect data-synthetic fit as long as the scattered-wave and noise contributions are small.

The basic requirement to the waveform-fitting procedure is that it does not allow ‘chance’ fits of noise or scattered waves with JWKB synthetics. We achieve this in part by computing synthetics for 1-D seismic-velocity profiles (2) perturbed using a relatively small number of basis functions (4). The depth span of the basis functions increases from  $O(10)$  km in the crust and uppermost mantle to over 100 km in the lower mantle. No additional smoothness constraints are applied. The parametrization is adequate for modelling realistic 1-D profiles but it does not allow obviously unreasonable ones—for example, with very thin [ $O(1)$  km] layers in the mantle with very large, alternating positive and negative anomalies in each—which could result in exotic perturbations of modes’ group velocities and possibly produce chance fits. With 15–30 basis functions the 1-D waveform-fitting problem is underdetermined, and our choice of a particular number of parameters in the range has little effect on the success of waveform fitting. The eventual number of significant independent equations (5) is determined through the orthogonalization (6).

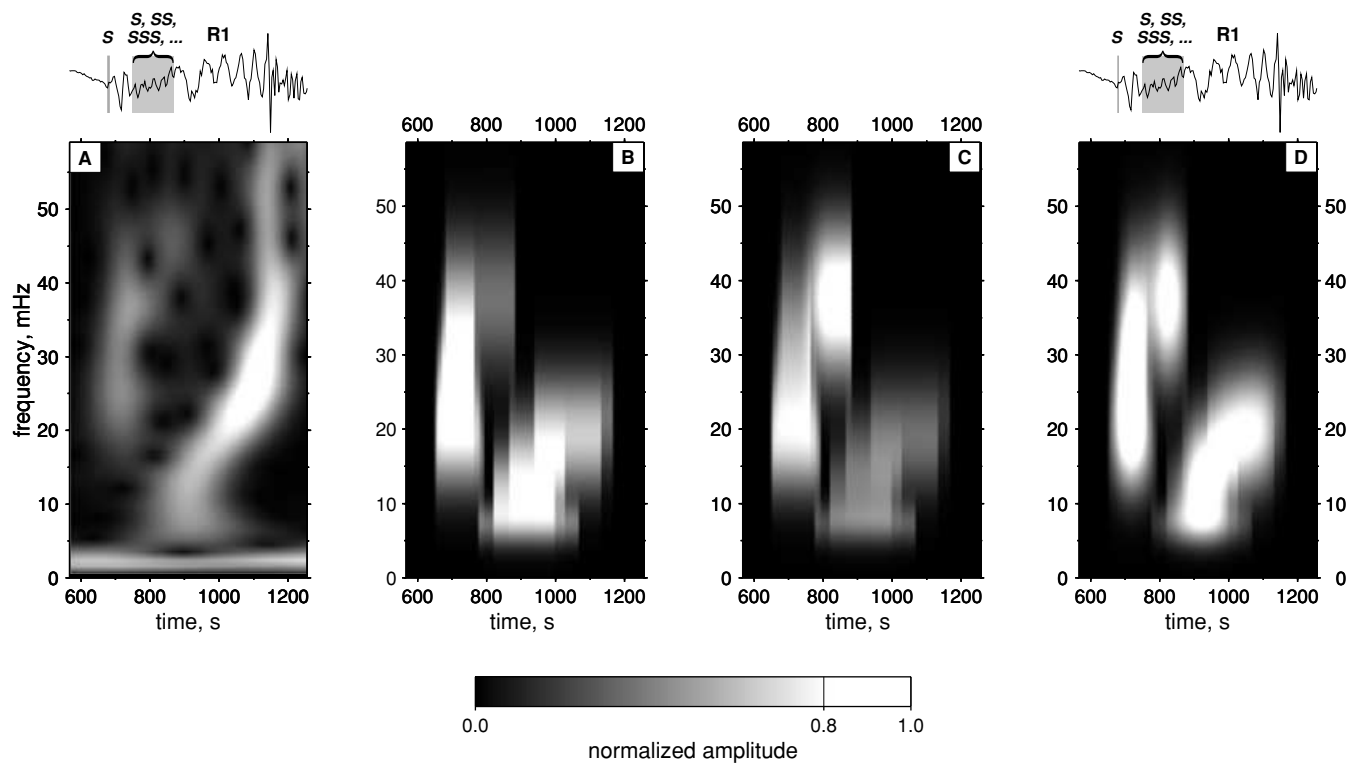
Selecting the time–frequency windows and fitting the waveforms within them proceed simultaneously, with new windows added one by one. Additional selection criteria are applied to avoid waves radiated at near-nodal azimuths, to guard against chance fits, to prevent cycle skipping, and to ensure the validity of the far-field and point-source approximations.

Before waveform inversions, we run automated checks of data quality and identify and discard clipped and otherwise defective traces. In the following we assume that seismograms are recordings of the actual vibrations of the Earth’s surface beneath seismometers.

Figs 1 and 2 show an example of the windowing-weighting operator  $\mathbf{R}$  designed for one seismogram and the corresponding waveform fit. In the following we describe the criteria and rules used by the algorithm in order both to ensure the validity of well-known approximations—such as the far-field and point-source approximations (Dahlen & Tromp 1998)—and to enable an accurate, fully automated multimode waveform inversion.

#### 3.1 Near-nodal radiation

If the source–station azimuth is near a node in the radiation pattern of a wave, this wave’s portion of the seismogram is likely to be dominated by scattered energy. For our purposes, this portion is to be avoided. Another reason to avoid such waves is that the synthetics computed to model them will be affected the strongest by phase errors due to source-mechanism uncertainties (see Appendix). Also, sensitivity volumes of seismic waves become increasingly more



**Figure 1.** The construction of the time–frequency filter for a seismogram. The earthquake occurred on 1997 November 3, at 29°N, 85°E and a 36-km depth. The vertical-component recording is from the station SSE (Sheshan) operated by the China Digital Seismograph Network (CDSN); it is response-corrected to displacement. The broad-band seismogram is shown above the frames A and D. The theoretical arrival time of the short-period *S* wave computed in AK135 (Kennett *et al.* 1995) is shown with a vertical line. Arrivals of triplicated multiple-*S* waves interfere with each other within the grey band. R1 indicates the fundamental Rayleigh mode. (a) Amplitude as a function of time and frequency, computed with a multiple-filter technique (e.g. Meier *et al.* 2004). (b) The set of selected time–frequency windows. The amplitude shown at a point is a sum of the amplitudes of all Gaussian-filter curves (Fig. 2) corresponding to the time–frequency windows that the point falls within. (c) Weighted time–frequency windows. The amplitude at a point is a sum of the amplitudes of Gaussian-filter curves as in (b) but multiplied with weighting functions, different in each time–frequency window. (d) The signal (a) passed through the weighted filter (c). The resulting amplitude at a point shows the contribution of different time–frequency portions of the signal to the misfit function.

complex towards near-nodal azimuths, with the Earth’s structure that is sampled being mostly away from the source–station great-circle plane. An example in Fig. 3 shows the fundamental-mode sensitivity areas computed using the Born approximation as by Meier *et al.* (1997). At the 50° azimuth the area bends away from the great circle, and at the nodal 90° its separate parts are on the sides of it. Unless the sensitivity is modelled explicitly, this complexity represents yet another reason to identify and avoid near-nodally radiated waves.

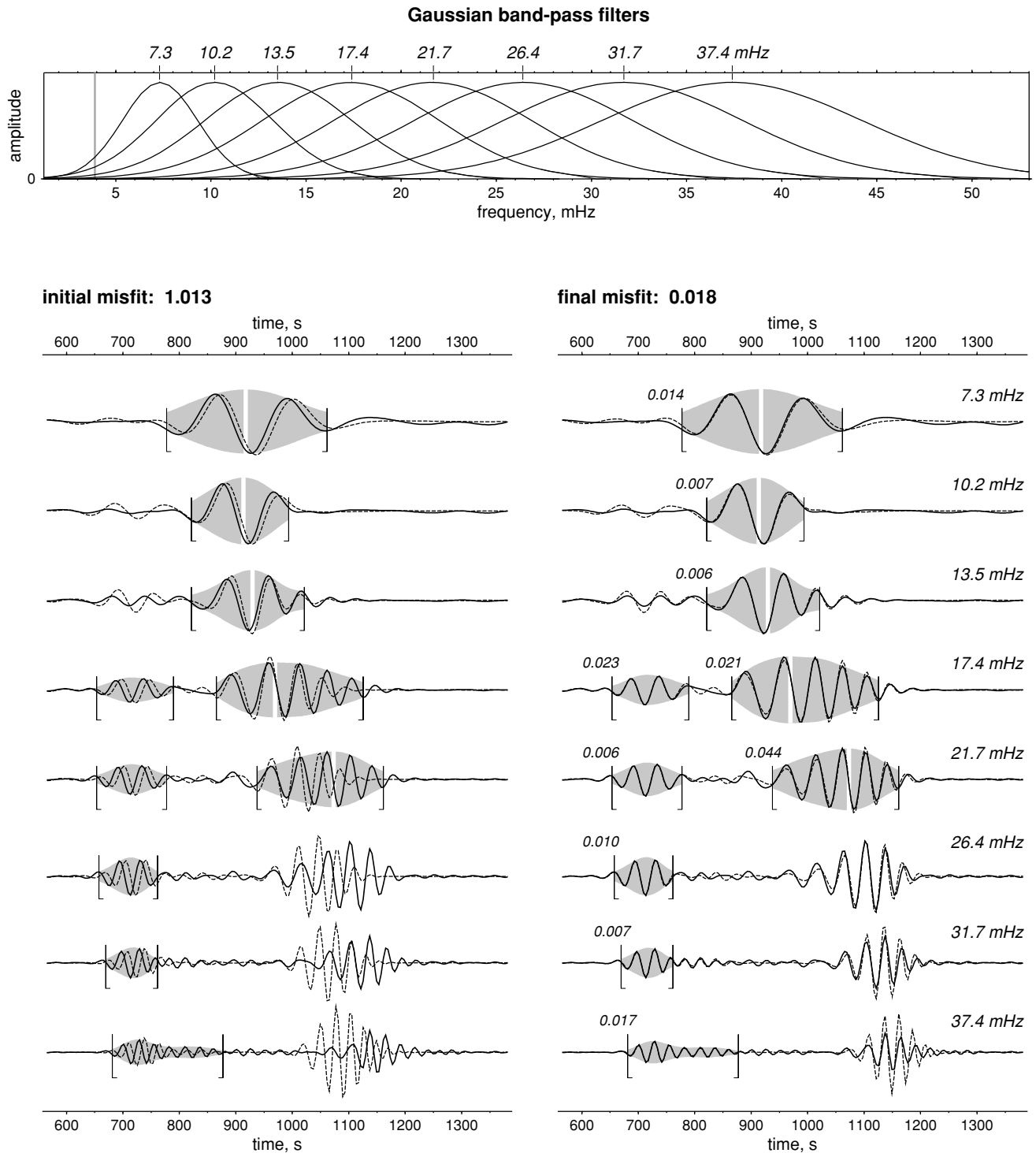
Before the waveform inversion we compute the frequency-dependent radiation patterns of the modes. For each source–station pair, we select the frequency bands in which the amplitude of the waves radiated at the azimuth at each frequency is at least 0.5 of the maximum amplitude at the frequency. Only the signal within these bands is used in the waveform inversion. Applying the 0.5 threshold in the example in Fig. 3, we would select the signal at azimuths between 0° and 50° but not between 50° and 90°.

The radiation pattern of the fundamental mode and that of the sum of the higher modes are generally different and are computed separately. The node-free band selected for the higher modes is used for all higher mode windows (containing both *S* and multiple-*S* waves). Depending on the source mechanism and the source–station azimuth, the node-free frequency bands for the fundamental and higher modes can have zero, finite or infinite widths.

### 3.2 Definition of the time–frequency windows

In the course of processing a seismogram we select a series of densely spaced Gaussian filters (Fig. 2). Their relative width increases with frequency so that the resolution in the time domain remains approximately constant (Dziewonski & Hales 1972). The resolution in the time domain trades off with that in the frequency domain: the narrower the filters, the wider the seismic-wave trains. We choose filter widths such that a seismogram recorded at a distance of a few thousand kilometres from the source contains a few separate wave trains after filtering (e.g. the *S* and fundamental-mode wave trains in Fig. 2).

We define a time–frequency window as a boxcar time window applied after bandpassing through a Gaussian filter. The signal around every maximum of the signal envelope (i.e. between every pair of neighbouring envelope minima) is considered a wave train of a seismic phase, such as the *S* wave or the fundamental mode. Time-window boundaries cannot cut in the middle of a wave train, so that the windows can contain only complete wave trains: the value of the envelope of the signal at a window boundary must be at least 4–5 times smaller than at the nearest envelope maximum inside the window. Under these basic rules, time–frequency windows are selected in the course of waveform fitting (Section 3.4).



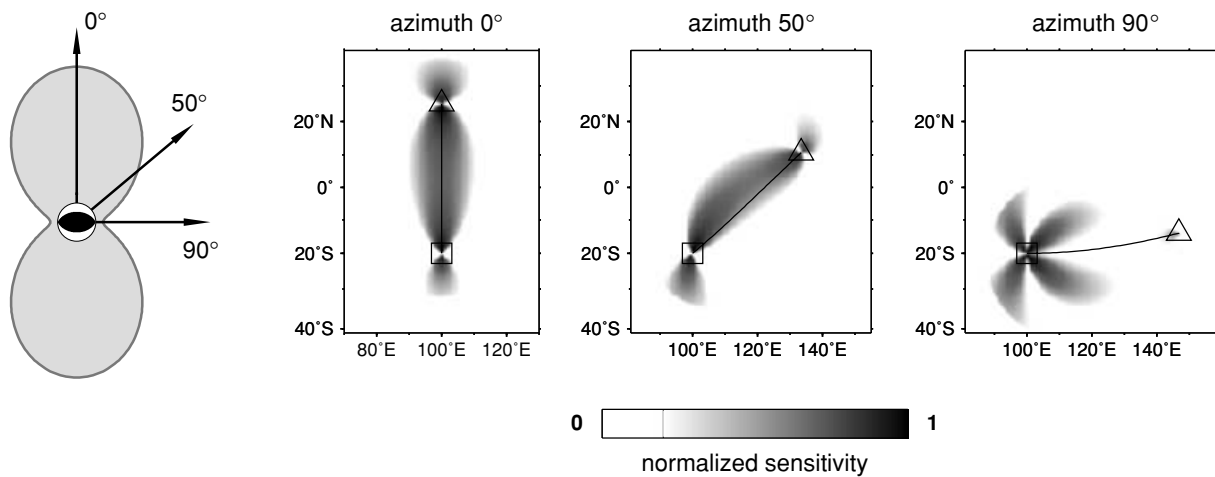
**Figure 2.** Waveform fitting. The broad-band seismogram shown in Fig. 1 is filtered using eight closely spaced Gaussian filters (top). The resulting waveforms (solid lines) are matched with synthetics (dashed lines) in ten time–frequency windows simultaneously. The time windows are indicated by half-brackets and by shading of the signal envelope. Fundamental-mode wave trains are identified by white vertical lines at the maxima of the envelope. The column of waveforms on the left shows the initial fit with synthetics computed for a 1-D background model (Fig. 4). In the column on the right, the misfit is minimized through path-averaged perturbations  $\delta\beta(r)$ . The values of the misfit are given next to each window; initial and final global misfits are given above the waveform columns.

### 3.3 Time- and frequency ranges

The maximum frequency range defined by the node-free frequency bands from Section 3.1 may need to be narrowed further. To ensure the validity of the far-field approximation, the minimum frequency

is set such that at least 3–4 wavelengths of the fundamental mode at this frequency fit between the source and the station.

A high-frequency limit is set at  $1/(3\tau)$ , where  $\tau$  is the estimated duration time of the event taken from a catalogue (e.g. Dziewonski *et al.* 1994). The purpose of this criterion is to ensure the validity



**Figure 3.** The radiation pattern of a 100-s Rayleigh wave (left) and waveform-sensitivity areas (Meier *et al.* 1997) at three azimuths, each averaging over a 65- to 105-km depth range. The source (shown on the maps with a square) was placed at a 10-km depth within a source-region model with a 25-km-deep Moho.

of the point-source approximation: waves at periods comparable to  $\tau$  will likely be affected by unmodelled complexity of the source, finite in space and time.

We put all or almost all of the Gaussian filters' bands within the allowed frequency range, that is, the centre frequency of the lowest-frequency filter is above the low-frequency limit; at the limit the filter curve has a value of 0.2–0.3 (the maximum value at the centre is 1.0). The same rule applies at the high-frequency limit.

For the time windows, the right (latest-time) limit is set immediately after the fundamental-mode wave train (identified on every trace, in every frequency band) and is intended to cut out the scattered waves of the coda. The left (earliest-time) limit depends on the distance: below 35° it is just before the expected arrival of the *S* wave train, between 35° and 70°—after the *S* and before the *SS* wave train, etc. (*S* and multiple-*S* arrival times are computed in AK135 (Kennett *et al.* 1995)). This eliminates body waves that bottom deep in the lower mantle, and are less sensitive to structure in the upper mantle which is the focus here.

### 3.4 Window selection and waveform fitting

We first compute the lowest frequency Gaussian filter within the pre-selected range (Section 3.3), bandpass the seismogram through it, and measure the amplitude of the envelope of the noise before the *P*-wave arrival. If the noise level exceeds a fraction of the maximum amplitude of the signal envelope within the *S*-wave–fundamental-mode window, we increment the centre frequency of the filter until detecting an acceptable signal-to-noise ratio. If it is not detected within the frequency range, the seismogram is discarded.

Having chosen the first frequency we attempt to obtain a waveform fit at it and, if successful, add higher frequency filters one after another. Waveform inversion within a time–frequency window is considered successful—and, thus, the window is accepted—if the relative data–synthetic misfit within it is smaller than 5 per cent (see Section 3.6). A single, widest possible (Sections 3.1 and 3.3) time window is initially attempted at every frequency. If the fit is acceptable only in parts of the window then it is either narrowed to include only the well-fit wave train(s) (as the window at 13.5 mHz in Fig. 2) or split into two or more windows with an acceptable fit within each (17.4 mHz in Fig. 2). Each time window contains at least one complete wave train (the criterion in Section 3.2); the

wave trains usually look very similar to those in Fig. 2. Waveforms in all time–frequency windows are inverted together, that is, we search for the  $\delta\beta(r)$  that minimizes the cumulative misfit over all windows.

The fundamental mode can interfere with its own coda and other scattered waves. To reduce the likelihood of chance fits in such cases we require that all accepted fundamental-mode windows are defined by successive Gaussian filters and that there are at least three such fundamental-mode windows for every seismogram. These criteria are probably conservative; our visual inspection of many hundreds of waveform fits has never detected any signs of a chance fit of fundamental-mode wave trains distorted by scattering.

Phase shifts (or distortions) resulting from interference with scattered waves are not necessarily visible by eye. For example, the coda following the fundamental mode usually becomes more energetic with increasing frequency, but the fundamental-mode wave trains often appear unaffected. In spite of the appearance, scattered waves may have caused phase shifts, and if so AMI should not be expected to reproduce those in JWKB synthetics. Indeed, we have observed that the growth of the coda usually correlates with the deterioration of the quality of the fundamental-mode fit, even if the coda appears to follow rather than overlap in time with the fundamental-mode wave train.

The only suspicious fits to pass the selection procedure have been detected within the higher mode part of seismograms at relatively long distances (above 6000–7000 km or so) and at periods below 40 s or so. Interference of numerous arrivals of triplicated multiple-*S* waves can produce complex waveforms, with numerous envelope maxima. Occasionally, a window around only one of such maxima is selected, with the fit just outside the window being surprisingly poor. The most likely reason for this is that our 1-D parametrization of structural perturbations  $\delta\beta(r)$  (4) ignores the 3-D structure of the multiple-*S* waves' sensitivity volumes, essential in these cases. A simple solution to this problem is setting a high-frequency limit for higher modes at long distances; a better—future—solution may be in a 3-D parametrization of the waveform inversion and an explicit computation of the 3-D sensitivity volumes (e.g. Meier *et al.* 1997; Friederich 2003). We expect the signal-selection procedure developed here to be even more effective when used with such a 3-D parametrization.

### 3.5 Cycle-skip prevention

By starting the fitting at the lowest frequency, we decrease the likelihood of an error in resolving the  $2\pi$  phase ambiguity (cycle skipping). Once the phase is anchored in the first frequency band, it will also be determined correctly in the following, higher frequency bands. We also require that the lowest frequency (that of the first accepted window) is low enough so that no more than a number (we use 12) of the fundamental-mode wavelengths can fit between the source and the station; if no acceptable fit can be obtained at sufficiently low frequencies, the seismogram is discarded. This makes it unlikely that a phase difference (due to the difference  $\delta C_m(\omega)$  between our initial-guess values  $C_m^0(\omega)$  and the observed ones) exceeding  $\pi$  could have been accumulated at the lowest frequency over the length of the path. As a final—and main—safeguard, in every window we first fit the envelope of the synthetic to that of the data, so as to align the wave trains before fitting the waveforms. Inspection of numerous waveform fits has confirmed that these measures are sufficient to prevent cycle skipping.

### 3.6 Energy equalization

The misfit function for an  $i$ th window is defined as

$$F_i = \frac{\int_{[l_i]} w_i(t) (d_i(t) - f_i^{\text{eq}} s_i(t))^2 dt}{\int_{[l_i]} w_i(t) d_i^2(t) dt}, \quad (7)$$

where  $d_i(t)$  and  $s_i(t)$  are the observed and synthetic seismograms passed through the same Gaussian filter, the integrals are over the length of the time window,  $w_i(t)$  is a weighting function (Section 3.7) and the factor  $f_i^{\text{eq}}$  equalizes the energy of the synthetic with that of the data:

$$f_i^{\text{eq}} = \sqrt{\frac{\int_{[l_i]} d_i^2(t) dt}{\int_{[l_i]} s_i^2(t) dt}}. \quad (8)$$

Without energy equalization, the misfit function would depend on the absolute amplitudes of the recorded and synthetic seismograms. Amplitudes of the data from the Global Seismographic Network (GSN) appear to be accurate to within a few percent (Tibuleac *et al.* 2003). Amplitudes of synthetics are proportional to the scalar moments of the events, which we take from a moment–tensor catalogue (e.g. Dziewonski *et al.* 1994). The uncertainty in the scalar-moment values in catalogues reaches tens of per cents (Helffrich 1997), so that  $F_i$  would also be typically tens of per cents. Misfit values so large, however, could be caused not only by amplitude uncertainty but also by distortion of waveforms due to scattering or noise. The value of  $F_i(\delta\beta(r))$  would not thus be a sufficient measure of the quality of the waveform fit.

The removal of the absolute amplitude by means of energy equalization makes  $F_i$  depend only on the shape of the waveforms. The shape depends on the phase and relative amplitude of the modes at frequencies within the pass band and is very sensitive to interference with scattered waves or noise. This makes

$$F_i < 0.05 \quad (9)$$

an effective selection criterion.

The computed equalization factors  $f_i^{\text{eq}}$  (8) usually turn out to be different for different windows selected for the same seismogram. This shows that errors in the synthetics' amplitude are not entirely due to scalar-moment errors—the effect of those is equivalent to

multiplication of the signal by a constant factor. The computed amplitudes of the modes are also affected by frequency-dependent errors due to uncertainties in: their excitation (that depends on the assumed Earth structure at and near the source and on the centroid–moment–tensor (CMT) solution, including centroid depth); the attenuation model assumed for the path; the computed amplification beneath the sensor (that depends on the assumed Earth structure at and near the receiver). In addition, JWKB synthetics do not account for possible focussing of seismic energy due to lateral heterogeneity around the path. Because the effect of all of these errors on the modes' phase is generally much smaller than on their amplitude, we allow variations in  $f_i^{\text{eq}}$ , within limits.

Large amplitude discrepancies can be caused by large errors in source mechanisms or path attenuation models, and such errors are likely to also substantially affect the phase of the synthetic seismograms. We use the condition

$$\frac{1}{f_{\text{max}}^{\text{eq}}} < f_i^{\text{eq}} < f_{\text{max}}^{\text{eq}}, \quad (10)$$

with an empirical  $f_{\text{max}}^{\text{eq}}$  set to 1.5–2, to exclude time–frequency windows (and whole seismograms if necessary) with amplitude discrepancy too large.

### 3.7 Weights

The global misfit function minimized in waveform fitting is a sum of the misfits within all  $N$  selected time–frequency windows normalized by a weighted data norm:

$$F_{\text{glob}}(\delta\beta(r)) = \frac{\sum_{i=1}^N (\omega_{\text{ref}}/\omega_i)^2 \int_{[l_i]} w_i(t) (d_i(t) - f_i^{\text{eq}} s_i(t))^2 dt}{\sum_{i=1}^N (\omega_{\text{ref}}/\omega_i)^2 \int_{[l_i]} w_i(t) d_i^2(t) dt}. \quad (11)$$

The ratio of a reference frequency  $\omega_{\text{ref}}$  to the centre frequency of the  $i$ th window  $\omega_i$  accounts for the fact that the phase shift at a frequency due to a small perturbation  $\delta C_m$  will be proportional to the frequency.

The contributions from different parts of the signal to  $F_{\text{glob}}$  are tuned by weights  $w_i(t)$ , each of them a product of three different weighting functions:

$$w_i(t) = w_i^{\text{env}}(t) w_i^{\text{type}}(t) w_i^{\text{err}}(t). \quad (12)$$

The purpose of the ‘envelope’ weighting  $w_i^{\text{env}}(t)$  is to prevent the wave trains with largest amplitudes (e.g. the fundamental-mode wave trains in Fig. 2) from dominating  $F_{\text{glob}}$ . This weight is initially computed at each maximum of the signal envelope  $e(t)$ . If a  $j$ th maximum within an  $i$ th window occurs at a time  $t_{ij}^{\text{max}}$ , then

$$w_i^{\text{env}}(t_{ij}^{\text{max}}) = \frac{1}{e_i(t_{ij}^{\text{max}})/e_{\text{glob}}^{\text{max}} + \epsilon^{\text{env}}}, \quad (13)$$

where  $e_{\text{glob}}^{\text{max}}$  is a global maximum of the envelope over all windows and  $\epsilon^{\text{env}}$  (we use  $\epsilon^{\text{env}} = 0.06$ ) prevents division by values close to zero.

The ‘wave-type’ weight  $w_i^{\text{type}}(t)$  is also initially computed at the envelope maxima  $t_{ij}^{\text{max}}$  and is designed to tune the contributions of different wave types to  $F_{\text{glob}}$ . Within the time–frequency windows in Fig. 2, there are five wave trains of the  $S$  wave (at 17.4, 21.7, 26.4, 31.7 and 37.4 mHz) but only one multiple- $S$  wave train (at 37.4 mHz). In this case,  $w_i^{\text{type}}(t)$  gives a five times higher weight to the multiple- $S$  wave train than to each of the  $S$  wave trains: 1 and 0.2, respectively. Multiple wave trains of the dispersive fundamental mode are also downweighted but not so strongly, to recognize

that they contain substantially different information at different frequencies. The total fundamental-mode weight is proportional to the number of its wave trains so that three wave trains have a weight of 1.4 for all of them (0.43 each), six wave trains have a weight of 2.0 (0.33 each), etc. The energy of different wave trains (depending, in particular, on their length) is also accounted for, with longer wave trains downweighted.

After the weighting functions  $w_i^{\text{env}}(t)$  and  $w_i^{\text{type}}(t)$  have been computed at all envelope maxima they are determined for signal elsewhere within the windows. The signal between a boundary of a window and the nearest envelope maximum has the same value of the weight as at that maximum. This is because we assume that this signal belongs to an arrival of the same seismic wave, so that giving additional weight to the lower-amplitude beginning or end of the wave train would not yield additional information. Instead, it could enhance the impact of possible interference with the waves just outside the window, rejected by the selection procedure and therefore to be avoided. All the windows in Fig. 2 except for the one at 37.4 mHz have constant  $w_i^{\text{env}}(t)$  and  $w_i^{\text{type}}(t)$  within them. If there is more than one envelope maximum within a window—as at 37.4 mHz in Fig. 2—the weights are interpolated between the maxima with cosine functions. The weighting thus defined aims at tuning the contribution of different waves to  $F_{\text{glob}}$ , while not increasing the impact of noise or scattered waves.

The ‘error-suppressing’ weight  $w_i^{\text{err}}(t)$  is designed to suppress the propagation of errors in the data into the tomographic model. The error due to noise is estimated using the ratio of the amplitude of the signal envelope at each point within a window to the amplitude of the envelope of noise, the latter measured in every frequency band before the  $P$ -wave arrival. In the Appendix, we discuss evaluation of an error of a different type, that in the phase of the synthetic seismogram due to the CMT uncertainty. For a data error of any type, we first estimate the error that it would translate into in a hypothetical phase-velocity measurement. This error is a function of time and frequency, and depends on the source–station distance. Such ‘phase-velocity error’ estimates are used because it is straightforward to compute their variations with time and frequency and because phase-velocity values are directly related to the waveforms (1). Using all error estimates, cumulative errors are computed as functions of time in every window, and portions of the signal with larger associated errors are downweighted.

### 3.8 Filtering, windowing, weighting: the example

The selection and weighting of the signal can be seen as an application of a single operator  $\mathbf{R}$  (3). Fig. 1 illustrates the construction of  $\mathbf{R}$ .

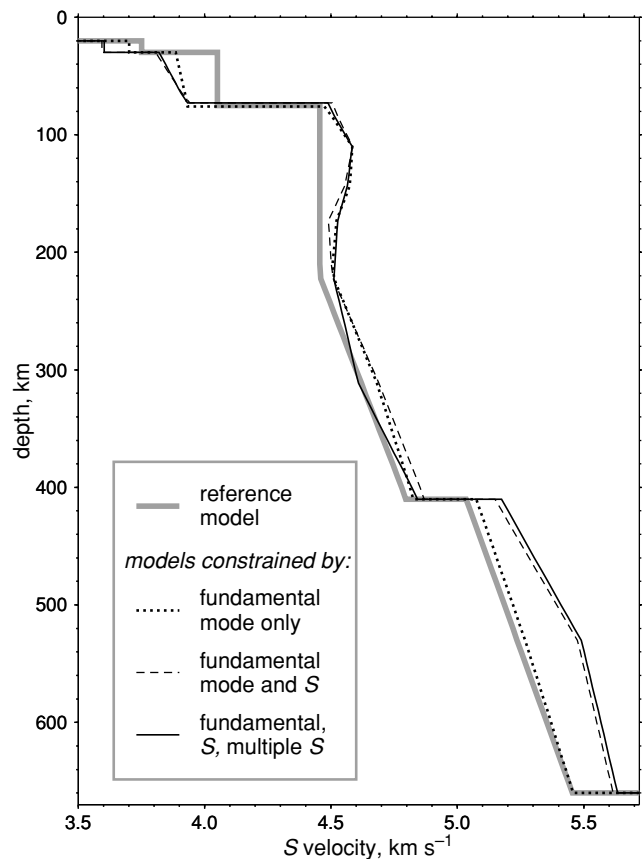
The broad-band seismogram is shown above the frames of Figs 1(a) and (d). The arrival of the  $S$  wave is followed by a wave train composed of a number of arrivals of triplicated- $S$  and multiple- $S$  waves; the dispersed fundamental Rayleigh mode (R1) arrives later. The time–frequency transform shows that below 5 mHz the seismogram is contaminated by noise. Above 5 mHz the fundamental mode is the dominant arrival, most energetic between 20 and 35 mHz. The  $S$  wave is seen the clearest between 20 and 45 mHz; the multiple- $S$  signal is weak.

Fig. 1(b) shows the time–frequency windows (same as in Fig. 2) selected according to the set of criteria we have defined. According to our assumptions, the windows isolate the portion of the signal in the time–frequency plane for which the JWKB approximation is valid.

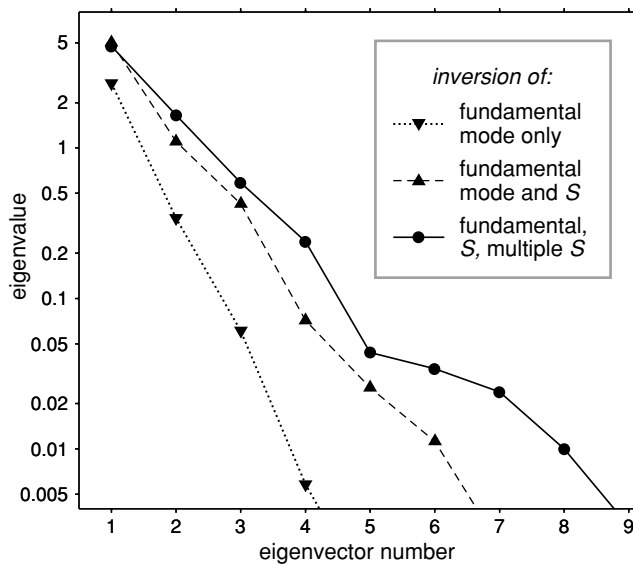
If we pass the signal in Fig. 1(a) through the filter in Fig. 1(b) and fit it with a synthetic as is, the misfit function will be dominated by the energetic fundamental mode at 20–35 mHz, with the low-frequency fundamental mode and the  $S$  wave having smaller contributions and the multiple- $S$  waves—almost no contribution. In Fig. 1(c) we multiply the filter by the weights (Section 3.7). Passed through this weighted filter, the signal from Fig. 1(a) becomes that in Fig. 1(d), with the contribution from all selected wavegroups approximately equal.

### 3.9 Result of the waveform inversion

The result of the waveform inversion of a seismogram is a set of linear equations with uncorrelated uncertainties (5), to be combined with equations obtained from all other seismograms of the data set and solved for a 3-D Earth model. For the purpose of illustration, equations yielded by the inversion of one seismogram can be solved for an (underdetermined) 1-D model that averages over the source–station path. Fig. 4 shows such 1-D models obtained from three different waveform inversions of the seismogram in Figs 1 and 2. We first invert only the fundamental mode (the signal only within the five fundamental-mode windows, Fig. 2) and find, as expected, that it constrains  $S$ -velocity anomaly down to a depth of 200 km or so. At greater depth, the sensitivity of the fundamental-mode



**Figure 4.** The reference model used for the waveform inversion and the path-average-model estimates obtained from the inversion of all (solid line) or subsets (dashed lines) of the data within the selected time–frequency windows (Figs 1 and 2). The models are solutions of systems of linear equations—underdetermined for a single path—yielded by the waveform inversion.



**Figure 5.** The eigenvalues corresponding to the linear equations yielded by the waveform inversions of all (solid line) or subsets (dashed lines) of the data within the selected time–frequency windows (Figs 1 and 2).

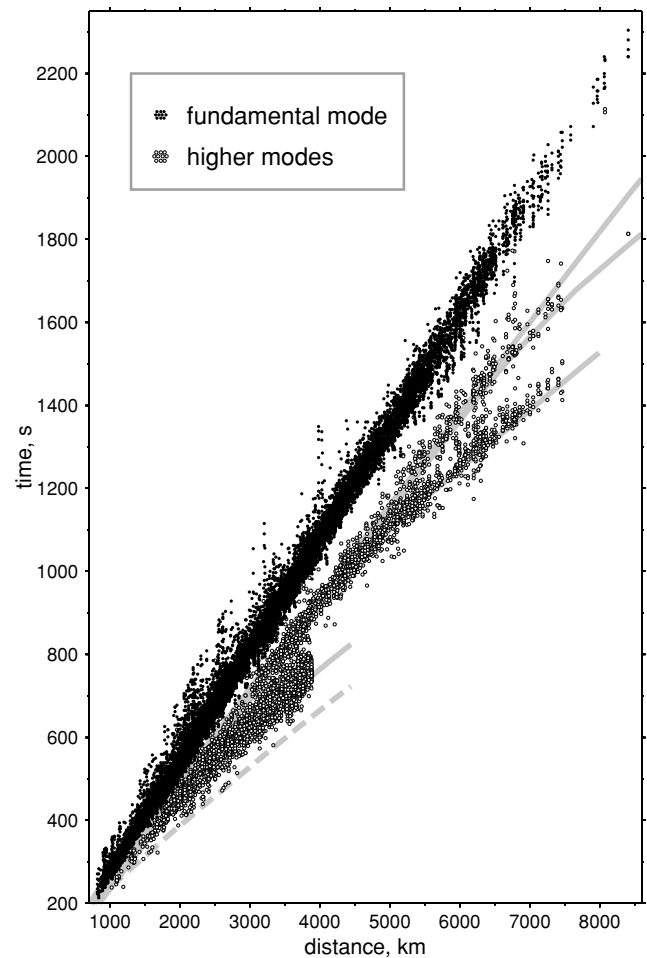
wave trains decreases and cannot provide any vertical resolution; the weak anomaly seen in the fundamental-mode model below 200-km depth is simply smeared from above. When we also include the *S* wave (the *S*-wave windows at 17.4, 21.7, 26.4, and 31.7 mHz) we extend the sensitivity of our data down to the transition zone. Finally, by including the multiple-*S* window as well (37.4 mHz) we are able to distinguish the anomaly between 200 and 400 km from that above and below. This increase in the amount of information given by additional data types can also be seen in the increase of the eigenvalues corresponding to the linear equations yielded by the waveform inversions (Fig. 5). Figs 4 and 5 illustrate the importance of careful weighting of the different parts of the signal: without it, the multiple-*S* fit would contribute very little to the misfit function and the information carried by these waves would be lost.

#### 4 APPLICATION OF AMI TO WESTERN PACIFIC DATA

The first application of AMI has been to thousands of vertical-component seismograms recorded in the Western Pacific and South-east Asia. The resulting high-resolution, *S*-velocity model has been described and interpreted by Lebedev & Nolet (2003). Here, we focus on the results of the automated signal selection in order to verify the consistency of the procedure and to validate the assumptions it is based on.

According to the selection criteria, a seismogram was included in the data set only if an acceptable fit could be obtained in three or more consecutive fundamental-mode windows, with higher mode wave trains optional. Within the windows selected for the 4038 accepted seismograms, there are 18 701 fundamental-mode and 7855 higher mode wave trains.

In Fig. 6 we plot the arrival times of all the wave trains as a function of the source–station distance. An arrival time is measured at every maximum of the signal envelope within every window. As expected, fundamental-mode arrivals plot approximately along a straight line in the distance–time diagram. Late arrivals at distances below

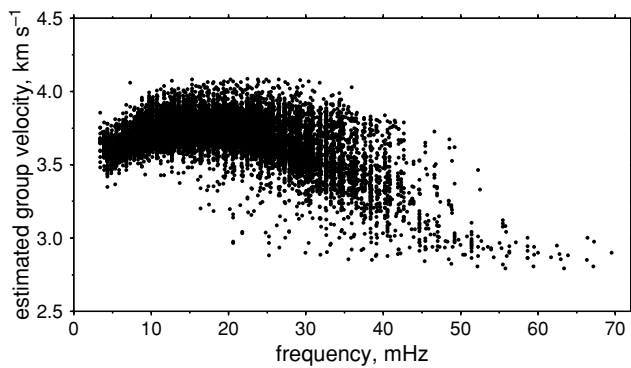


**Figure 6.** Arrival times of the 18 701 fundamental- and 7855 higher-mode wave trains, all successfully matched with JWKB synthetics. Arrival times are measured at maxima of the signal envelope. Traveltime curves of the *S*, *SS*, *SSS* and *SSSS* waves for a surface source in AK135 (Kennett *et al.* 1995) are shown with solid grey lines; the *S*-wave curve for a 650-km-deep source is shown with a dashed line.

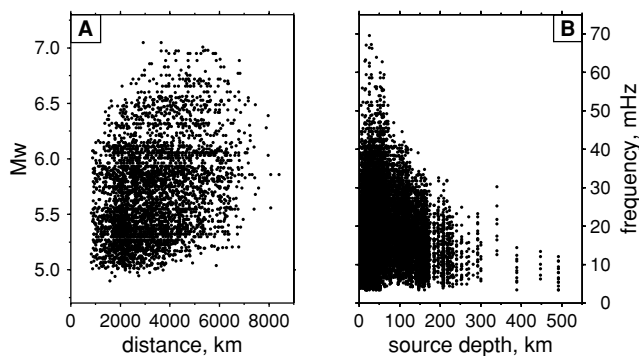
4000 km are those of the relatively slow, higher frequency wave trains (see also Fig. 7).

Higher mode arrivals form a likeness of the branches of the *S*, *SS* and *SSS* waves. The variability in the arrival times is due not only to lateral heterogeneity in the Earth but also to variability in the earthquake depths and to the complexity of the interference patterns of bandpassed triplicated waves. The *S* waves are cut out at distances exceeding 35°, as intended (Section 3.3).

A different perspective on the fundamental-mode arrivals is given by their distribution in the frequency–group-velocity plane (Fig. 7). A group velocity estimate is yielded by the arrival time of the fundamental-mode wave train divided by the distance; this is not an exact value because the influence of the source mechanism on the group arrival time is neglected (Levshin *et al.* 1999). Below 10–15 mHz, all group velocities are relatively large; their average increase with frequency is due to the increase of their average sensitivity to the typically high *S* velocities within the lithosphere. Extrema of group-velocity curves correspond to Airy phases in the seismograms; Airy phases occur often near 15–20 mHz. At 20–40 mHz, the group-velocity variability increases, while average values decrease. Above 50 mHz group velocities fall, most of them to below 3 km s<sup>−1</sup>: Rayleigh waves at these frequencies sample primarily the crust.



**Figure 7.** Group-velocity estimates from 18 701 fundamental-mode wave trains. As well as arrival times in Fig. 6, the estimates are by-products of waveform inversion, used only to examine the overall results of the data processing.



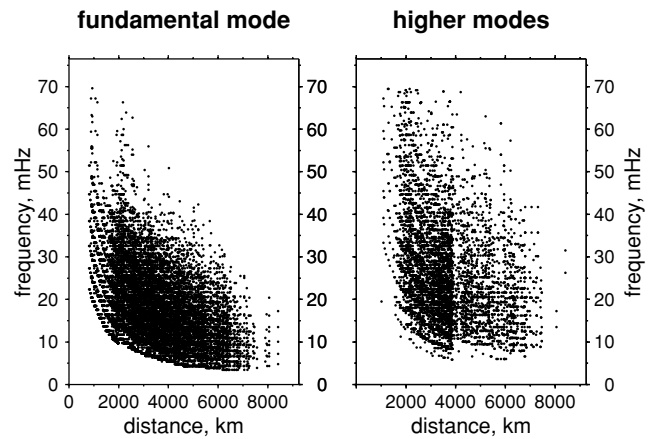
**Figure 8.** (a) Distribution of the event moment magnitudes vs. source-station distances for the 4038 seismograms with any signal selected. (b) Distribution of the source depths vs. filters' centre frequencies for the 18 701 fundamental-mode wave trains selected.

In Fig. 8(a) we plot a dot for each of the 4038 seismograms with any signal selected, in the plane of the source–station distances and the events' moment magnitudes  $M_W$ . No dots appear in the upper-left corner of the frame. This is because large events are associated with large duration times, so that higher frequency signal is ruled out by our point-source-approximation criterion, while lower-frequency signal at short distances is ruled out by the far-field-approximation criterion (Section 3.3). Lower-right corner of the frame is also empty: signal from small events is drowned in noise at long distances. The boundary of this empty corner gives a magnitude cut-off as a function of distance; this can be used for an initial selection of seismograms in Rayleigh wave studies.

The maximum frequency of the automatically identified fundamental-mode windows decreases with increasing source depth (Fig. 8b) as expected: deep events excite only long-period fundamental mode.

In Fig. 9, we plot centre frequencies of the selected fundamental- and higher mode windows as a function of distance. The lower-left corners of the frames are empty, in agreement with the far-field-approximation criterion (Section 3.3). In the higher mode frame, the decrease in the density of dots just below 4000-km distance is due to the  $S$ -wave cut-off at  $35^\circ$  (Section 3.3).

The most interesting pattern in Fig. 9 is the decrease in the maximum frequency of the windows with distance. According to our assumptions, the windows contain only the signal that can be modelled with the JWKB approximation and, therefore, the distributions



**Figure 9.** Distributions of source–station distances versus filters' centre frequencies for the fundamental- and higher mode wave trains selected.

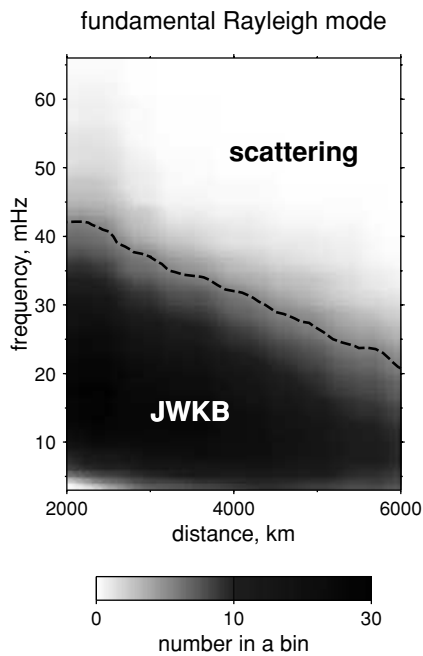
in Fig. 9 can be viewed as an empirical mapping of the field of validity of the approximation.

As a surface wave propagates further and further from the source, it experiences more and more interactions with the heterogeneity within the Earth's crust and mantle. The proportion of the wave's energy likely to have been scattered increases. For a given point on the surface (e.g. the location of a receiver recording a seismogram) it implies that at a longer distance more scattered waves are likely to pass through the point and interfere with the direct wave. The likelihood of the validity of the JWKB approximation is thus expected to decrease with distance, the decrease being faster at higher frequencies due to both shorter wavelength of the waves and higher heterogeneity in the shallower layers of the Earth which higher frequency waves sample. The pattern in Fig. 9 is consistent with these expectations. For example, for the 40-mHz (25-s) fundamental mode the approximation is often valid at 2000-km distance, rarely valid at 4000 km, and almost never valid at distances exceeding 6000 km.

In a particular case of a wave that traverses first a heterogeneous region and then a homogeneous one, the accuracy of the JWKB approximation may increase with distance, a phenomenon we could call 'waveform healing' by analogy to the related wavefront healing. Such cases, however, cannot be expected to change the dominant pattern of the decrease of the approximation accuracy with frequency.

Apart from the validity of (1), the distribution of the points in Fig. 9 is also affected by our chosen implementation of the relation (2), including the computation of the Fréchet derivatives and the parametrization of  $\delta\beta(r)$ . For some source–station paths, for example those sampling strong crustal heterogeneity, our 1-D parametrization may be insufficient and the Fréchet derivatives inaccurate for higher frequency fundamental mode. This probably decreases the point density at higher frequencies. Nevertheless, the fact that no points at all appear in the upper-right half of the frame in Fig. 9(a), in spite of thousands of times that waveform fits were attempted at these frequencies and distances, shows the robustness of the main pattern—a decrease with distance of the likelihood of the validity of (1)—while also verifying the absence of chance fits (of noise or scattered waves).

Figs 6–9 display the results of a fully automated window selection. The distribution patterns match the selection criteria as well as general expectations, confirming the accuracy and consistency of the automated processing.



**Figure 10.** An estimated empirical field of validity of the JWKB approximation for the fundamental Rayleigh mode. Point density of the distribution in Fig. 9(a) is plotted using the number of points in distance–frequency bins.

## 5 DISCUSSION

The validity of the JWKB approximation is warranted only in the limit of a smooth lateral heterogeneity in elastic parameters (Dahlen & Tromp 1998). In the Earth's upper mantle (notably, in subduction zones) heterogeneity is known to be rough—as measured against the Fresnel-zone widths (Wang & Dahlen 1995) of long- and intermediate-period surface waves. For modelling waves that may sample heterogeneous regions, the validity of the JWKB approximation is not warranted, that is, the approximation may or may not be valid for a given time–frequency portion of a given seismogram.

The accuracy of (1) for a given wave train will depend on the way that the modes that contribute to it (one or more of them) sample heterogeneity within the Earth. That, in turn, depends on the locations of the source and the station, on the modes' frequency, and on their frequency-dependent radiation patterns. AMI enables case-by-case selection of the time–frequency portions of seismograms that can be accurately modelled using eqs (1) and (2).

In Fig. 10 we plot the distribution of the fundamental-mode fits using the point density (hit count in distance–frequency bins) from Fig. 9(a) and map the fields of the JWKB regime (the approximation is likely to be valid) and the scattering regime (the approximation is unlikely to be valid). There is no sharp boundary between the two, with the likelihood decreasing gradually with increasing distance and frequency. Kennett & Nolet (1990) modelled wave propagation over 3350-km distance in realistic upper-mantle models and inferred a threshold of 20 mHz for the validity of the JWKB approximation for the fundamental mode. In Fig. 10, the density (hit count) at 3350 km starts decreasing above 20–25 mHz (so does the likelihood of the approximation validity), consistent with this being a conservative threshold value. In many cases, however, the approximation will still be valid well above this threshold, and in other cases it will breakdown below the threshold. The most attractive solution is thus a case-by-case signal selection—as implemented here.

Inverting one seismogram with AMI takes on the order of 1 min on a typical modern processor. Rejecting seismograms too noisy (of which there often is a majority) takes on the order of 1 s each. Using a few processors, one can process a data set of tens of thousand seismograms in a matter of weeks.

AMI is formulated here using (1) and (2), but the signal selection procedure will be effective in more theoretically elaborate schemes as well, including, for example, waveform inversions using 3-D waveform sensitivity kernels (e.g. Meier *et al.* 1997; Lebedev & Nolet 1998). Even in inversions of scattered waves as such, an initial ray-theoretical estimation of the smooth (or of a path-averaged) structure of the medium (Meier *et al.* 1997; Friederich 2003) can be done quickly and accurately with AMI. And the selection of the scattered waves to be inverted with a first-order approximation (neglecting multiple scattering) can be formulated as an extension of the window-selection procedure presented here, although new criteria will need to be developed.

## 6 CONCLUSIONS

The Automated Multimode Inversion of surface- and *S*-wave forms enables a fast and accurate extraction of phase information from regional and teleseismic seismograms. As in Nolet (1990), the information is output in the form of linear equations with uncorrelated uncertainties. The equations are suitable to constrain high-resolution tomographic models on a regional (Lebedev & Nolet 2003) and potentially on a global scale.

AMI fits observed seismograms with JWKB synthetics and identifies time–frequency windows containing the signal that can be modelled accurately with the JWKB approximation. The rigorous window selection procedure allows neither chance fits nor waveform fits with cycle skips.

Elaborate weighting of the contributions of waves of different amplitudes and different types to the misfit functions ensures that the information on deep upper-mantle structure carried by less energetic waves (*S*, multiple-*S*, very-long-period fundamental mode) is extracted fully. Weighting can also suppress the propagation of data errors into the tomographic model.

## ACKNOWLEDGMENTS

We gratefully acknowledge numerous discussions with F. A. Dahlen. Insightful comments from two anonymous referees and associate editor Gabi Laske have helped us to improve the clarity of presentation and are greatly appreciated. Figures were generated with Generic Mapping Tools (Wessel & Smith 1995). Waveform data was obtained from facilities of the Incorporated Research Institutions for Seismology (IRIS) Data Management System (DMS), funded through the National Science Foundation (NSF) GEO Directorate, Instrumentation and Facilities Program, under Cooperative Agreement EAR-0004370. Global Seismographic Network (GSN) is operated jointly by the IRIS, the United States Geological Survey (USGS), and the NSF. This work was supported by the NSF through the grant EAR0309298 awarded to Guust Nolet and—at the final stage—by the David and Lucile Packard Foundation through a Fellowship awarded to RDvdH.

## REFERENCES

Aki, K. & Richards, P.G., 1980. *Quantitative seismology*, Freeman, New York.

- Beucier, E., Stutzmann, E. & Montagner, J.P., 2003. Surface wave higher-mode phase velocity measurements using a roller-coaster-type algorithm, *Geophys. J. Int.*, **155**, 289–307.
- Cara, M. & Lévêque, J.J., 1987. Waveform inversion using secondary observables, *Geophys. Res. Lett.*, **14**, 1046–1049.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical global seismology*, Princeton University Press, Princeton.
- Debayle, E., 1999. SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion, *Geophys. J. Int.*, **137**, 747–754.
- Debayle, E. & Kennett, B.L.N., 2000. The Australian continental upper mantle: structure and deformation inferred from surface waves, *J. geophys. Res.*, **105**, 25 423–25 450.
- Dziewonski, A.M. & Hales, A.L., 1972. Numerical analysis of dispersed seismic waves, in *Methods in Computational Physics*, Vol. 11, ed. Bolt, B.A., Academic Press, New York.
- Dziewonski, A.M. & Steim, J.M., 1982. Dispersion and attenuation of mantle waves through waveform inversion, *Geophys. J. R. astr. Soc.*, **70**, 503–527.
- Dziewonski, A.M., Ekström, G. & Salganik, M.P., 1994. Centroid-moment tensor solutions for January–March 1994, *Phys. Earth Planet. Inter.*, **86**, 253–261.
- Friederich, W., 2003. The S-velocity structure of the East Asian mantle from inversion of shear and surface waveforms, *Geophys. J. Int.*, **153**, 88–102.
- Gaherty, J.B., Kato, M. & Jordan, T.H., 1999. Seismological structure of the upper mantle; a regional comparison of seismic layering, *Phys. Earth planet. Inter.*, **110**, 21–41.
- Gee, L.S. & Jordan, T.H., 1992. Generalized Seismological Data Functionals, *Geophys. J. Int.*, **111**, 363–390.
- Helffrich, G.R., 1997. How good are routinely determined focal mechanisms? Empirical statistics based on a comparison of Harvard, USGS and ERI moment tensors, *Geophys. J. Int.*, **131**, 741–750.
- Kennett, B.L.N. & Nolet, G., 1990. The interaction of the S-wavefield with upper mantle heterogeneity, *Geophys. J. Int.*, **101**, 751–762.
- Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, **122**, 108–124.
- Knopoff, L. & Schwab, F.A., 1968. Apparent initial phase of a source of Rayleigh waves, *J. geophys. Res.*, **73**, 755–760.
- Lebedev, S. & Nolet, G., 1998. Sensitivity of multimode surface waves and the scale of heterogeneity in the Western Pacific upper mantle, *EOS, Trans. Am. geophys. Un.*, **79** (17), Spring Meet. Suppl., S206.
- Lebedev, S. & Nolet, G., 2003. Upper mantle beneath southeast Asia from S velocity tomography, *J. geophys. Res.*, **108**, doi:10.1029/2000JB000073.
- Lebedev, S., Chevrot, S. & van der Hilst, R.D., 2002. Seismic evidence for olivine phase changes at the 410- and 660-kilometer discontinuities, *Science*, **296**, 1300–1302.
- Levshin, A.L., Ritzwoller, M.H. & Resovsky, J.S., 1999. Source effects on surface wave group travel times and group velocity maps *Phys. Earth planet. Inter.*, **115**, 293–312.
- Masters, G., MINOS. at: Reference Earth Model web page: <http://mahi.ucsd.edu/Gabi/rem.dir/surface/minos.html>.
- Megnin, C. & Romanowicz, B., 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *J. geophys. Res.*, **105**, 709–728.
- Meier, T., Lebedev, S., Nolet, G. & Dahlen, F.A., 1997. Diffraction tomography using multimode surface waves, *J. geophys. Res.*, **102**, 8255–8267.
- Meier, T., Dietrich, K., Stockhert, B. & Harjes, H.P., 2004. One-dimensional models of shear wave velocity for the eastern Mediterranean obtained from the inversion of Rayleigh wave phase velocities and tectonic implications, *Geophys. J. Int.*, **156**, 45–58.
- Muyzert, E. & Snieder, R., 1996. The influence of errors in source parameters on phase velocity measurements of surface waves, *Bull. seism. Soc. Am.*, **86**, 1863–1872.
- Nolet, G., 1990. Partitioned waveform inversion and two-dimensional structure under the Network of Autonomously Recording Seismographs, *J. geophys. Res.*, **95**, 8499–8512.
- Nolet, G., van Trier, J. & Huisman, R., 1986. A formalism for non-linear inversion of seismic surface waves, *Geophys. Res. Lett.*, **13**, 26–29.
- Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 2004. Global transition zone tomography *J. geophys. Res.*, **109**, doi:10.1029/2003JB002610.
- Simons, F.J., van der Hilst, R.D., Montagner, J.P. & Zielhuis, A., 2002. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle, *Geophys. J. Int.*, **151**, 738–754.
- Tibuleac, I.M., Nolet, G., Michaelson, C. & Koulakov, I., 2003. P wave amplitudes in a 3-D earth, *Geophys. J. Int.*, **155**, 1–10.
- Takeuchi, H. & Saito, M., 1972. Seismic surface waves, in *Methods in Computational Physics*, Vol. 11, pp. 217–295, ed. Bolt, B.A., Academic, San Diego, California.
- van Heijst, H.J. & Woodhouse, J.H., 1997. Measuring surface-wave overtone phase velocities using a mode-branch stripping technique, *Geophys. J. Int.*, **131**, 209–230.
- Wang, Z. & Dahlen, F.A., 1995. Validity of surface-wave ray theory on a laterally heterogeneous earth, *Geophys. J. Int.*, **123**, 757–773.
- Wessel, P. & Smith, W.H.F., 1995. New version of the Generic Mapping Tools released, *EOS, Trans. Am. geophys. Un.*, **76**, 329.
- Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89**, 5953–5986.
- Yoshizawa, K. & Kennett, B.L.N., 2002. Non-linear waveform inversion for surface waves with a neighbourhood algorithm—application to multimode dispersion measurements, *Geophys. J. Int.*, **149**, 118–133.
- Yoshizawa, K. & Kennett, B.L.N., 2004. Multimode surface wave tomography for the Australian region using a three-stage approach incorporating finite frequency effects, *J. geophys. Res.*, **109**, doi:10.1029/2002JB002254.
- Zielhuis, A. & Nolet, G., 1994. Shear wave velocity variations in the upper mantle beneath central Europe, *Geophys. J. Int.*, **117**, 695–715.

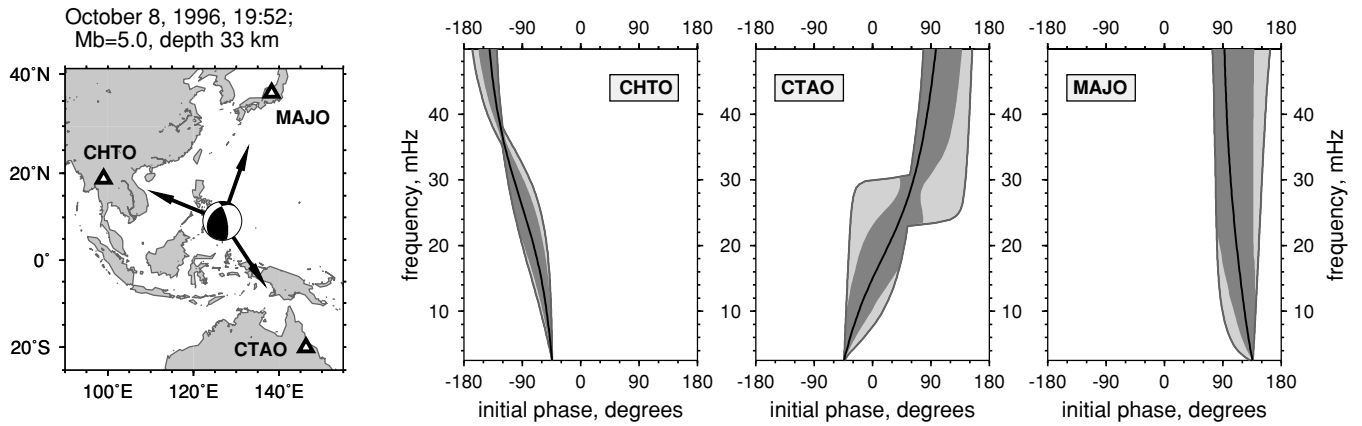
## APPENDIX A: INITIAL-PHASE ERRORS

The phase of a recorded surface wave is a sum of the initial phase at the source and a phase shift accumulated over the source–station path (e.g. Knopoff & Schwab 1968). The propagation phase delays contain information on Earth structure that waveform inversion techniques—AMI included—aim to extract. This information, however, is only as accurate as the values of the initial phase used to isolate the propagation delays.

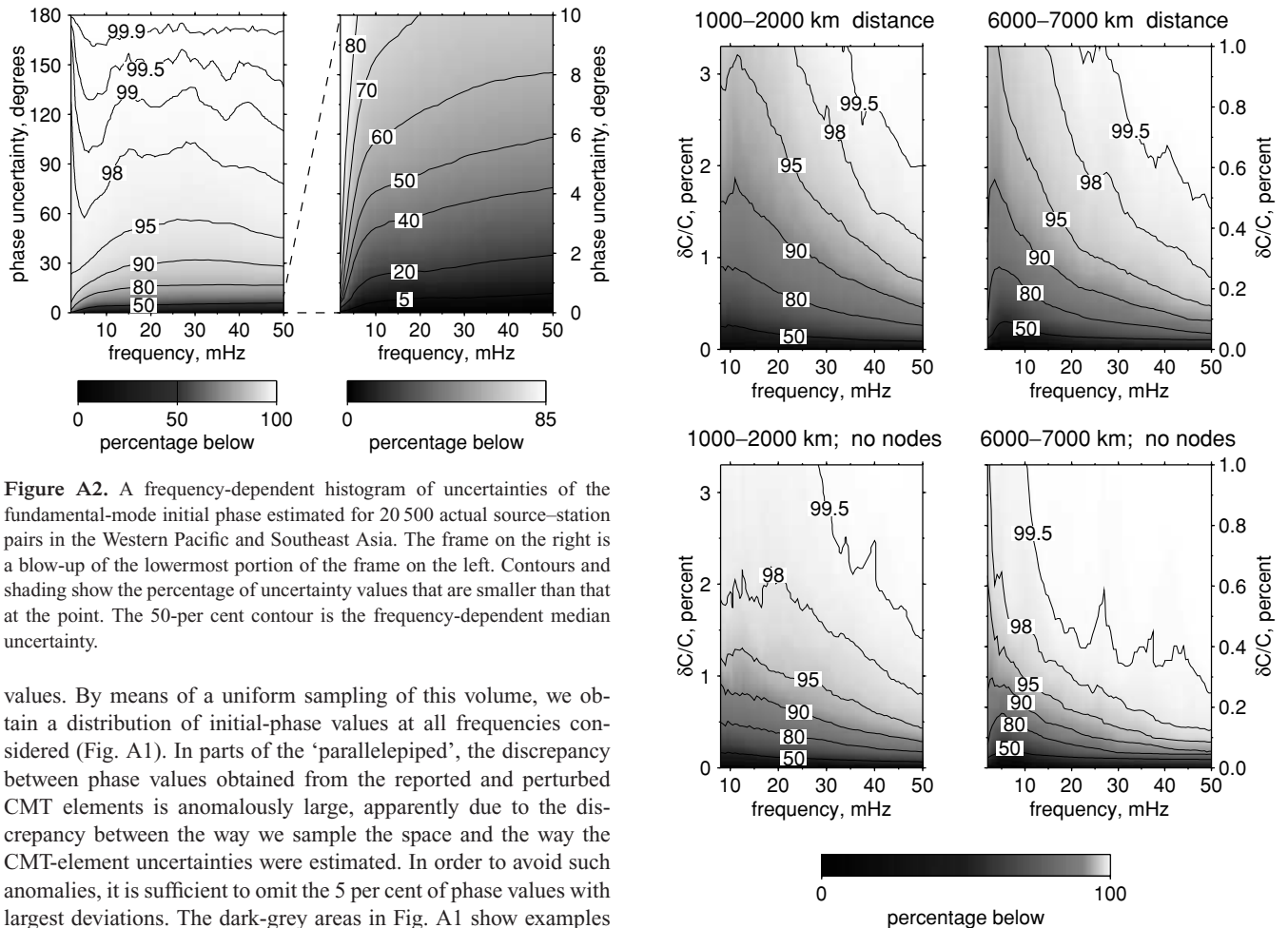
A CMT solution for the source mechanism of an event (e.g. Dziewonski *et al.* 1994) allows to compute the initial phase of any mode at any frequency and azimuth (in the relation (1) the complex amplitude  $A_m(\omega)$  contains the initial phase information). Through this calculation, moment tensors taken from catalogues enter waveform inversions as a type of data. Uncertainties of the CMT solutions translate into uncertainties of measurements, and CMT errors can propagate into the resulting models. Here we evaluate the magnitude of the uncertainties and propose to suppress their propagation using error-dependent weights.

Muyzert & Snieder (1996) suggested to compute derivatives of the phase with respect to source parameters and discard source–station pairs if the derivatives turned out to be above a threshold. This assumes that the uncertainties of the source parameters (which would multiply the derivatives to yield phase-error estimates) are the same in every case. In reality, some source mechanisms are constrained better than others, depending, in particular, on the location and size of the event and the related amount and quality of the data available to constrain the CMT.

Here we estimate the initial-phase uncertainties directly, using the uncertainties reported for the six elements of the Harvard CMT. In the 6-D space of the moment–tensor elements, the values of their uncertainties define a ‘parallelepiped’, centred at the reported CMT



**Figure A1.** The locations of a source and three stations (left) and the frequency-dependent uncertainties in the initial phase of the fundamental mode estimated at azimuths towards each of the stations. Bold solid lines show the initial phase at the source computed from the six elements of the Harvard CMT. Surrounding light-grey bands cover all the values of the initial phase allowed by the reported uncertainties in the CMT elements. Dark-grey bands sample all but 5 per cent of the 6-D volume defined by the CMT uncertainties, excluding anomalously large phase deviations (see text).



**Figure A2.** A frequency-dependent histogram of uncertainties of the fundamental-mode initial phase estimated for 20 500 actual source-station pairs in the Western Pacific and Southeast Asia. The frame on the right is a blow-up of the lowermost portion of the frame on the left. Contours and shading show the percentage of uncertainty values that are smaller than that at the point. The 50-per cent contour is the frequency-dependent median uncertainty.

values. By means of a uniform sampling of this volume, we obtain a distribution of initial-phase values at all frequencies considered (Fig. A1). In parts of the ‘parallelepiped’, the discrepancy between phase values obtained from the reported and perturbed CMT elements is anomalously large, apparently due to the discrepancy between the way we sample the space and the way the CMT-element uncertainties were estimated. In order to avoid such anomalies, it is sufficient to omit the 5 per cent of phase values with largest deviations. The dark-grey areas in Fig. A1 show examples of the resulting frequency-dependent estimates of the initial-phase uncertainty.

Computing the uncertainties for 20 500 actual source-station pairs in the Western Pacific and Southeast Asia, we observe that a typical initial-phase uncertainty is a few degrees (Fig. A2), although in a few per cent of cases it reaches tens of degrees.

**Figure A3.** Frequency-dependent histograms of estimated errors in hypothetical fundamental-mode phase-velocity measurements. Only values from source-station pairs with distances in the 1000–2000 and 6000–7000 km ranges are included in the frames on the left and right, respectively. Top: estimates at all frequencies for each path within a range. Bottom: near-nodally radiated waves (Section 3.1) are excluded.

At a given frequency and distance, the initial-phase uncertainty translates into an uncertainty of a hypothetical phase-velocity measurement, a quantity more relevant for tomographic imaging. In Fig. A3 we plot the relative phase-velocity uncertainties  $\delta C/C$  in a regional (1000–2000 km) and a teleseismic (6000–7000 km) distance ranges. In the top frames, we include all source–station pairs within the distance ranges. In the bottom frames, we exclude about 28 per cent of data that is radiated at near-nodal azimuths, according to the criterion from Section 3.1. Near-nodal radiation is clearly associated with much larger errors of phase-velocity measurements. Consistent removal of near-nodal data is an effective way to reduce the potential impact of initial-phase errors.

At short distances, the uncertainty at 50–100 s even away from the nodes exceeds 1 per cent in over 5 per cent cases. The propagation of the remaining errors into the tomographic model can be suppressed through error-dependent weights, computed as functions of the error estimates so that the signal is downweighted if larger errors are associated with it.

Errors in event locations and centroid origin times can be treated in the same fashion, although good estimates of these are rare. An attractive solution would be to update the source parameters (including the CMTs and centroid locations and origin times) using high-resolution models of global heterogeneity; this is to be the focus of future work.