INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/480-7

WORKSHOP ON EARTHQUAKE SOURCES & REGIONAL LITHOSPHERIC STRUCTURES FROM SEISMIC WAVE DATA

19 - 30 November 1990

Nonlinear Inversion of Seismic Waveforms

Guust Nolet

University of Utrecht
Dept. of Theoretical Geophysics
Utrecht, The Netherlands

NONLINEAR INVERSION OF SEISMIC WAVEFORMS

Guust Nolet

Lecture notes, Trieste 1990

Abstract

Only about 5% of the Earth's surface has a dense enough coverage with seismic stations and a high enough level of seismicity to make delay-time tomography of the Upper Mantle possible. On the one hand, we may try to improve this situation by increasing the number of seismic sensors, especially in underdeveloped countries and in oceanic areas. But for the near future, we must try to image the Upper Mantle structures using more of the information contained in seismograms that are already available.

Waveforms and wave arrival times of the fundamental and higher mode surface waves are sensitive to Upper Mantle shear velocity structure and can be used to image this structure. For high frequency waves (> 10 mHz) fully nonlinear inversion is much more powerful than repeated linearized inversion. With present-day availability of broad-band seismic records the station separation is the limiting factor for the resolution. Contrary to delay-time tomography, an acceptable depth resolution is also obtained in regions of weak or absent seismicity.

Introduction

Everyone of us would like to take a look deep inside the Earth - except perhaps Edgar Allen Poe, who wrote a sonnet on Science and accuses it to rob us of our romantic illusions:

Science! True daughter of old time thou art
Who alterest all things with thy peering eyes
Why preyest thou thus upon the poet's heart
Vulture, whose wings are dull realities?

Probably also except captain John Cleves Symmes, who proposed to the American Congress as late as 1825 to send an expedition to the hollow interior of the Earth by way of a supposed hole near the North Pole. The idea of a hollow Earth remained very much alive until the end of the last century and inspired Jules Verne, who usually cannot be accused of ignoring valid scientific facts, to write his 'Voyage au centre de la Terre' (figure 1).

Je m'imaginais voyager à travers un diamant (p. 184).

Figure 1: Jules Verne imagined the Earth as hollow and habitable.

Since the turn of this century, seismology has given us a reasonably good idea of the average structure of our planet. Unfortunately, the romantic pictures that illustrated Jules Verne's story were almost immediately replaced by *graphs* of seismic velocity versus depth. Interesting for the few who are happy enough to stick to the facts, but one cannot help to sympathize with Edgar Allan Poe. The scientific mind had gained over the romantic instinct.

With hindsight we may now suspect that both the romantic and the somewhat puristic scientist were wrong. The romantic because he drew pictures where no factual evidence was available. The scientist because he used Ockham's razor to strip the pictures of anything that was not warranted by the data available. And since the data were insufficient to reveal three-dimensional structure, seismologists reduced the Earth to a one-dimensional *model*. This is, in my opinion, one of the main reasons that seismology has been out of

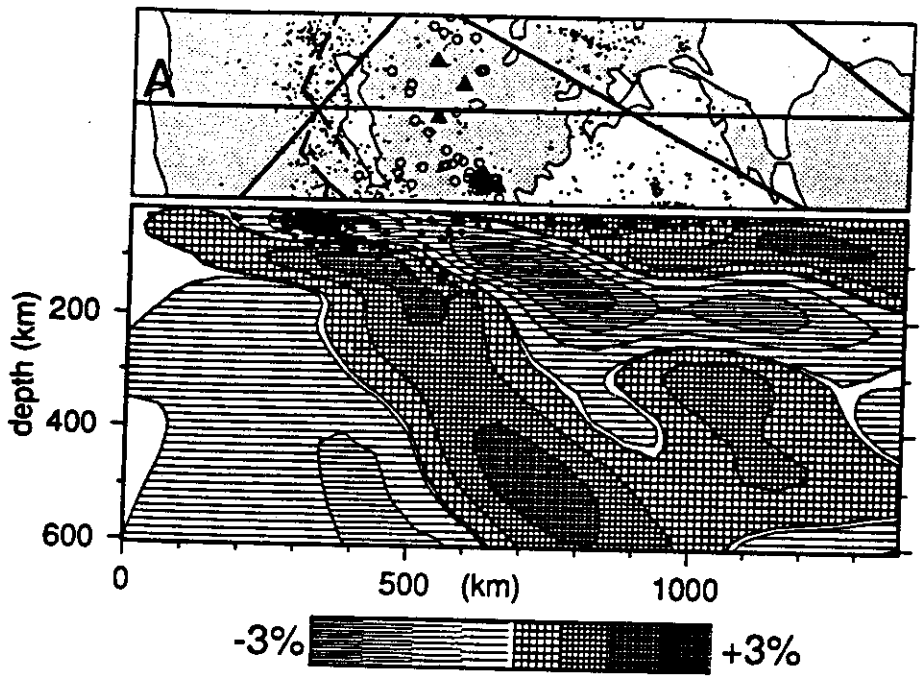

Figure 2: A tomographic image from ISC P-delays of the Hellenic subduction zone. Along a cross-section given by the line labeled A, deviations of the P-wave velocity with respect to the Jeffreys-Bullen model are mapped. Seismicity is indicated by the black dots (Spakman et al., 1988).

touch with geology for a long time.

Fairly recently, however, we have witnessed a change of scene. Suddenly seismologists were producing three-dimensional pictures, like the one in figure 2. Because such images are like a slice, cut through the Earth, the procedure to obtain them is often called 'seismic tomography', a term borrowed from the medical practice. Figure 2 was obtained by measuring 500,550 arrival times of P-waves (as published in the bulletins of the International Seismological Centre) and inverting these for the velocity structure of the Earth using a linearized relationship (Nolet, 1987) for the delay times δT_i:

$$\delta T_i = T_i - T_i^0 = \int_{S_i} \frac{ds}{c} - \int_{S_i^0} \frac{ds}{c_0} \approx \int_{S_i^0} \left(\frac{1}{c} - \frac{1}{c_0}\right) ds$$

or

$$\delta T_i = -\int_{S_i^0} \frac{\delta c(\mathrm{r})}{c_0(\mathrm{r})^2} ds \qquad (1)$$

where $\delta c = c - c_0$ s the perturbation of the seismic velocity c with respect to a starting model c_0, S_i is the ray trajectory of ray i, and T_i^0 is the calculated travel time for the starting model. If we choose the starting model symmetric, this makes our calculation very much more efficient. Note that we have used Fermat's principle to substitute the ray path S_i^0as calculated for the starting model for the (unknown) ray path in the true Earth: we only make a second order error in calculating the time delay.

The general procedure to solve systems like (1) is to parametrize the model $\delta c(\mathrm{r})$ in some way and to use an iterative matrix solver. For details we refer to Nolet (1987a). There are many problems with this approach. First of all, the delay times are often highly erroneous. Secondly, the coverage with seismic ray paths is usually insufficient to

determine $c(\mathbf{r})$ with a high resolution. Consequently, we need many (redundant) data to allow the errors to average out, and we need a robust and stable solver for the matrix system (1). It turns out that we can obtain a useful resolution only in those areas where we have both a high density of seismic stations *and* a high, preferably also deep, seismicity. Only very few areas on Earth qualify for such high-resolution tomography. An optimistic estimate is that at most 5% of the Upper Mantle can be studied using delay time tomography. Not even the very highly active subduction zones qualify easily: there are generally no seismic stations on the oceanic side of the zone, whereas the stations on the other side are often concentrated on island arcs - and the resulting unbalanced distribution of rays is very difficult to handle.

One very attractive possibility is to use hydrophones attached to freely floating sonobuoys for the recording of P-waves in oceanic areas. Both the communication with the buoys and the determination of their location can be done using polar-orbiting satellites. Satellite transmission has already been used with land based data collection platforms (Poupinet, 1987), but measurements in the ocean pose some very serious noise problems. These are directly associated with the wave motion at the water surface, which is 2-3 meters in amplitude under moderate conditions (wind 5 Beaufort), and with the natural pressure gradient in the water: a vertical motion of a hydrophone of only 1 mm causes a pressure fluctuation of 10 Pa, roughly what we would expect for a P-wave of a magnitude 6 event at teleseismic distances. Although there is a reasonable hope that the technical problems can be overcome and that a useful signal-to-noise ratio can be obtained (Nolet, 1988), this is obviously a long term goal.

What, then, are the possibilities for seismic tomography using presently available data? Obviously, a seismogram contains much more information than the arrival time of the first wavelet, and we should wish to include the whole seismic time series in the data set, and use this in an imaging technique. There are two obstacles to such an approach. The first is the enormity of the numerical problem. The second is that it is often difficult or impossible to develop a manageable approximation to the solution of the wave equation, especially one that is adequate enough to model the response of complicated media.

This lecture shall deal in depth with the first problem. I do not claim to have solved the second problem. But reasonable approximations to the exact theory of wave propagation are available for a number of possible situations:

- We may use the WKBJ approximation (Chapman, 1978) to compute simple body wave synthetic seismograms.
- We may use a great-circle approximation (Woodhouse, 1974; Jordan, 1978) to compute low-frequency surface wave arrivals. We may even include the focussing (Woodhouse and Wong, 1986; Snieder, 1988) due to lateral heterogeneity. Low-frequency S waves can be computed accurately if many modes are summed.
- We may use the Born approximation to compute scattered surface waves in a heterogeneous Earth (Snieder and Nolet, 1987).

In this lecture we shall restrict ourselves to the computation of low frequency surface waves using the great-circle approximation. We summarize the formalism in Appendix A. Using this method, we shall investigate how to find an Earth model that predicts waveforms with an aceptable fit to the observed signals. Especially, we shall show how to interprete a large number of waveform observations in terms of a 2D or 3D model of Upper Mantle heterogeneity. The key to an efficient inversion is the introduction of a very specific parametrization of the Earth - we shall intially describe the Earth in terms of velocity averages along wavepaths. We shall show how we can obtain uncorrellated estimates for such averages and their variance. A second step then translates these averages in terms of a 2D or 3D model of the S-velocity in the Earth.

Path-integral representation of the Earth model

We shall assume that the surface waves follow narrowly defined paths along the surface with the energy spread out in the depth direction. In developing the theory we shall assume these paths to be of infinitesimally small width, although in principle the paths can have a finite width. The essential point of our approach here is that there is a well defined region that influences a particular seismic time series, and that the influence of the the remaining Earth upon this particular datum can be ignored. In the interest of clarity, we shall assume the Earth model to be adequately represented by one velocity only - $\beta(r)$ - which is not a bad approximation for the inversion of surface waves and their predecessors in the seismogram like S,SS,SSS, etc. Extension of the theory to models of more than one physical parameter is straightforward (Nolet, 1989).

In a laterally homogeneous Earth, we may write the seismic wave as a sum of higher modes of surface waves. Denoting the the mode number by n, the complex excitation coefficient of a mode by $A_{nj}^0(\omega)$, and the wavenumber at circle frequency ω by $k_n^0(\omega)$, we then write for a component of the seismic spectrum $S_j(\omega)$ in a particular station at epicentral distance Δ_j, where j numbers different time series or data:

$$S_j^0(\omega) = \sum_{n=0}^{\infty} A_{nj}^0(\omega)\exp\left[ik_n^0(\omega)\Delta_j\right]$$

The expression for A_{nj}^0 is given in Appendix A. As is well known, deviations from lateral heterogeneity will perturb the phase factor. Neglecting effects of second order, such as backscattering, and variations in the excitation factor $A_{nj}^0(\omega)$, we may then model the seismogram using a phase integral (or WKBJ) approximation as:

$$S_j(\omega) = \sum_{n=0}^{\infty} A_{nj}^0(\omega)\exp\left[i\int_0^{\Delta_j} k_n[\theta(\Delta),\phi(\Delta);\omega]d\Delta\right] \qquad (2)$$

where θ and ϕ are the colatitude and longitude of the surface ray path. The wavenumber $k_n(\theta,\phi;\omega)$ is now the local wavenumber, i.e. it represents the dispersion of this mode in a hypothetical laterally homogeneous Earth with the properties found in the heterogeneous Earth along the radius vector pointing into the direction of (θ,ϕ).

In the following we shall omit the Δ-dependence in the notation of the path given by θ and ϕ, and use the symbol P_j with the integral sign to indicate integration along the surface wave path P_j. We define the average wavenumber perturbation along P_j by:

$$\overline{\delta k_n(\omega)^{(j)}} = \frac{1}{\Delta_j}\int_{P_j}[k_n(\theta,\phi,\omega) - k_n^{(j)}(\omega)]\,d\Delta \qquad (3)$$

where $k_n^{(j)}(\omega)$ is the dispersion for a background model $\beta(r)^{(j)}$ that gives an adequate representation of the average properties along the path P_j. If the true Earth does not deviate too much from this averaged model, a first-order Taylor expansion of k_n as a functional of the background model leads to a very accurate prediction of the average wavenumber perturbation. Defining the average velocity perturbation along P_j as:

$$\overline{\delta\beta(r)^{(j)}} = \frac{1}{\Delta_j}\int_{P_j}[\beta(\theta,\phi,r) - \beta^{(j)}(r)]\,d\Delta \qquad (4)$$

we have the following linearized relationship:

$$\overline{\delta k_n(\omega)^{(j)}} = \int_0^a \left[\frac{\partial k_n^{(j)}(\omega)}{\partial\beta^{(j)}(r)}\right]\overline{\delta\beta^{(j)}(r)}\,dr \qquad (5)$$

where the Frechet derivatives of k_n with respect to β are given in Appendix A. We now have the following relationship between the wavenumber perturbation and the seismic signal that serves as datum j:

$$S_j(\omega) = \sum_{n=0}^{\infty} A_{nj}^{0}(\omega)\exp\left[i\,(k_n^{(j)}(\omega) + \overline{\delta k_n^{(j)}(\omega)})\Delta_j\right] \qquad (6)$$

Nolet et al. (1986) and Nolet (1987) used a discrete parametrization of $\beta(\mathbf{r})$ in (5) to compute the wavenumber perturbation. They used nonlinear optimization to fit the Fourier-transformed predictions (6) to observed time series. This scheme typically leads to $O(10^2)$ parameters in a 2D inversion, which is feasible but tedious, and would lead to $O(10^4)$ parameters, or even more, in a three-dimensional inversion, which is impossible to handle even with the fastest computers.

Expressions (5) and (6) show that the seismic signal along one path is a function of a far more restricted set of parameters. To see this even more clearly, let us develop the averaged velocity perturbation along P_j into a set of M basis functions $h_i(r)$:

$$\overline{\delta\beta^{(j)}(r)} = \sum_{i=1}^{M} \gamma_i^{(j)} h_i^{(j)}(r) \qquad (7)$$

M can be very small, since the basis $h_i^{(j)}$ spans the velocity field only in the vertical coordinate, and is valid only along P_j. With (7) we obtain for (5):

$$\overline{\delta k_n(\omega)^{(j)}} = \sum_{i=1}^{M} \gamma_i^{(j)} \int_0^a \left[\frac{\partial k_n^{(j)}(\omega)}{\partial \beta^{(j)}(r)}\right] h_i^{(j)}(r)\,dr \qquad (8)$$

Note that we allow not only for different background models $\beta^{(j)}(r)$ for different paths, but also for different parameterizations. The first option increases the precision of the linearization involved in (5), since it allows us to choose different models for paths crossing oceans, shields or tectonic regions with often widely different dispersion characteristics. The second option allows us to adapt the depth functions to the expected resolution of the data along various paths in order to increase the efficiency of the inversion.

Partitioned inversion: the nonlinear step

With (8), we have arrived at the starting point of a partitioned inversion scheme. In this section we show how we can estimate the parameters $\gamma_i^{(j)}$ for each path separately, and how we may defer the estimation of the multi-dimensional model $\beta(\mathbf{r})$ to a second, linear inversion step.

Let us first investigate how we can determine the parameters γ_i for one or more seismic time series observed along path P_j. Since we deal with one path only, we shall temporarily omit the superscript j.

$$F(\boldsymbol{\gamma}) = \int [\mathbf{R}s(t,\boldsymbol{\gamma}) - \mathbf{R}d(t)]^2\,dt + \boldsymbol{\gamma}^T \mathbf{C}_\gamma^{-1}\boldsymbol{\gamma} \qquad (9)$$

where the Bayesian term with the (properly weighted) a priori covariance matrix $\mathbf{C}_\gamma$ is optional. As we shall show later, the inclusion of such a term is not needed to stabilize the inversion. The operator $\mathbf{R}$ allows for a time-dependent weighting as well as filtering of the data, which is often needed to prevent that energetic arrivals (such as the fundamental mode) completely dominate the properties of F. In practice we have varied the weighting of different parts of the seismogram until the behaviour of the objective function F agreed with our subjective notion of a 'good fit'.

In order to minimize F we shall use established methods of nonlinear optimization (Nolet et al. 1986, Nolet 1987). An example is given in figure 3. The method differs from other approaches in that no attempt is made to find local gradients of F analytically, but that these are calculated numerically using a realistic step size, and that a fully nonlinear search is conducted in parameter space along optimal directions. The method is very flexible and

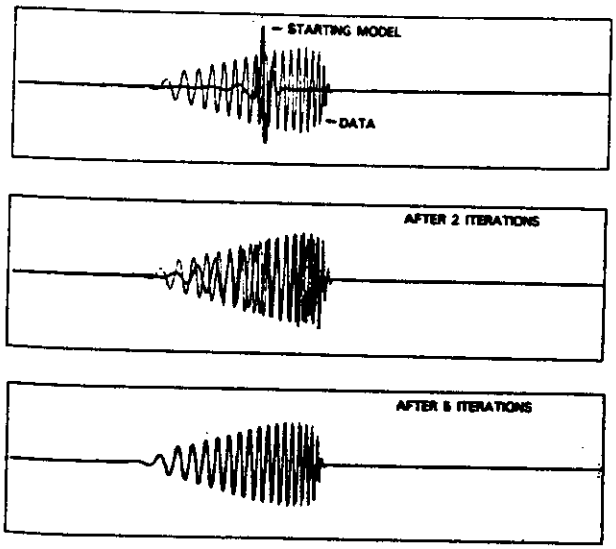

Figure 3: Waveform fitting of a synthetic fundamental Love mode, using nonlinear optimization (Nolet et al., 1986). Each iteration consists of a search in parameter space, after which eigenvectors are re-calculated. R is an operator that computes the signal envelope. The starting model is a simple one-layer-over-halfspace. The final model is a 4-layer model that fits velocities to 0.1% accuracy when compared with the target model.

allows for complicated filtering through R. The search procedure makes it very robust, and the method converges even when phase mismatches are large or when waveforms differ substantially. Finally, we avoid any problems with the fundamental non-differentiability of seismograms that was first recognized by Shaw and Orcutt (1985).

For the derivatives of F with respect to the model parameters, we find:

$$\mathbf{g}(\mathbf{p}) \equiv \nabla F = 2\int_0^T \mathbf{R}\nabla s(t,\boldsymbol{\gamma})[\mathbf{R}s(t,\boldsymbol{\gamma}) - \mathbf{R}d(t)]dt + \mathbf{C}\boldsymbol{\gamma}$$

For the Hessian matrix, the matrix of second derviatives $\partial^2 F/\partial\gamma_i\partial\gamma_j$ we find:

$$\mathbf{H}(\boldsymbol{\gamma}) \equiv \nabla\mathbf{g}(\mathbf{p}) = \int_0^T \left[\mathbf{R}\nabla s(t,\boldsymbol{\gamma})(\mathbf{R}\nabla s(t,\boldsymbol{\gamma}))^T + \mathbf{R}\nabla\nabla s(t,\boldsymbol{\gamma})[\mathbf{R}s(t,\boldsymbol{\gamma}) - \mathbf{R}d(t)]\right]dt + \boldsymbol{\gamma}\mathbf{C}$$

In general, the presence of filters R complicate the computation of H using such analytical expressions, and we shall prefer to use finite difference methods for the computation of g and H. However, this severely restricts the maximum number N of parameters that we can use, since we do need $2N$ evaluations of F to compute g (for sufficient accuracy we must take the difference of two points on either side of the central point p), and even $\frac{1}{2}N(N+1)$ to find H.

We now search for the minimum along one direction (to be defined later), and repeat this with new search directions until the residual has reached an acceptable level, or any other stopping criterion is met. These methods offer the big advantage that memory requirements are minimal, and that explicit calculation of H during the search is not needed. The important point is now to define the search directions in such a way that convergence is fast.

The *method of steepest descent* defines the search direction d as a vector in parameter space anti-parallel to the local gradient of the objective function:

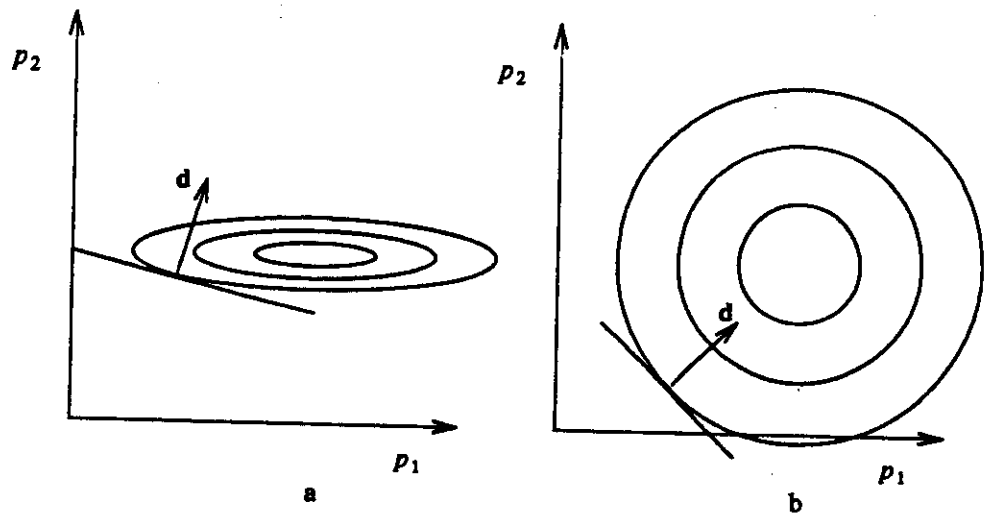

Figure 4: Isolines for two different configurations for the minimum of F in the case of a two-dimensional γ, and the steepest descent direction **d**.

$$\mathbf{d}^i = -\mathbf{g}(\gamma^i)$$

where i is the iteration number. Steepest descent methods have an intuitive appeal (a blind man walking in the mountains would use a steepest descent path to find his way to the beach) but their convergence is slow . The reason is that the vector **d** rarely points into the direction of the minimum. This is shown in figure 4a for a two-parameter model and ellipsoidal isolines for the objective function. The result is that the steepest descent method follows an indirect, zig-zag path towards the minimum. Only if the isolines are circular (figure 4b) does **d** point in the direction of the minimum. In multidimensional optimization, figure 4a corresponds to the case that the eigenvalues of $\mathbf{H}(\gamma)$ are widely different, whereas they ere equal in the 'circular' case of figure 4b. We may therefore expect that descent methods converge faster when $\mathbf{H}(\gamma)$ has one, or only a few, groups of closely spaced eigenvalues (see Gill et al., 1981, page 147). Knowing $\mathbf{H}(\gamma)$ or its estimate, this could be accomplished by preconditioning, i.e. by a scaling of the parameters γ_i. In the more common case that $\mathbf{H}(\gamma)$ is unknown, we have no better option than to scale the parameters intuitively to roughly equal magnitude, e.g. as a percentage of the estimated model value.

Another strategy to speed up the convergence is to use *conjugate directions* for subsequent iterations, i.e.:

$$\mathbf{d}^{iT}\mathbf{H}(\gamma_{min})\mathbf{d}^j = 0 \quad \text{if} \quad i \neq j \tag{10}$$

Eigenvectors of $\mathbf{H}(\gamma_{min})$ are orthogonal, and therefore conjugate, but there are other - nonorthogonal - sets of conjugate vectors that are easier to calculate. Conjugate vectors have the nice property that, for a quadratic objective function, a model change along the direction given by one of these vectors does not impair the function decrease which has been obtained along other directions. This is easily seen by expanding the parameter correction $\Delta\gamma$ and using the Taylor series for bF around the minumum location γ_{min}. Suppose we start from γ_0:

$$\gamma_0 = \gamma_{min} + \Delta\gamma$$

where we may expand:

$$\Delta\gamma = \sum_i \mu_i \mathbf{d}^i$$

Using a second order Taylor expansion around γ_{min}:

$$F(\gamma_0) = F(\gamma_{min}+\Delta\gamma) = F(\gamma_{min}) + \tfrac{1}{2}\sum_{ij} \mu_i \mu_j \, \mathbf{d}^{iT} \mathbf{H}(\gamma_{min}) \mathbf{d}^j$$

Because of (9), the complicated sum with cross terms over different direction vectors reduces to a simple sum:

$$F(\gamma_{min}+\Delta\gamma) = F(\gamma_{min}) + \tfrac{1}{2}\sum_{i} \mu_i^2 \, \mathbf{d}^{iT} \mathbf{H}(\gamma_{min}) \mathbf{d}^i \tag{10}$$

We now minimize in subsequent directions $\mathbf{d}^i$, slowly reducing $\Delta\gamma$, and we see that each new direction subtracts a positive amount from the objective function which is determined by that direction only.

Conjugate directions may be found by a simple recursive scheme (Fletcher and Reeves, 1964; Gill et al., 1981):

$$\mathbf{d}^{i+1} = -\mathbf{g}^{i+1} + \beta_i \mathbf{d}^i \tag{11}$$

where

$$\beta_i = \frac{|\mathbf{g}^{i+1}|^2}{|\mathbf{g}^i|^2}, \quad \beta_0 = 0$$

and $\mathbf{g}$ denotes the gradient of F for the i-th iteration model γ^i.

Convergence is fast and stable if search directions are determined with the method of conjugate gradients. In the following we shall apply this method to converge quickly to a minimum of F. Since the number of parameters is now quite small ($M \approx 10$), it is possible to calculate second derivatives of F using finite differences, and thus the Hessian matrix $\mathbf{H}$ after we have found the optimum γ.

Once we have found γ_{opt}, we have obtained M linear constraints for the Earth model $\delta\beta(r)$, and hence for $\beta(r)$ since we know the background model, through the expansion (7). If every γ_j were uniquely determined, we could assemble many such constraints for different paths, and finally solve the accumulated system of linear equations for various values of r. Unfortunately, there is no guarantee that the equations for γ_i are independent, and in our experience they are often highly dependent. If we were to treat them as independent linear constraints on the Earth model $\beta(r)$, we would not only be overly optimistic, we would also risk ending up with severely conflicting data in the linear inversion. We may however use the knowledge of $\mathbf{H}$ to transform the system (7) to a system of independent data with estimated variances. To this end we diagonalize $\mathbf{H}$:

$$\mathbf{H} = \mathbf{S}\Lambda\mathbf{S}^T \tag{12}$$

and we define a new datum

$$\eta = \mathbf{S}^T\gamma \tag{13}$$

Since $\mathbf{H}$ is symmetric, $\gamma = \mathbf{S}\eta$. We may now express our uncertainty in the solution by allowing the value of F to deviate by an amount ε from its minimum value and define a confidence region $\Delta\gamma = \gamma - \gamma_{opt}$ through:

$$\tfrac{1}{2}\Delta\gamma^T \mathbf{H} \Delta\gamma < \varepsilon \tag{14}$$

This implies for the transformed parameters that $\tfrac{1}{2}\Delta\eta^T \Lambda \Delta\eta < \varepsilon$, or

$$|\Delta\eta_i| \le \sqrt{\frac{2\varepsilon}{\lambda_i}} \tag{15}$$

and we can write the expansion (7) in terms of the new parameters:

$$\overline{\delta\beta(r)} = \sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij}\eta_j h_i(r) = \sum_{j=1}^{M} \eta_j g_j(r) \tag{16}$$

with transformed basis functions

$$g_j(r) = \sum_{i=1}^{M} S_{ij} h_i(r) \tag{17}$$

Equation (16) really represents an infinite system of equations constraining $\beta(r)$, since r can be arbitrarily specified in $0 \le r \le a$. Obviously only a finite subset of these equations is independent. We introduce the *dual* basis $\bar{g}_j(r)$ to reduce (16) to a finite number of independent equations with known variances in the r.h.s. The dual basis satisfies an orthogonality condition with the original set of basis functions $g_i(r)$:

$$\int_0^a \bar{g}_j(r) g_i(r)\, dr = \delta_{ij} \tag{18}$$

so that multiplication of (16) with subsequent members of the dual basis and integration over depth finally yields a usable set of linear constraints for the heterogeneous Earth model $\beta(\theta,\phi,r)$:

$$\frac{1}{\Delta_j} \int_0^a \int_{P_j} [\beta(\theta,\phi,r) - \beta^{(j)}(r)] \bar{g}_i(r)\, dr d\Delta = \eta_i \pm \Delta\eta_i \tag{19}$$

We have restored the path-dependent notation to emphasise that the resulting system of linear equations is in fact independent of the background model since the latter is a known quantity in (19), and to show that different paths may easily use different background models. In general, equation (19) will be employed in a linear inversion for a multi-dimensional model. This is the second stage of the partitioned inversion method. Note that the applicability of the partitioning does not depend on the particular algorithm used to generate synthetic seismograms. For example, instead of (5) we might use the WKBJ method of Chapman (1978) to compute body-wave synthetics, or we might include scattered waves.

The linear inversion step

Repeating the nonlinear inversion as described in the previous section for many different paths, we obtain a large number of constraints (19) for the Earth model $\beta(\mathbf{r})$, which we rewrite as:

$$\int_V G_i'(\mathbf{r}) \beta(\mathbf{r}) d^3\mathbf{r} = \frac{1}{\Delta_i} \int_0^a \int_{P_i} \beta^{(i)}(r) \bar{g}_i(r) dr d\Delta + \eta_i \pm \Delta\eta_i \qquad (i=1,\ldots,N) \tag{20}$$

with

$$G_i'(\mathbf{r}) = \bar{g}_i(r)\, \delta(\theta_i(\Delta) - \theta)\, \delta(\phi_i(\Delta) - \phi)$$

where $\theta_i(\Delta)$ and $\phi_i(\Delta)$ designate the surface ray for path i as a function of raypath length coordinate Δ. In general this system of equations will be underdetermined so that we shall wish to impose a minimum norm condition on the solution. For this it will be useful to define the velocity *deviation* $\delta\beta(\mathbf{r})$ with respect to some standard Earth model, say $\beta^{(0)}(r)$, so that (20) reduces to

$$\int_V G_i'(\mathbf{r}) \delta\beta(\mathbf{r}) d^3\mathbf{r} = \frac{1}{\Delta_i} \int_0^a \int_{P_i} [\beta^{(i)}(r) - \beta^{(0)}(r)] \bar{g}_i(r) dr d\Delta + \eta_i \pm \Delta\eta_i$$

We scale this equation by dividing by $\Delta\eta_i$ and rewrite the scaled and shifted datum as q_i and the scaled kernel as $G_i(r)$ to obtain:

$$\int_V G_i(\mathbf{r}) \delta\beta(\mathbf{r}) d^3\mathbf{r} = q_i \pm 1 \tag{21}$$

The fact that the integral kernels have a δ-function behaviour makes a straightforward application of the inverse method of Backus and Gilbert (1970) impossible. In order to

Table 1a: Epicentral data

Event	Date	T_0	Epicentre	ψ	Region
1	Dec 22, 1983	4:11:29.2	11.866N 13.529W	0°	NW Africa
2	Jan 14, 1987	11:03:48.7	42.565N 142.85E	+3°	Hokkaido

Table 1b: Source parameters

Event	Depth(km)	strike	dip	slip	M_0(Nm)	τ(s)
1	11	305°	24°	-88°	3.4×10^{18}	5.3
2	89	262°	83°	-86°	1.7×10^{19}	9.2

remedy this, we may either assume that the kernels have a finite width conforming to their Fresnel radius, or we may stick to ray theory, but impose an a priori smoothness on the model. Since we have, in this paper, assumed that ray theory is valid (eq. 2), the more consistent approach is the adoption of a priori smoothness and we expand the model as:

$$\delta\beta(\mathbf{r}) = \sum_{i=1}^{N} \alpha_i \tilde{G}_i(\mathbf{r}) \tag{22}$$

with

$$\tilde{G}_i(\mathbf{r}) = B(\theta,\phi) * G_i(\mathbf{r})$$

where * denotes the convolution over the (θ,ϕ) coordinates. A convenient choice for the convolving function is $B(\theta,\phi)=1$ if the arc-length between (0,0) and (θ,ϕ) is less than some specified value, and 0 otherwise. α is now found by solving:

$$\mathbf{A}\alpha = \mathbf{q} \tag{23}$$

where

$$A_{ij} = \int_V G_i(\mathbf{r}) \tilde{G}_j(\mathbf{r}) d^3\mathbf{r}$$

Note that **A** is not symmetric in this case. The system (23) is in general underdetermined, and one must use singular value decomposition, with a cut-off criterion at low values for the singular values of **A** to keep a stable solution.

Resolution analysis

Once we have a model and a satisfactory estimate of ε, we may compute the formal resolution of the model using the Backus-Gilbert method. Following Backus and Gilbert (1970), we estimate a linear local average of the model as

$$\langle\delta\beta(\mathbf{r}_0)\rangle = \sum_{i=0}^{N} a_j(\mathbf{r}_0) q_j = \int_V A(\mathbf{r}_0,\mathbf{r}) \delta\beta(\mathbf{r}) d^3\mathbf{r} \tag{23}$$

where it is understood that the integration should be over a vertical surface in the 2-D case. It follows from (23) that the averaging kernel is a linear combination of the data kernels:

$$A(\mathbf{r}_0,\mathbf{r}) = \sum_{i=1}^{N} a_i(\mathbf{r}_0) G_i(\mathbf{r}) \tag{24}$$

and the variance of the average is given by

$$\sigma(\mathbf{r})^2 = \sum_{i=1}^{N} a_i(\mathbf{r}_0)^2 \tag{25}$$

The factors $a_i(\mathbf{r}_0)$ must be found by minimizing the spatial region where $A(\mathbf{r}_0,\mathbf{r})$ is

significantly different from 0, while at the same time keeping the center of gravity of this kernel close to r_0 and the volume integral close to its zero-bias value of 1. Although Backus and Gilbert propose adding the condition of zero bias in the 1-D case, it is our experience that this may lead to strong numerical instabilities for large size problems in 2-D or 3-D. But a second type of criterion proposed by Backus and Gilbert works very well. We find the factors $a_i(r_0)$ by minimizing

$$J(r_0) = \int_V [A(r_0,r) - \delta(r_0-r)]^2 d^3r + w\sigma(r)^2 \qquad (26)$$

for various values of the trade-off parameter w. Such computations can be done very efficiently if singular value decomposition is applied to the resulting system of linear equations (Gilbert, 1972).

We note that this resolution analysis depends on the validity of the second-order expansion of $F(\gamma)$. It does not work correctly in the case of secondary minima with acceptable data misfit. Backus (1989) developed a comprehensive theory to estimate resolution, even in the case of nonlinear problems. The present analysis is not unlike that of Backus in that we avoid Bayesian assumptions and estimate the resolution from a confidence region in data space. However, since the emphasis in this paper is on computational efficiency, we have refrained from introducing hard constraints on $|\gamma|$ as well as on modelling errors.

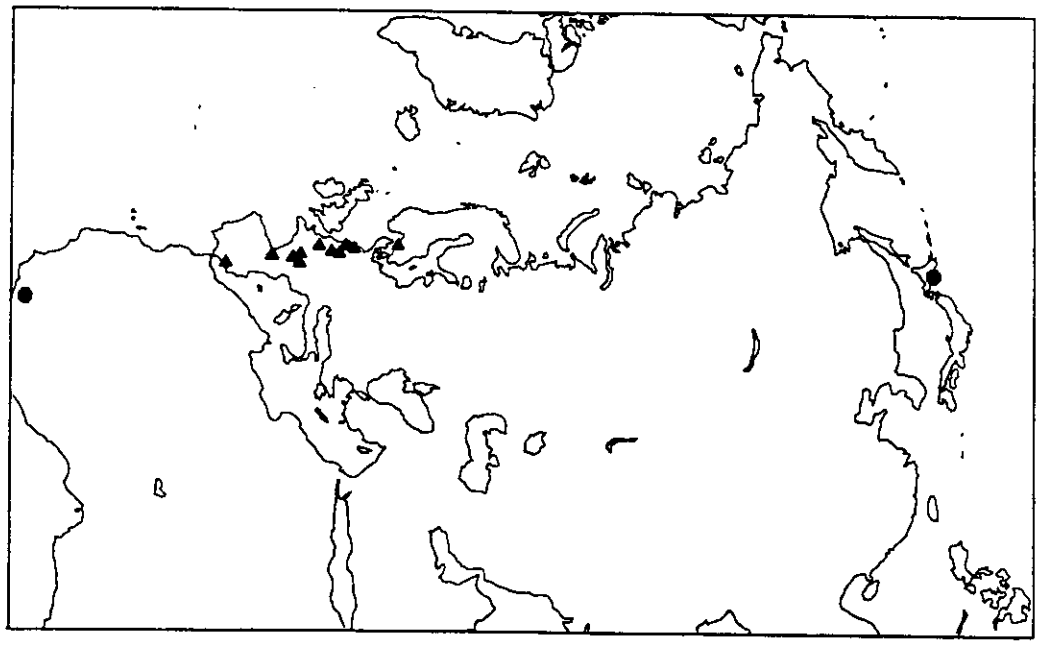

Figure 5: Locations of events (circles) and NARS stations (triangles)

Illustration

To illustrate the theory and test its assumptions, we have applied it to data from the NARS array (Dost et al., 1984; Nolet et al., 1985) in western Europe. With a station spacing of roughly 200 km, and digital, broad-band recording, the seismic data from this array form an ideal source of information on lateral heterogeneity in the upper mantle. Since the configuration of the array is approximately linear, we have searched the data set for events located on the array great-circle, and we aim for a 2D model of the upper mantle S-velocity directly under the array. At this moment, this is about as far as we can go with available data. We expect, however, that the density of similar seismic stations in and around

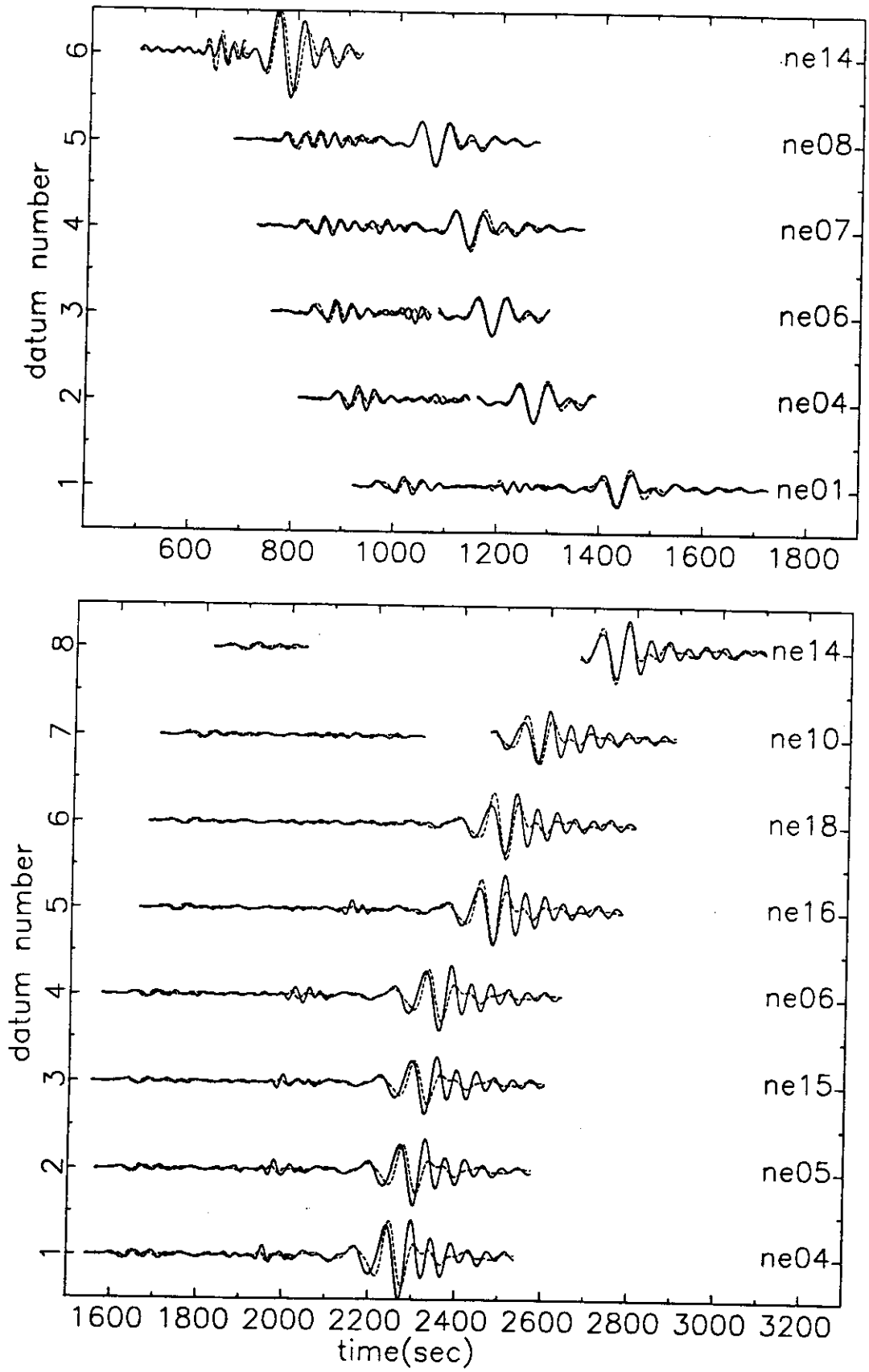

Figure 5: PREMC Synthetics (broken) and observed vertical displaments (solid) for the events in Africa (top) and Hokkaido (bottom).

western Europe will soon be large enough to attempt a full, 3D inversion.

The two events used are listed in Table 1, and plotted in figure 5. Both events are located within 3° of the NARS great circle. We used the HVRD source mechanisms as listed in the PDE monthly bulletins.

As a background model for the inversion, we chose a modified PREM model (Dziewonski and Anderson, 1981). We increased the thickness of the crust in this model to 30 km, lowered Q in the crust and imposed a uniform S-velocity of 4.47 km/s and Q of 143 between the Moho and 220 km. We call this model PREMC. The modifications are such that PREMC gives a good fit to the fundamental mode signal for event 1, as can be seen in figure 2. The synthetics were calculated by summing up to 20 modes at 50 frequency points below 60mHz.

We used vertical component recordings. The influence of the density ρ and the compressional velocity α on the seismogram is very much less than that of β in the time window of interest and these parameters cannot be determined independently. We therefore seek to resolve a 2D-model $\beta(r)$ that represents a cut through the Earth along the NARS great circle. We ignore the possible effects of anisotropy and multipathing because we do not expect to be able to really resolve these with the data now available.

The two selected events form a very severe test of the method and its assumptions. There is a strong overlap of raypaths for each event separately, and for both events under the array itself, and we expect that effects of multipathing will show up as inconsistencies between the linear constraints (19). Rays from both events, but especially from event 2, have travelled a considerable distance outside the array whereas we are mostly interested in the detailed structure under the array itself. The sub-array structure has therefore a smaller influence on the waveshape then the average structures under northern Eurasia and western Africa, respectively, and we definitely need the resolving kernels (24) to decide how reliable the outcome of the inversion is.

The individual fits that were obtained using the nonlinear inversion are shown in figure 4a and 4b. At this stage, we also modified the Q model once more to make it optimal for the new weighting. The fits to event 1 (figure 4a) are very satisfactory. The synthetic LR, S and SS are in phase with the observations at all stations, with the exception of the SS-S interference in ne07. Despite considerable effort we have not been able to fit this wave completely. For event 2, a similar good fit is obtained for the S wavegroups and the LR wave (figure 4b). To obtain the fits in figures 4a and 4b, perturbations of up to -6% were needed in the average sub-Moho velocity for event 2, a probable influence of the larger crustal thickness along the path over Eurasia. Also, the jump at 220 km was decreased by a velocity change below 220 of about -3% along this path. All other perturbations in path averages for event 2 were below 1.5%. With a few exceptions, parameter changes for event 1 were smaller than 2%. These values by themselves say very little, since we did not use any damping in the inversion but relied instead on the linear inversion to find us a compatible model with acceptably small velocity variations.

The confidence level ε in (14) was essentially determined for each event by generating synthetics for slight deviations $\Delta\gamma$ away from γ_{opt}, and deciding subjectively which ε draws the line between 'acceptable' and 'unacceptable' misfits.

Ideally, these 14 seismograms, modelled with 9 parameters each, could provide 126 constraints on the 2D S-velocity structure beneath the NARS array. In reality, many $\Delta\eta_i$ were extremely large. Some ill-constrained parameters even yielded (small) negative eigenvalues of the Hessian, a consequence of the omission of a Bayesian term in the objective function. We excluded ill-constrained η_i from the outset from (19). This gave us a total of 56 constraints. The singular values λ_i of A decrease rapidly, such that λ_9 is less then 1% of λ_1, and there were only 48 eigenvalues larger than $10^{-6}\lambda_1$. This is the numerical confirmation of our earlier statement that the application constitutes a severe test

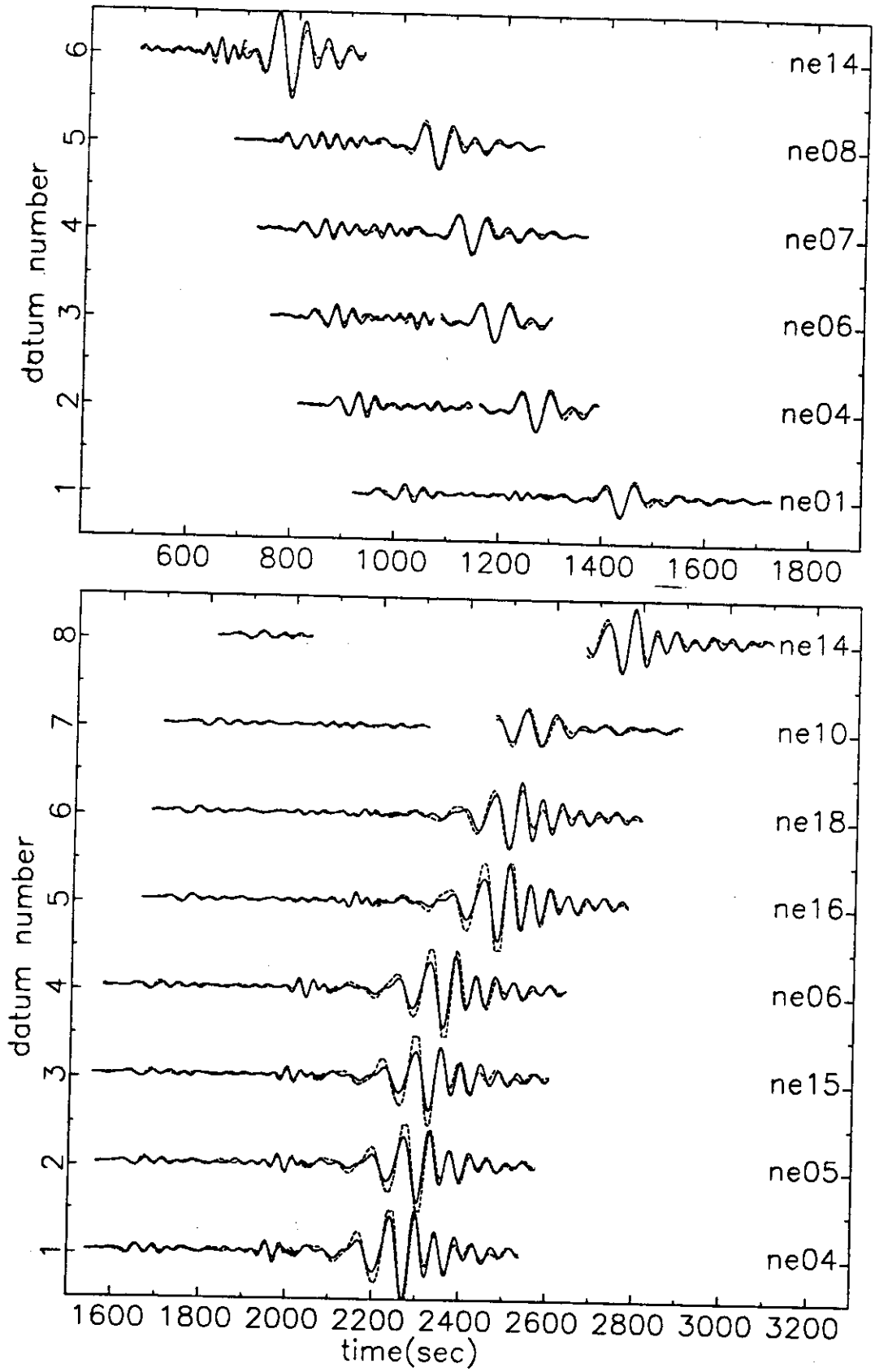

Figure 6: As figure 5, but individual seismogram fits after application of the nonlinear inversion step.

of the method, since the very low effective rank of **A** makes it impossible to fit inconsistencies in the data with plausible 2D models of the S-velocity. Such inconsistencies may arise because not all paths follow exactly the great circle and may actually sample different parts of the Earth outside the array ('out-of-plane propagation'). Horizontal refraction of waves may lead to similar effects.

We adopted a model constructed from the first 15 eigenvectors as our 'preferred' solution. This model is named WEPL3 and shown in figures 6a and 6b. When recalculating the fit for the WEPL3 (figure 7a and 7b), it is obvious that the fit has not suffered very much from damping. The misfit of the SS wave in station ne07 has remained, which indicates that this misfit cannot be ascribed to the nonlinear algorithm getting stuck in a local minimum. Rather, we suspect that multipathing and/or scattering from the nearby edge of the LVL (figure 6) is to blame. For event 2, only the fit to the S_a wave in ne16 has suffered. Note that stations ne07 and ne16 are the two stations that deviate most from the array great circle path. We suspect that our 2D approximation is to blame for the misfit, rather than a breakdown of the great-circle approximation itself - although this cannot be stated with full certainty without additional observational evidence. We note, however, that efforts to increase the maximum frequency beyond 50 mHz were generally unsuccesful; occasionally, 50 mHz proved even a bit too much for the shallow S energy, and it is likely that scattered energy is to blame for the failure to find adequate snthetics at high frequencies.

From the resolution calculations we infer that the horizontal resolution is of the order of 1000 km in the south, 400-700 km in the north where the station are more densely spaced. The vertical resolution is about 100 km at 80 km depth and 100-150 km at deeper levels. Below 400 km we cannot obtain much better standard deviations in the velocities than about 0.1 km/s, with 700 km resolution horizontally.

Discussion

Through the partitioning of the nonlinear inverse problem, we have brought the inversion of large datasets with waveforms within reach of the power of present computing technology. Essentially this is caused by a considerable reduction in the number of evaluations of $F(\gamma)$ that is needed: for every seismogram we have only $O(10)$ parameters, for the whole Earth we would have $O(10^5)$ or more parameters, and even for regional problems this number easily exceeds 10^3.

An additional adantage of the partitioning is that we are able to treat one event or region at a time. We simply save up our constraints. If, at a later stage, more data become available, we simply add these to the linear inversion step. Also, we may combine constraints from waveform inversion with other linear constraints. Ultimately, seismology n the 1990's will progress by the accumulation of large data banks with linear constraints on Earth structure!

The use of waveforms in imaging allows us to study regions with no or weak seismicity that are crossed by surface wave paths. If we have many crossing paths, we may even obtain a reasonable resolution in regions of poor station coverage such as the oceans. This is a consequence of a theorem due to Radon (see, Chapman, 1987). However, the horizontal resolution is ultimately determined by the number of seismic stations. These stations must be broad-band to allow for a reliable recording of low frequencies, and digital to allow for processing of large data sets. The progress of seismology in the 1990's for unraveling the deep Earth structure is therefore very much dependent on cooperative observational projects such as POSEIDON in Japan, ORFEUS is Europe, and the IRIS and GSN initiatives in the US.

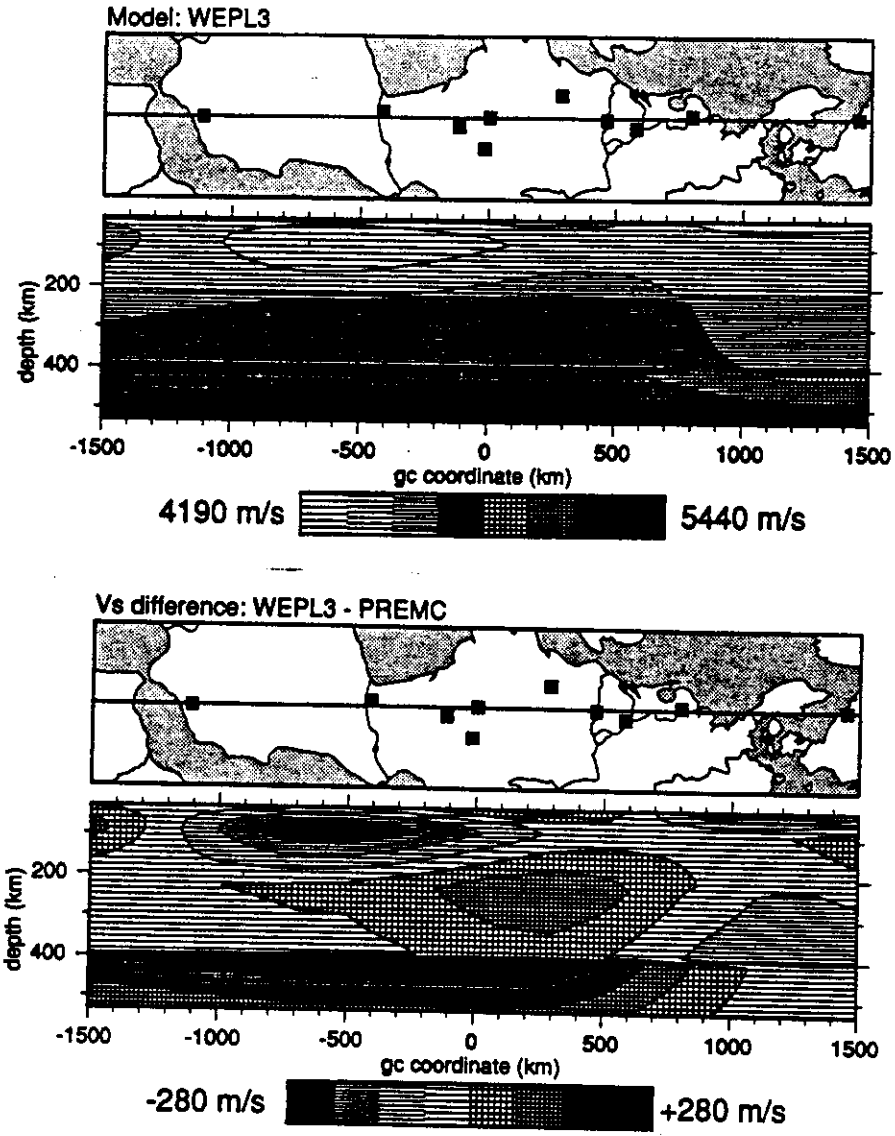

Figure 7: Model WEPL3 *S* velocity, absolute (top) and difference with the starting model (bottom)

Acknowledgements

The theory of partitioned nonlinear inversion was developed while the author was a Cecil H. and Ida Green Scholar at the Institute of Geophysics and Planetary Physics, Scripps Institute of Oceanography. He gratefully acknowledges this support, as well as the many interesting discussions with colleagues at Scripps and in Utrecht.

References

Backus, G., Confidence set inference with a prior quadratic bound, preprint, 1989.

Backus, G. and G. F. Gilbert, Uniqueness on the inversion of inaccurate gross Earth data, Phil. Trans. Roy. Soc. Lond., A266, 123, 1970.

Chapman, C. H., A new method for computing synthetic seismograms, Geophys. J. Roy. astr. Soc., 54, 481-518, 1978.

Chapman, C.H., The Radon transform and seismic tomography, in: Nolet, G. (ed), *Seismic Tomography*, Reidel, 25-48, 1987.

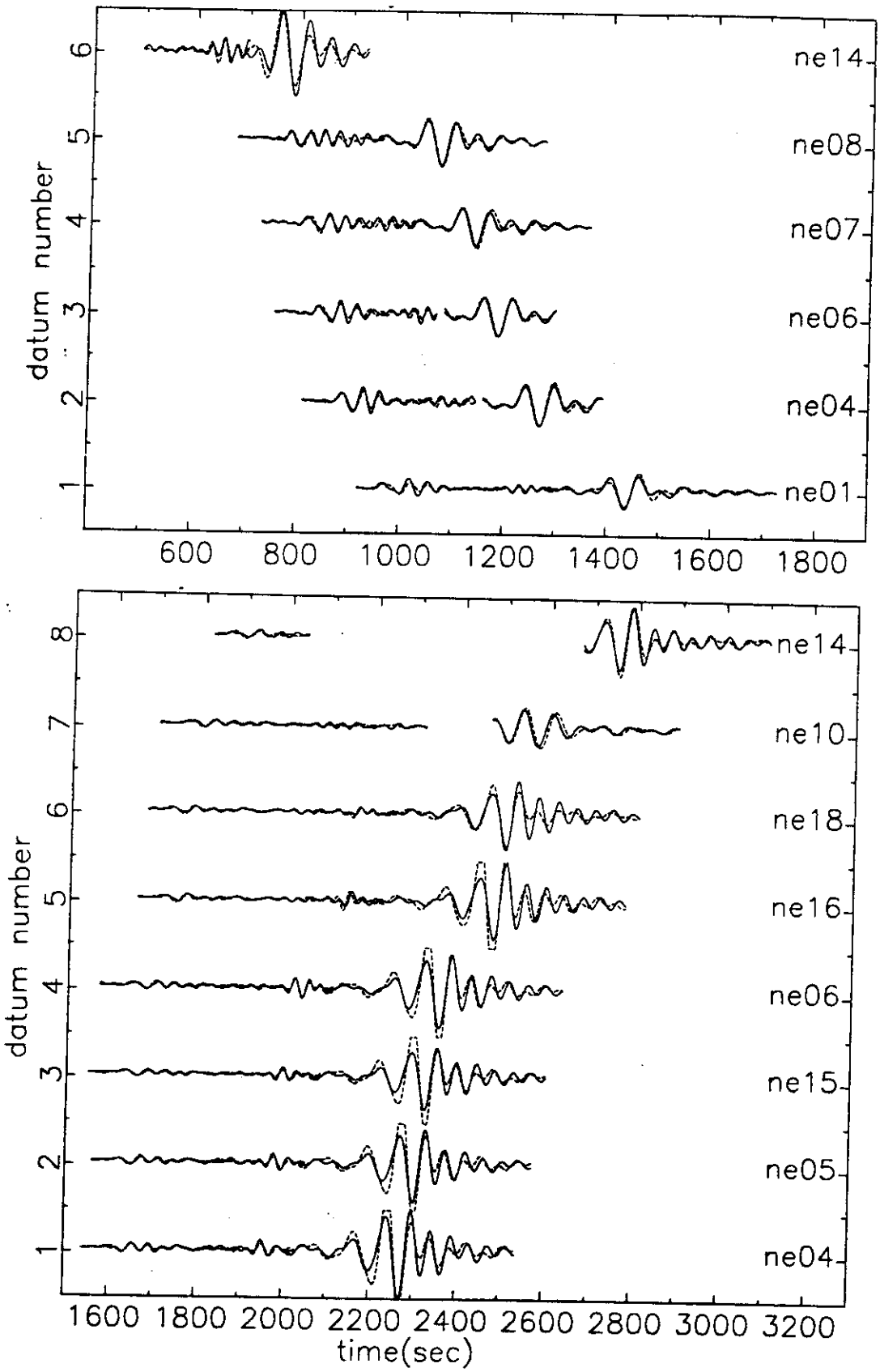

Figure 8: As figure 5, but seismogram fits for WEPL3 after application of the final linear inversion step.

Dost, B., The NARS array, PhD thesis, Utrecht University, 1987.
Dost, B., A. van Wettum and G. Nolet, The NARS array, Geol. Mijnb., 63, 318-386, 1984.
Dziewonski, A. M. and D. L. Anderson, Preliminary reference Earth model, Phys. Earth Plan. Int., 25, 297-356, 1981.
Fletcher, R. and C.M. Reeves, Function minimization by conjugate gradients, Computer J., 7, 149-154, 1964.
Gilbert, J. F., Ranking and winnowing gross Earth data for inversion and resolution, Geophys. J. Roy. astr. Soc., 43, 125, 1971.
Gill, P.E., W. Muray and M.H. Wright, *Practical Optimization*, Acad. Press, London, 401pp., 1981.
Jordan, T. H., A procedure for estimating lateral variations from low frequency eigenspectra data, Geophys. J. Roy. astr. Soc., 54, 571-610, 1978.
Nolet, G., Waveform tomography, in: Nolet, G., ed., *Seismic Tomography*, Reidel, 301-322, 1987.
Nolet G., Can teleseismic arrivals be recognized with sonobuoys?, Internal rep., Scripps Inst. Ocean., 1988.
Nolet, G., Partitioned wave-form inversion and 2D structure under the NARS array, J. Geophys. Res., 95, 8499-8512, 1990.
Nolet, G., B. Dost and H. Paulssen, Intermediate wavelength seismology and the NARS experiment, Ann. Geophys. 4, 305-314, 1985 (with errata in Ann. Geophysicae 4, 593, 1986).
Nolet, G., J. van Trier and R. Huisman, A formalism for nonlinear inversion of seismic surface waves, Geophys. Res. Lett., 13, 26-29, 1986.
Poupinet, G., Seismic data collection platforms for satellite transmission, in: Nolet, G. (ed), *Seismic Tomgraphy*, Reidel, 239-250, 1987.
Shaw, P. R. and J. A. Orcut, Waveform inversion of seismic refraction data and applications to young Pacific crust, Geophys. J. Roy. astr. Soc., 82, 375-414, 1985.
Snieder, R. K., On the connection between ray theory and scattering theory for surface waves, in: Vlaar, N. J. et al. (eds.), *Mathematical geophysics*, Reidel, 77-83, 1988.
Snieder, R. and G. Nolet., Linearized scattering of surface waves on a spherical Earth, J. Geophys., 61, 55-63, 1987.
Spakman, W., M. J. R. Wortel and N. J. Vlaar, The Hellenic subduction zone: a tomographic image and its geodynamic implications, Geophys. Res. Lett., 15, 60-63, 1988.
Woodhouse, J. H., Surface waves in a laterally varying layered structure, Geophys. J. Roy. astr. Soc., 37, 461-490, 1974.
Woodhouse, J. H. and Y. K. Wong, Amplitude, phase and path anomalies of mantle waves, Geophys. J. Roy. astr. Soc., 87, 753-774, 1986.

Appendix A

A seismogram can be approximated by summing a finite number of modes at each frequency component. Snieder and Nolet (1987) model the spectrum of a far field seismic wave as:

$$S_R(\gamma|\mathbf{r},\omega) = -\frac{1}{2}\sum_n \frac{[\hat{\mathbf{r}}U_n(r) + i\hat{\boldsymbol{\theta}}\lambda_n(\omega)V_n(r)]}{u_n(\omega)\sqrt{\sin\theta}} A_{Rn}(\phi) e^{i\lambda_n\theta - \omega\theta/2Q_n u_n} \qquad \text{(A1)}$$

for Rayleigh waves, and for Love waves:

$$S_L(\gamma|\mathbf{r},\omega) = \frac{1}{2}\sum_n \frac{i\hat{\boldsymbol{\phi}}\lambda_n(\omega)W_n(r)}{u_n(\omega)\sqrt{\sin\theta}} A_{Ln}(\phi) e^{i\lambda_n\theta - \omega\theta/2Q_n u_n} \qquad \text{(A2)}$$

where $\lambda_n(\omega)$ is the wavenumber and $u_n(\omega)$ is the group velocity, both scaled with the Earth's radius a: $\lambda_n(\omega)=ak_n(\omega)$ and $u_n(\omega)=(\partial\lambda_n/\partial\omega)^{-1}$. We note that $\lambda_n(\omega)$ is related to the angular order l of the wave by $\lambda_n(\omega)=\sqrt{l(l+1)}$. The amplitude of the surface wave is linearly related to the components M_i of the moment tensor:

$$A_{RLn}(\phi) = (\lambda_n/2\pi)^{1/2}\sum_{k=1}^{6} B_{kn}^{RL} M_k \qquad \text{(A3)}$$

with the convention: $M_1=M_{rr}, M_2=M_{r\theta}, M_3=M_{r\phi}, M_4=M_{\theta\theta}, M_5=M_{\phi\phi}, M_6=M_{\theta\phi}$, and:

$$B_1^R = \partial_r U_z\, e^{-i\pi/4}$$

$$
\begin{aligned}
B_2^R &= -\lambda_n[\partial_r V_s + r_s^{-1}(U_s - V_s)]\cos\zeta_w\, e^{i\pi/4} \\
B_3^R &= -\lambda_n[\partial_r V_s + r_s^{-1}(U_s - V_s)]\sin\zeta_w\, e^{i\pi/4} \\
B_4^R &= \tfrac{1}{2}r_s^{-1}(2U_s - \lambda_n^2 V_s)\, e^{-i\pi/4} - \tfrac{1}{2}\lambda_n^2 r_s^{-1} V_s \cos 2\zeta_w\, e^{-i\pi/4} \\
B_5^R &= \tfrac{1}{2}r_s^{-1}(2U_s - \lambda_n^2 V_s)\, e^{-i\pi/4} + \tfrac{1}{2}\lambda_n^2 r_s^{-1} V_s \cos 2\zeta_w\, e^{-i\pi/4} \\
B_6^R &= -\lambda_n^2 r_s^{-1} V_s \sin 2\zeta_w\, e^{-i\pi/4}
\end{aligned}
\tag{A4}
$$

$$
\begin{aligned}
B_1^L &= 0 \\
B_2^L &= -\lambda_n[\partial_r W_s - r_s^{-1} W_s]\sin\zeta_w\, e^{i\pi/4} \\
B_3^L &= \lambda_n[\partial_r W_s - r_s^{-1} W_s]\cos\zeta_w\, e^{i\pi/4} \\
B_4^L &= -\tfrac{1}{2}\lambda_n^2 r_s^{-1} W_s \sin 2\zeta_w\, e^{-i\pi/4} \\
B_5^L &= -B_4^L \\
B_6^L &= \lambda_n^2 r_s^{-1} W_s \cos 2\zeta_w\, e^{-i\pi/4}
\end{aligned}
\tag{A5}
$$

Where ζ_w denotes the source-to-station azimuth, measured counterclockwise from South. In (A1)-(A4) we assumed a Fourier convention $s(\omega) = \int s(t) e^{+i\omega t} dt$. r_s denotes the source location, and the normalization used is

$$
\omega^2 \int_0^a \rho[U_n^2 + \lambda_n^2 V_n^2] r^2 dr = 1, \qquad \omega^2 \int_0^a \rho \lambda_n^2 W_n^2 r^2 dr = 1 \tag{A6}
$$

In the following we shall use a shorthand notation, and use for one of the components of the displacement spectrum (the j-th datum):

$$
S_j(\omega) = \sum_n S_{nj}(\omega) = \sum_n A_{nj}(\omega) e^{i\chi_{nj}(\omega) - \alpha_{nj}(\omega)} \tag{A7}
$$

To accomodate slight variations in phase velocity along the wave path, we replace the simple expressions for χ_{nj} and α_{nj} by integrals along the surface ray path. This approximation, which is widely applied but may not be valid for strong gradients in velocity or for certain wavelengths (see section 1) is essentially a phase integral, or WKBJ approximation:

$$
\chi_{nj}(\omega) = \int_{S_j} [k_n(\omega) + \delta k_n(\omega,\theta,\phi)] ds \tag{A8}
$$

and, changing the temporal Qaulity factor Q^T in (A1,A2) to the spatial factor $Q^X = Q^T k_n u_n / \omega$ we find:

$$
\alpha_{nj}(\omega) = \frac{1}{2} \int_{S_j} [k_n(\omega) + \delta k_n(\omega,\theta,\phi)][Q_n^{-1}(\omega) + \delta Q_n^{-1}(\omega,\theta,\phi)] ds \tag{A9}
$$

where S_j denotes the surface ray path, $\delta k_n(\omega,\theta,\phi)$ and $\delta Q_n^{-1}(\omega,\theta,\phi)$ are the *local* perturbations in wavenumber and quality factor of the mode, respectively, as a function of colatitude θ and longitude ϕ.

To a very good approximation, $\delta k_n(\omega,\theta,\phi)$ and $\delta Q_n^{-1}(\omega,\theta,\phi)$ are linearly related to the local perturbations in the starting model:

$$
\begin{aligned}
\delta k_n(\omega,\theta,\phi) &= \int_0^a \left\{ \left[\frac{\partial k_n}{\partial \alpha}\right]_{\beta,\rho} \delta\alpha + \left[\frac{\partial k_n}{\partial \beta}\right]_{\alpha,\rho} \delta\beta + \left[\frac{\partial k_n}{\partial \rho}\right]_{\alpha,\beta} \delta\rho \right\} dr \\
\left[\frac{\partial k_n}{\partial \alpha}\right]_{\beta,\rho} &= D_1^{L,R}\, 2\alpha\rho \\
\left[\frac{\partial k_n}{\partial \beta}\right]_{\alpha,\rho} &= (D_2^{L,R} - 2D_1^{L,R})\, 2\beta\rho
\end{aligned}
\tag{A10}
$$

$$\left[\frac{\partial k_n}{\partial \rho}\right]_{\alpha,\beta} = (D_1^{L,R} - 2D_2^{L,R})\beta^2 + D_2^{L,R}\alpha^2 + D_3^{L,R}$$

where, for Love and Rayleigh waves, respectively:

$$D_1^L = 0$$

$$D_2^L = -\frac{\omega}{2V_g}[r^2(\partial_r W_n - \frac{1}{r}W_n) + (\lambda_n^2 - 2)W_n^2 \qquad \text{(A11)}$$

$$D_3^L = \frac{\omega^3}{2V_g} r^2 W_n^2$$

$$D_1^R = -\frac{\omega}{2V_g}[r^2(\partial_r U_n)^2 + 2r\,\partial_r U_n(2U_n - \lambda_n^2 V_n) + (2U_n - \lambda_n^2 V_n)^2]$$

$$D_2^R = -\frac{\omega}{2V_g}[2r^2(\partial_r U_n)^2 + (2U_n - \lambda_n^2 V_n)^2 + \lambda_n^2(r\,\partial_r V_n + U_n - V_n)^2 + \lambda_n^2(\lambda_n^2 - 2)V_n^2] \qquad \text{(A12)}$$

$$D_3^R = \frac{\omega^3}{2V_g}[U_n^2 + \lambda_n^2 V_n^2]$$

and where V_g denotes group velocity $(\partial k_n(\omega)/\partial\omega)^{-1}$

Assuming zero bulk loss, the spatial quality factor $Q_n^{-1}(\omega)$ of the mode is linearly related to the intrinsic $Q_\mu^{-1}(r)$ of the Earth:

$$Q_n^{-1}(\omega) = -\frac{1}{k_n(\omega)}\int_0^a \beta(r)\left[\frac{\partial k_n}{\partial \beta}\right]_{\alpha,\rho} Q_\mu^{-1}(r) r^2 dr \qquad \text{(A13)}$$

so that

$$\delta Q_n^{-1}(\omega,\theta,\phi) = \frac{1}{k_n(\omega)}\int_0^a \left[\frac{\partial k_n}{\partial \beta}\right]_{\alpha,\rho} [\beta\delta Q_\mu^{-1}(r,\theta,\phi) + \delta\beta(r,\theta,\phi) Q_\mu^{-1}(r)] r^2 dr \qquad \text{(A14)}$$

The term with $\delta\beta$ dissappears when we have a perfectly elastic starting model. If we define δQ_μ^{-1} rather than δQ_μ as our model, inversion for $Q_n^{-1}(\omega)$ is linear in $Q_\mu^{-1}(r)$.

The linear expressions (A8) and (A9) enable us to calculate $\chi_{nj}(\omega)$ and $\alpha_{nj}(\omega)$ in a very efficient manner. We only need to solve the eigenvalue problem for the $k_n(\omega)$ once, and store the partial derivatives with respect to model perturbations. Thus, if we are willing to ignore perturbations in the amplitudes A_n due to perturbations in the elastic structure, it is a very simple matter to calculate the spectrum of the synthetic signal.