

Towards global earth tomography using the spectral element method: a technique based on source stacking

Y. Capdeville,* Y. Gung and B. Romanowicz

Department of Earth and Planetary Science and Seismological Laboratory, University of California at Berkeley, CA, USA

Accepted 2005 May 16. Received 2005 March 2; in original form 2004 July 16

SUMMARY

We present a new tomographic method based on the non-linear least-squares inversion of seismograms using the spectral elements method (SEM). The SEM is used for the forward modelling and to compute partial derivatives of seismograms with respect to the model parameters. The main idea of the method is to use a special data reduction scheme to overcome the prohibitive numerical cost of such an inversion. The SEM allows us to trigger several sources at the same time within one simulation with no incremental numerical cost. Doing so, the resulting synthetic seismograms are the sum of seismograms due to each individual source for a common receiver and a common origin time, with no possibility to separate them afterward. These summed synthetics are not directly comparable to data, but using the linearity of the problem with respect to the seismic sources, we can sum all data for a common station and a common zero time, and we perform the same operation on synthetics. Using this data reduction scheme, we can then model the whole data set using a single SEM run, rather than a number of runs equal to the number of events considered, allowing this type of inversion to be feasible on a reasonable size computer.

In this paper we present tests that show the feasibility of the method. It appears that this approach can work owing to the combination of two factors: the off-path sensitivity of the long-period waveforms and the presence of multiple-scattering, which compensate for the loss of information in the summation process. We discuss the advantages and drawbacks of such a scheme.

Key words: global seismology, inverse problem, spectral element method, tomography.

1 INTRODUCTION

Global seismic tomography is one of the most powerful tools to study the earth's interior structure (for a recent review, see Romanowicz 2003). Its basic principle involves two steps. The first step is a forward modelling step, in which a starting 1-D or 3-D model is chosen, and a seismic wave propagation theory is used to compute a synthetic data set to be compared with real observations. Depending on the approach, the observed data set can consist of time domain waveforms, that is, entire seismograms or portions thereof, or extracted 'observables' such as body wave traveltimes or surface wave phase velocities, collected from earthquake records at seismic stations distributed around the globe. In the second step, an inverse problem is solved, in which perturbations to the starting model are sought in order to explain differences between the synthetic and observed data sets. The procedure can be iterated until convergence.

Practically all forward modelling approaches used so far to account for 3-D structure at the global scale are based on first order perturbation theory (Born approximation), which is limited in its domain of validity to weak lateral seismic velocity contrasts. Even within this theory, further approximations are considered, such as, in the context of normal mode perturbation theory, the Path Average approximation (PAVA, Woodhouse & Dziewonski 1984), which limits the sensitivity to the average 1-D structure beneath the great circle path containing the source and the receiver, and is therefore strictly only a good approximation for the analysis of phase velocities of fundamental mode surface waves. In the context of the analysis of body wave traveltimes, the approximation often used is ray theory, which is an infinite frequency approximation and distributes the sensitivity to structure uniformly along the infinitesimal ray path. These standard approximations carry with them another drawback, which is that only a fraction of the information contained in whole seismograms can be utilized, well-separated body wave phases in the case of ray theory (except diffracted waves), fundamental mode surface wave phase velocities and mode frequency shifts for modes that are well separated in the frequency domain, in the case of normal mode perturbation theory. The advantage of

*Now at: Département de sismologie de l'Institut de Physique du Globe de Paris, Paris, France. E-mail: capdevil@ipgp.jussieu.fr

these approaches is that they are very fast computationally. Recently, the introduction of more sophisticated, higher order approximations to Born seismograms has allowed to perform tomographic inversions of complete long-period seismograms, containing waveforms of mixed body wave phases, diffracted waves, fundamental and higher mode surface waves, with more accurate forward modelling step and structure sensitivity kernels. For example, the Non-linear Asymptotic Coupling Theory (NACT, Li & Tanimoto 1993; Li & Romanowicz 1995) considers across-branch mode coupling to zeroth order asymptotically, resulting in 2-D broadband sensitivity kernels appropriate for body waves and diffracted waves, leading to the first global 3-D models based entirely on whole seismogram, long-period waveforms (Li & Romanowicz 1996; Mégnin & Romanowicz 2000; Gung & Romanowicz 2004). Still, sensitivity is limited to the 2-D vertical plane containing the source and the receiver. An extension to a higher order asymptotic approximation (Romanowicz 1987), which allows the introduction of sensitivity to the third dimension (focusing/defocusing), is still computationally effective, but restricted in validity to relatively smooth heterogeneity and observations away from the source and its antipode (Capdeville *et al.* 2002) or fundamental mode surface waves (Zhou *et al.* 2004). On the other hand, broadband kernels have recently been introduced for the computation of long-period body wave traveltimes. Most recently, with the advent of more powerful computers, full Born computations have started to be put in practice (e.g. Zhao *et al.* 2000), however, at the global scale, the corresponding broadband kernels have only been used for tomography based on traveltimes (Montelli *et al.* 2004), again, limiting the type of information utilized in seismic records, and therefore the sampling of the earth's interior that can be obtained.

The main advantage of the approximations currently used in tomography is that their computational speed is fast enough to allow inversion within a reasonable time frame (several days to several weeks). There are also serious drawbacks, especially as we seek to constrain increasingly finer details in the models. Because source station distribution is limited around the globe, restricting the data set to a few well-isolated body wave phases in the seismogram limits the sampling within the earth, leaving large gaps in areas inaccessible in other ways than through the illumination by multiply reflected/converted phases (in the vertical direction), or scattered waves interacting with lateral heterogeneity (in the horizontal direction). Also, the large data processing effort involved in measuring traveltimes or phase/group velocities is somewhat disproportionate with the limited sampling obtained—which is not likely to improve significantly unless major data collection efforts such as USArray of Earthscope are systematically extended to the whole globe, including the ocean floor.

Therefore, we look to future progress in mantle tomography through the combined use of full time domain seismograms and an accurate wave propagation theory in a 3-D earth, with fewer limits of validity. Until recently, the latter has not been available. In recent years, progress has been made in two directions: the development of higher order perturbation theory in the context of normal mode theory (e.g. Lognonné 1989; Lognonné & Romanowicz 1990), as well as numerical approaches (e.g. Cummins *et al.* 1997). While higher order perturbation theory is currently being explored as a possible tool for mantle tomography (e.g. Millot-Langet *et al.* 2003), we will here consider a numerical approach, the Spectral Element Method (SEM), recently introduced in seismology in Cartesian Geometry (Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999), and in spherical geometry for global earth scale applications (Chaljub 2000; Capdeville 2000; Komatitsch & Tromp 2002;

Capdeville *et al.* 2003a; Chaljub *et al.* 2003; Capdeville *et al.* 2003b). This method has the advantage of being able to model with accuracy the entire seismogram at any location with respect to the source, without any *a priori* assumptions on the velocity contrasts within the Earth. The availability of this new tool allows us to address the issue of full waveform tomography. Obviously, the main difficulty is the computing power required, which may be so large that the inverse problem would not be solved in practice with this tool for many years to come. We will investigate the inverse problem in the framework of classical non-linear least squares formalism (Tarantola & Valette 1982). This choice, in contrast to full space search approaches, already limits the type of models that can be obtained, but is arguably a reasonable approach at the global scale. We first show that, even with a least squares inversion technique, the complete inverse problem is numerically too expensive to be solved with presently accessible computers. Then, we present an approach, based on a specific data reduction scheme, which makes this problem more tractable. We illustrate the feasibility and potential of this approach through several synthetic tests.

2 NUMERICAL COST OF SOLVING THE INVERSE PROBLEM WITH SEM

Our aim is to find an Earth model with the minimum number of parameters that can explain our seismic data set, as well as data not used in the inversion but obtained under similar conditions. By Earth model, we mean the 3-D variations of elastic parameters, anelasticity and density. Let us assume that we wish to solve the inverse problem using a classical least square inversion (Tarantola & Valette 1982) and with a complete modelling theory (i.e. the SEM applied to the wave equation). Let \mathbf{p} be the set of parameters which describe our model. The data set \mathbf{d} is comprised of seismic time traces of N_s events recorded by N_r three component seismometers yielding $3 \times N_s \times N_r$ time series. We call \mathbf{g} the forward modelling function that allows us to model the data for a given set of model parameters: $\mathbf{d} = \mathbf{g}(\mathbf{p})$. In our case \mathbf{g} represents the SEM, which is able to compute a precise set of synthetics in any given model. The inverse problem has to minimize the classical cost function Φ ,

$$\Phi(\mathbf{p}) = {}^t[\mathbf{g}(\mathbf{p}) - \mathbf{d}]C_d^{-1}[\mathbf{g}(\mathbf{p}) - \mathbf{d}] + {}^t(\mathbf{p} - \mathbf{p}_0)C_p^{-1}(\mathbf{p} - \mathbf{p}_0), \quad (1)$$

where \mathbf{p}_0 is the *a priori* value of the model parameters, C_d and C_p are the covariance matrices of data and model parameters respectively. If \mathbf{g} is a nonlinear function, the minimum, or the closest local minimum to the starting model, of Φ , can be found by the Gauss–Newton method iterative process (Tarantola & Valette 1982). Given the model at iteration i , we can obtain model at the iteration $i + 1$:

$$\mathbf{p}_{i+1} = \mathbf{p}_i + ({}^t\mathbf{G}_i C_d^{-1} \mathbf{G}_i + C_p^{-1})^{-1} [{}^t\mathbf{G}_i C_d^{-1} (\mathbf{d} - \mathbf{g}(\mathbf{p}_i)) - C_p^{-1} (\mathbf{p}_i - \mathbf{p}_0)], \quad (2)$$

where \mathbf{G}_i is partial derivative matrix

$$\mathbf{G}_i = \left[\frac{\partial \mathbf{g}(\mathbf{p})}{\partial \mathbf{p}} \right]_{\mathbf{p}=\mathbf{p}_i}. \quad (3)$$

Usually, the forward problem is solved using first-order approximations such as, for example, the Born approximation within the normal mode framework (e.g. Woodhouse & Dziewonski 1984) or arrival time Frechet kernels (Dahlen *et al.* 2000). This leads to a linear relation matrix (\mathbf{G}_0) between the set of parameters and the synthetic data. In that case, only one iteration of (2) is required and the partial derivative matrix is built in the forward theory and

computed at a relatively low numerical cost. Some tomographic approaches are slightly non-linear (e.g. Li & Romanowicz 1996) but are still based on the Born approximation. They also have the advantage of providing naturally the partial derivative matrix with no extra numerical cost, at least at the *a priori* model stage ($i = 0$). When the SEM is used, the partial derivative matrix or kernels cannot be computed naturally. We must use a ‘brute force’ finite difference formulation.

At iteration i of the inversion scheme (3), the line l of the partial derivative matrix \mathbf{G}_i is given by

$$\left[\frac{\partial \mathbf{g}(\mathbf{p})}{\partial \mathbf{p}_l} \right]_{\mathbf{p}=\mathbf{p}_i} \simeq \frac{\mathbf{g}(\mathbf{p}_i + \delta \mathbf{p}_l) - \mathbf{g}(\mathbf{p}_i)}{\delta \mathbf{p}_l}, \quad (4)$$

where \mathbf{p}_i is the parameter vector set in which only the component l is nonzero. Therefore, in order to compute the partial derivative matrix, one needs to compute synthetic data for each seismogram, which involves N_s runs, and for each parameter of the model. This will obviously be the most expensive part of the inversion.

Let us estimate the computing time of such a tomography. Here we assume that only \mathbf{G}_0 is needed and that it can be used throughout the iteration of (2). This assumption is probably valid if the starting model is not too far from the solution and if the problem is only weakly non-linear (Note that if the problem is strongly non-linear, the least squares inversion would probably not converge toward the right solution anyway). Obviously, the determination of \mathbf{G}_0 will dominate the computation time. Let us assume that we wish to build a model with a lateral resolution equivalent to a degree 12 spherical harmonic expansion, which is indeed a modest objective given that current global tomographic models consider expansions up to or beyond degrees 20–24. We consider 10 vertical parameters and invert only for one elastic parameter (S velocity), which means that our model is roughly described by 3000 parameters. We also assume that very long-period waveforms (160 s and above) that will be used here are sensitive up to degree 24 in the sense of a spherical harmonics expansion. This is true in theory, but in practice it is not obvious that the effect of the highest degree is large enough to overcome the background noise. One may have to use higher frequencies to obtain a good result which means our estimation will be optimistic, since numerical cost increase as a power four with the frequency (for a given number of parameters and a given trace length). Let us assume that our data set is comprised of seismograms for 100 events recorded at a large number of receivers (the exact number is not relevant to the numerical cost). Finally, let us assume that we can simulate a 3-hr waveform at a frequency cut-off of 1/160 Hz for a single event in 1 hr, which is roughly what we can do now using a state-of-the-art 16 processors PC cluster. With such a hardware, computing the partial derivative matrix would take $100 \times 3000 \times 1 \text{ hr} \simeq 34 \text{ yr}$. Of course, using a 100 or 1000 times faster computer would reduce the computing time to several weeks, but even if such a computer exists nowadays (i.e. the Earth Simulator, Japan), it would require to use 100 per cent of the machine’s capacity for weeks. Therefore, such an approach is not realistic for the moment.

The seismic exploration community has faced such a problem for waveform tomography, and the Gauss–Newton method to solve the inverse problem is not used in practice because it requires to compute the partial derivative matrix explicitly. Instead, the gradient method is used, which does not require to compute the partial derivative matrix. Indeed, it has been shown that the gradient of the cost function $\Phi (\mathbf{G}_i \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{g}(\mathbf{p}_i)))$ in eq. (2), can be computed with only two forward modelling computations per source using the adjoint problem (Lailly 1983; Tarantola 1984; Tarantola 1988, or more recently, Pratt *et al.* 1998; Tromp *et al.* 2005), which considerably

reduces the number of forward modelling runs despite the fact that the gradient method requires a larger number of iterations than the Gauss–Newton method to converge. The solution can perfectly be applied to our problem but we would not do so in this article. We will investigate another solution that can be cumulated with the adjoint problem to compute gradients. So when reading the next sections, keep in mind that the numerical cost can be even more reduced than what we describe here.

3 DATA REDUCTION

We propose in this paper to use two SEM properties to reduce the numerical cost of a non-linear least square inversion with this tool. First, because the SEM computes the wave field at any location of the Earth, the numerical cost of an inversion is independent of the number of receivers. Second, it is possible in SEM (and in most other direct solution methods) to input several sources in the scheme, to trigger them simultaneously, without increasing the numerical cost. Of course, the resulting traces on the receivers side will be the sum of the traces due to each individual source and there is no possibility to separate them once the computation is done. If we cannot recover the individual synthetic seismograms after the computation, we can perform an equivalent stack of data for common seismometers assuming a common origin time for all the events, and use that reduced data set instead of traces of individual events, as the stacked data are directly comparable to the stacked synthetic seismograms. This operation is possible thanks to the linearity of the wave equation with respect to seismic sources which means that computing traces for one seismometer for each source separately and then summing them is equivalent to computing one trace of all the sources triggered simultaneously. This data reduction scheme allows us to model the whole data set with one SEM simulation with respect to N_s when traces for a common station are not stacked. Finally, note that summing the traces with a common zero time is not necessary (i.e. the sources can be staggered in time), but it is used here to provide a simpler explanation.

If we apply this to our example of Section 2, the 34 yr of computation reduce to 4 months. Of course, this data reduction is not without drawbacks and some information that is contained in independent seismograms will be lost in the summation process. However, we hope that this loss of information will be compensated by the fact that we are able to use all the information present in a long time series for each trace.

4 VALIDATION TESTS

In this section, we present several numerical experiments to assess the robustness of the inversion when the stack data reduction is applied. These tests are circular tests in the sense that the ‘data’ to be inverted are generated with the same forward theory as the one used to invert. We name the model used to generate the data to be inverted the input model or the target model. These tests only provide information on the ability of the process to converge toward the solution under some circumstances (e.g. amplitude of velocity contrast, data coverage, presence of noise etc.) They do not provide any information on the behaviour of the inversion in the case of an incomplete theory, like, for example, how an isotropic inversion would map an anisotropic medium or how high degree horizontal spherical harmonics components (or equivalent) would leak or alias in a low degree inversion. Nevertheless, these tests provide valuable

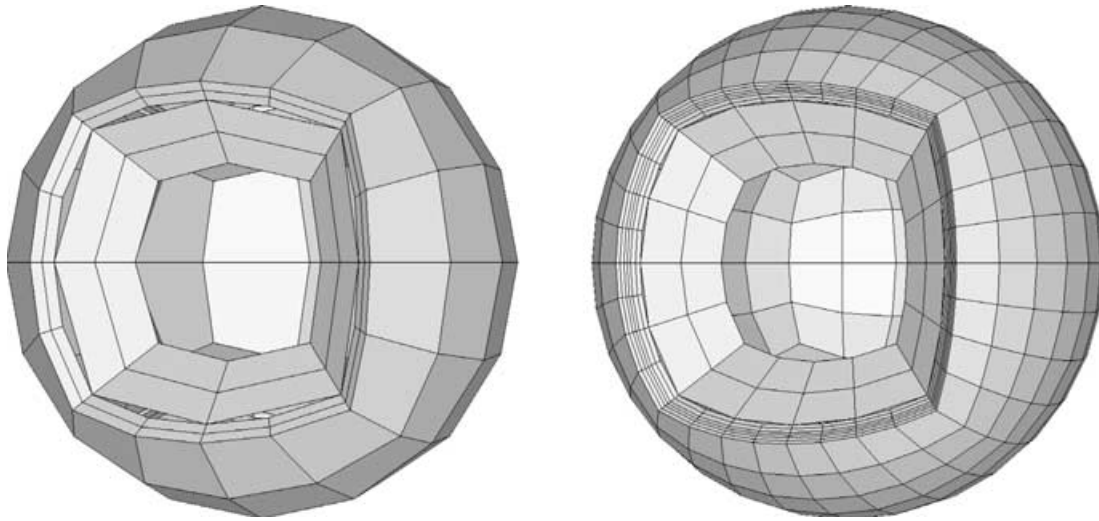


Figure 1. Two examples of meshes of the sphere used to parametrize the velocity model: one with 274 free parameters (left) and one with 2610 free parameters (right).

information on the feasibility of the process. At least if the process failed in these tests, there is little chance that it will ever succeed.

In the following tests, no damping is applied ($\mathbf{C}_d = \mathbf{I}$ and $\mathbf{C}_p^{-1} = 0$) so that the least squares inversion process is simply a Gauss–Newton method to invert \mathbf{g} :

$$\mathbf{p}_{i+1} = \mathbf{p}_i + (\mathbf{G}_i \mathbf{G}_i)^{-1} [\mathbf{G}_i(\mathbf{d} - \mathbf{g}(\mathbf{p}_i))]. \quad (5)$$

In order to limit the numerical cost of these experiments, only \mathbf{G}_0 will be computed and will be used instead of \mathbf{G}_i at iteration i . We will see that this approximation does not hinder the convergence, at least for these tests. Note that if the starting model is spherically symmetric, normal mode perturbation theory would provide an exact solution for \mathbf{G}_0 (Woodhouse 1983) and will be computationally more efficient. In practice we already use that possibility, but this work will be presented in a later publication. Of course, this normal mode perturbation approach is only an option when the starting model is spherically symmetric, which may not be desirable with the present level of sophistication in tomography. Nevertheless, an interesting possibility for 3-D starting models may be to combine the adjoint problem solution mentioned in Section 2 to compute an accurate gradient of the cost function and normal mode perturbation theory to compute the approximate Hessian ($\mathbf{G}_i \mathbf{G}_i$).

4.1 Parametrization

Instead of spherical harmonics or block parametrization, we use a piecewise-polynomial approximation description based on our spectral element discretization (Sadourny 1972; Ronchi *et al.* 1996; Chaljub *et al.* 2003). The sphere is discretized in non-overlapping elements and each of these elements can be mapped on a reference cube. On the reference cube, a polynomial basis is generated by the tensor product of a 1-D polynomial basis of degree $\leq N$ in each direction. The continuity of the parametrization between elements is assured. More details on this discretization mesh can be found in Chaljub *et al.* (2003). Fig. 1 presents two examples of meshes on the sphere used for this parametrization with a polynomial degree over elements $N = 2$. The first mesh (left) has 274 free parameters and roughly corresponds to a spherical harmonic degree 8 horizontally in the upper mantle and a degree 4 horizontally in the lower mantle. The second mesh (right) has 2610 free parameters and roughly

corresponds to a spherical harmonic degree 16 horizontally in the upper mantle and a spherical harmonic degree 8 horizontally in the lower mantle. In practice, this parameterization may not be a good choice, because parameters at the corner of elements have a different spatial spectral content than parameters at the centre of an element. However, for the tests presented here, as the input model is represented on the same mesh as the inversion mesh, this choice does not affect the results.

4.2 Experiments setup and input models

For computation cost reasons, the following experiments have been carried out with the small mesh (274 free parameters) and only one elastic parameter has been inverted (S velocity). We choose a realistic source-receiver configuration of 84 well-distributed events recorded at 174 three-component stations of the IRIS and GEOSCOPE networks (Fig. 2). The corner frequency used here is 1/160 Hz and each trace has a duration of 12 000 s. For each test, the starting model is the spherically symmetric PREM (Dziewonski & Anderson 1981). The partial derivatives matrix \mathbf{G}_0 is therefore the same for all tests and requires 275 SEM runs to be built, which is reasonable in terms of numerical cost (11 days using our hardware example, Section 1).

Two input models will be used. For both of them, the reference background model is PREM to which a 3-D V_s velocity contrast field is added. This 3-D V_s velocity contrast field is generated on the same mesh as the one that will be used for the inversion (Fig. 1 left). The first model is named BIDON (Fig. 3) and is a very simple model: all the parameters are set to zero except one in the upper mantle and one in the lower mantle. The amplitude of the velocity fluctuations is large (9 per cent) compared to what we expect for the Earth for such a long spatial wavelength. On Fig. 3 (left) we plot a depth cross section of the model and on Fig. 3 (right) we plot the V_s velocity as function of the parameter number of the mesh (from 1 to 274). This 1-D representation does not provide a precise idea of what a map of the model would actually look like, but it gives accurate information on the precision of the inversion, which a geographical map does not. The parameter indexes are sorted such that the lower mantle is predominantly on the left side of the plot and the upper mantle predominantly on the right side of the plot to give some information

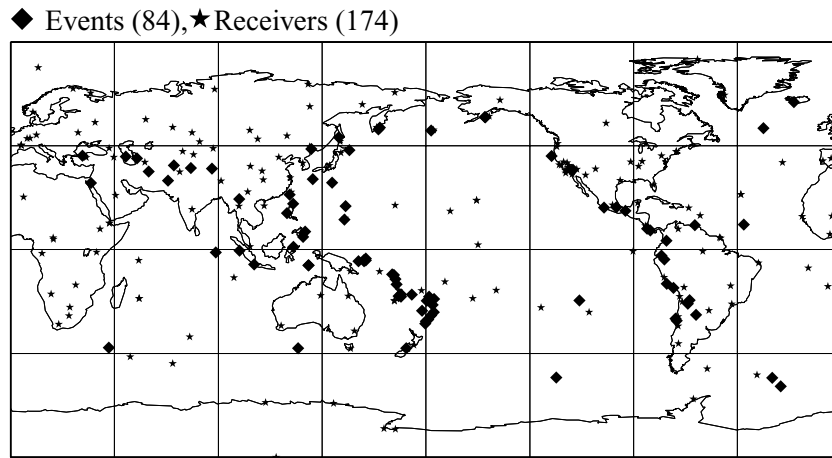


Figure 2. Sources (stars) and receivers (diamonds) configuration used to test the inversion process in this article. A total of 84 earthquakes recorded over 174 three components stations are used.

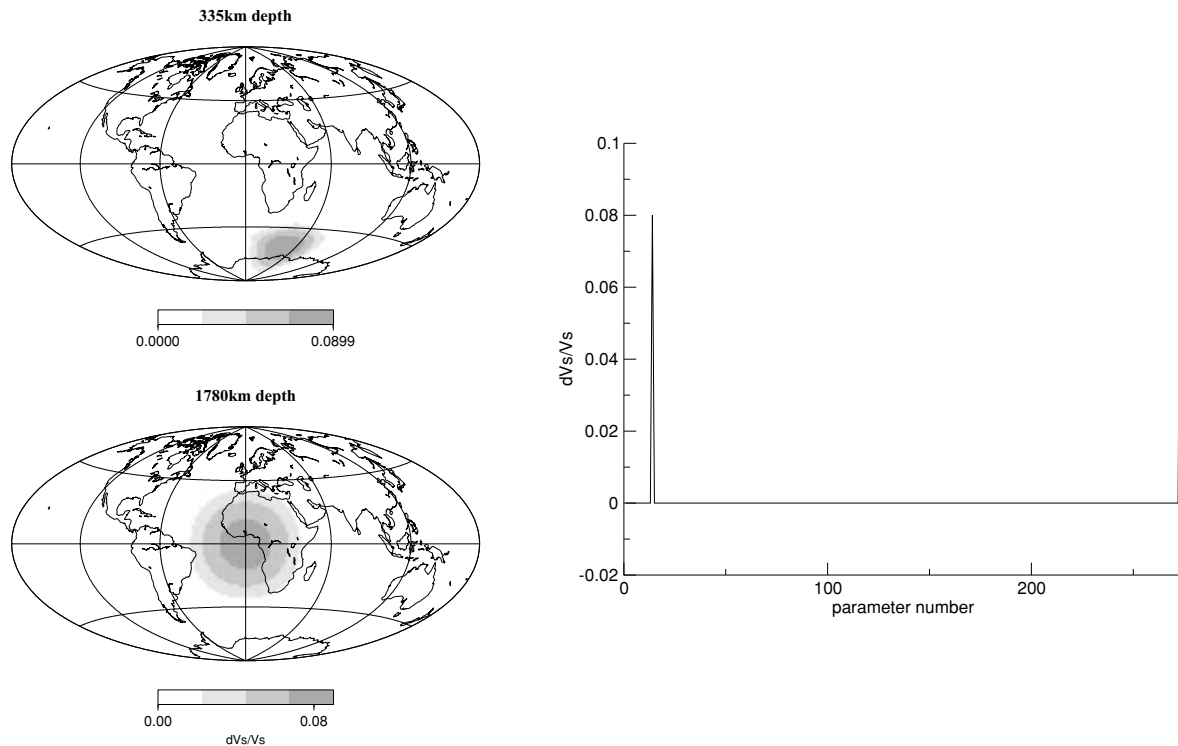


Figure 3. Earth model BIDON. Only two parameters have a velocity contrast with respect to the spherically symmetric reference model (PREM). Left panel shows maps at two different depths and on the right is shown a 1-D representation on the model where the V_s velocity contrast is plotted as a function of the parameter number (from 1 to 274).

about the location of potential errors when looking at these plots. The second model is named SAW6 (Fig. 4) and is more realistic than BIDON. This model is derived from the tomographic model SAW24B16 (Méginn & Romanowicz 2000), truncated at degree 6 and mapped on the 274 parameter mesh (Fig. 1 left). The maximum amplitude velocity contrast is much lower (about 3 per cent) than in BIDON which is typical of long-wavelength mantle heterogeneity. In this case all the parameters have non-zero values as it can be seen on the right plot of Fig. 4.

4.3 Test in BIDON model

Stacked data are generated with SEM in the model BIDON and are inverted following the inversion scheme presented in this paper. The results of the first three iterations of inversion are shown in Fig. 5. The first iteration already gives a velocity contrast very close to the correct value for the two parameters with non-zero velocity contrast, but for the other ones the result is very noisy. The second iteration gives a much better result and the third one has converged

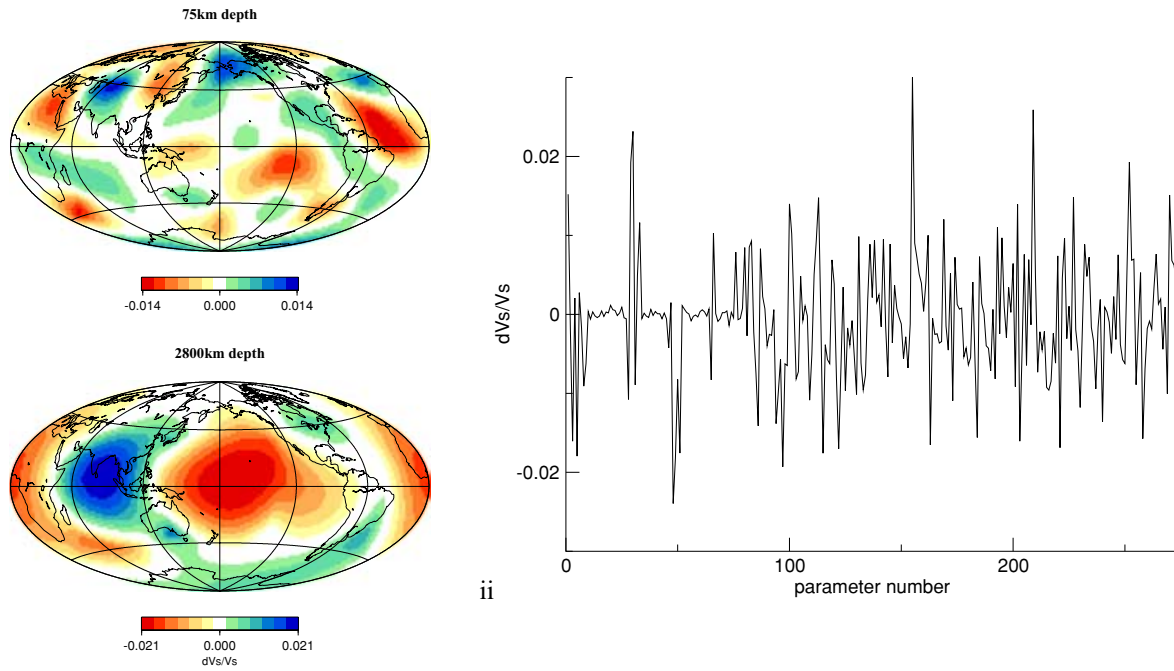


Figure 4. Earth model SAW6. This model is derived from the tomographic model SAW26B16 (Mégnin & Romanowicz 2000). Maps at two different depths (left) and a velocity contrast of each parameters as a function of the parameter number (right) are represented.

toward the correct result. This first experiment is satisfactory and shows that the process can work, at least in simple models. The fact that the first iteration is relatively far from the correct model is interesting because it means that a method based on the first order Born approximation would give a very poor result in that case. The non-linearity is here strong enough to justify a non-linear scheme, but it is weak enough to allow the convergence toward the right solution and not toward a wrong local minimum model, and this without updating the partial derivative matrix G_i at each iteration.

4.4 Test in saw6 model

We now perform the same test but with data generated in the more realistic model SAW6. Results of the first two iterations of the inversion are shown Fig. 6 and already present a good convergence toward the input model for the second iteration. This faster convergence compared to the first test can be explained by the lower velocity contrast of the input model, which implies smaller non-linear effects. All model parameters, from the lower mantle to the surface, are well retrieved.

4.5 Test in saw6 model with noisy data

The purpose of this experiment is to assess the noise sensitivity of the inversion scheme. This kind of test reflects how stable the inversion is, and in this experiment, we are not in a favourable case. Indeed, data with periods 160 s and above have a very poor depth resolution, and to obtain very good results with such an experiment, one should use higher frequency data or decrease the number of vertical parameters. We nevertheless perform the test with SAW6 input model again, but this time synthetic noise is added to the data. To do so, we generate a random noise corresponding to a realistic background noise in this frequency band (the noise spectrum is a

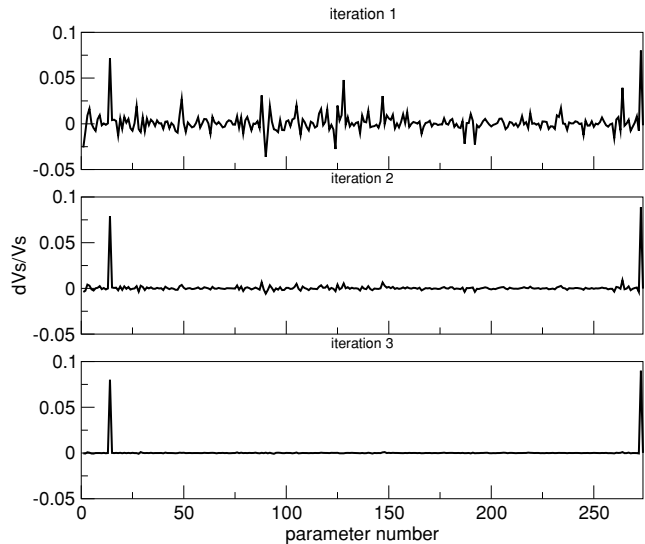


Figure 5. Inversion results for the three first iterations for data generated in the model BIDON (Fig. 3). The velocity contrast with respect to PREM is plotted as a function of the parameter number. The input model is accurately retrieved after three iterations.

slope from -175dB and -165dB in the 100 to 300-s period range) and for each event-station pair, stack them and then add them to the synthetic data. The result of the inversion is shown Fig. 7. The noise affects the results of the inversion, but the scheme is still able to retrieve the target model correctly. The deepest parameters of the model are the most affected by noisy data, which is not a surprise knowing the poor sensitivity to deep layers of long-period data. Fig. 8 shows that, despite the noise, the inversion is able to retrieve a model that explains the data far beyond the noise level. The fact that we are able to fit the data so well, even though the

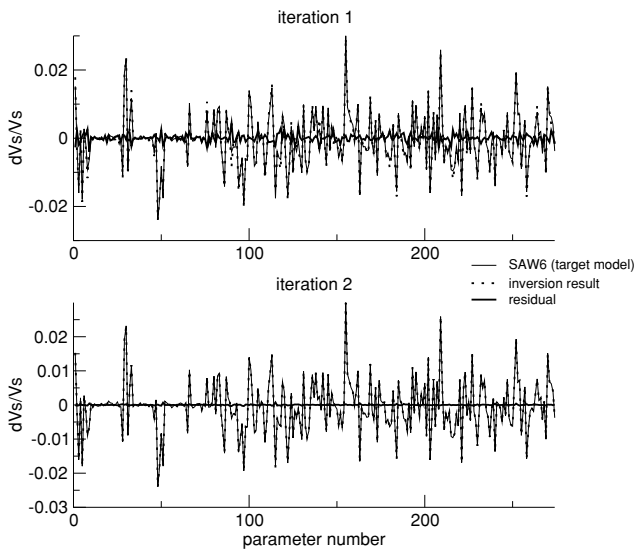


Figure 6. Inversion results for the two first iterations for data generated in the model SAW6 (Fig. 4). After two iterations, the inversion result match very well the input model for all model parameters.

model is not perfect, is also due to the lack of depth sensitivity of long wavelength data.

4.6 Test in SAW6 model with only one station

We perform an extreme test to assess the ability of the process to recover information in the case of very poor data coverage. To do so, we perform a test with only one receiver, the GEOSCOPE station KIP in Hawaii (Fig. 9). This time, no noise is added to traces. The input model is SAW6 (Fig. 4) and the results of the inversion are shown in Fig. 10, in the 1-D representation for iterations 1, 5 and 10. The output model for the first iteration is very far from the input model and at this point it seems that the inversion scheme has no chance to recover it. However after a large number of iterations, the process finally converges toward the input model. It is impressive that the process is still able to converge without updating the partial derivative matrix \mathbf{G}_0 at any iteration. The conclusion of this experiment is that, what allows us to retrieve the input model is not only the wide off path sensitivity of the theory, but also the non-linearity or, in other words, the multiple scattering. Indeed, a Born theory with no geometrical approximation has the same wide off path sensitivity as a direct solution method like SEM, but would give a wrong model (similar to the one which is obtained at iteration one). Clearly, the inversion is in that case highly unstable and a very high data precision is required to allow the inversion to converge toward the right model. An application to real data would be a disaster due to the presence of noise or, equivalently, of physical processes not included in the theory (anisotropy, attenuation, effect of atmospheric pressure etc.).

4.7 Test in SAW6 model with moment tensor errors

So far, a perfect knowledge of the source location, origin time and moment tensor have been assumed. When applying the method to real data, this will not be the case and significant errors on the source parameters can be expected. In order to partly address this issue, we perform a test where the *a priori* moment tensors are not well known but we keep the locations and origin times perfectly known. This

reflects the fact that, at least at very long period, the location and origin time errors are small compared to the wavelength. In this test we generate data in the SAW6 model with each component of each *a priori* moment tensor perturbed by a random coefficient lying between -30 per cent and $+30$ per cent. The moment tensors used to generate the partial derivatives and to compute the forward modelling part of the inversion are, therefore, not the ones that have been used to generate the synthetic data to be inverted. The result of the inversion after three iterations is shown Fig. 11. The scheme can clearly not retrieve the input model. The unknown moment tensors create a large noise that can not be overcome with a reduced number of data. A solution to this problem can be to increase the number of data, and therefore, because the number of stations can not be significantly increased, to use multiple stacked data sets. Another solution is to invert also for the moment tensor at the same time as V_s velocity. In this case, a difficulty due to the stacked data set, is that, for sources close to each other, only the sum of these moment tensors can be retrieved, but not individual moment tensors. If the primary goal of the inversion is to retrieve V_s field, an accurate sum of the moment tensors of sources very close to each other is enough. Indeed, an accurate sum of the moment tensors of sources very close to each will give a correct prediction of the stacked displacement at stations, which is all what we need for a V_s tomography with stacked data. Now, if we are also interested in individual moment tensors, a solution can be to separate sources in the time domain by introducing time delays between close sources. Doing so, different sources located at the same place will have a different effect on stacked data. In this example, we will only focus on retrieving V_s field. In the case of sources very close to each other, we therefore wish to invert only for the sum of the moment tensors. In order to do so, we generate partial derivatives of individual components of moment tensors. The Hessian matrix ($\mathbf{G}_i \mathbf{G}_i$) for moment tensors only is then build, an eigenvalue analysis of this matrix is performed and only the 75 per cent larger eigenvalues are kept. This is equivalent to a damping that removes the instabilities, but it only affects the moment tensor inversion part. Of course the choice of 75 per cent is not precise and will prevent us from explaining the signal perfectly. Therefore, we expect a small error due to this choice to spread into the V_s inverted field. We finally invert for the V_s field at the same time as the moment tensors cleaned from its 25 per cent lower eigenvalues. Fig. 12 shows the result of the inversion. Thanks to the inversion for the moment tensors, we are able to retrieve very well the input model. The remaining errors are due to the 75 per cent choice in the number of kept eigenvalues for the source inversion. Indeed, this choice is not optimum and some signal that should be explained by the moment tensors is not and slightly degrades the V_s inversion.

5 TOWARD REAL CASES: DEALING WITH MISSING DATA

Thanks to the success of all preliminary tests of this inversion scheme, we have already started to work on real data to get a preliminary long-period model. The results of this work will be presented in a later publication. When working with real data, a problem with this kind of approach immediately appears: the missing data. Indeed when trying to gather data for a reasonable number of events (let's say around 50) recorded at a large number of stations (around 80), there are always between 10 and 20 per cent of missing data whatever the configuration is. The reason why the data are sometimes not available at a given station varies from case to case, but there are very few stations that have 100 per cent availability for 50 events.

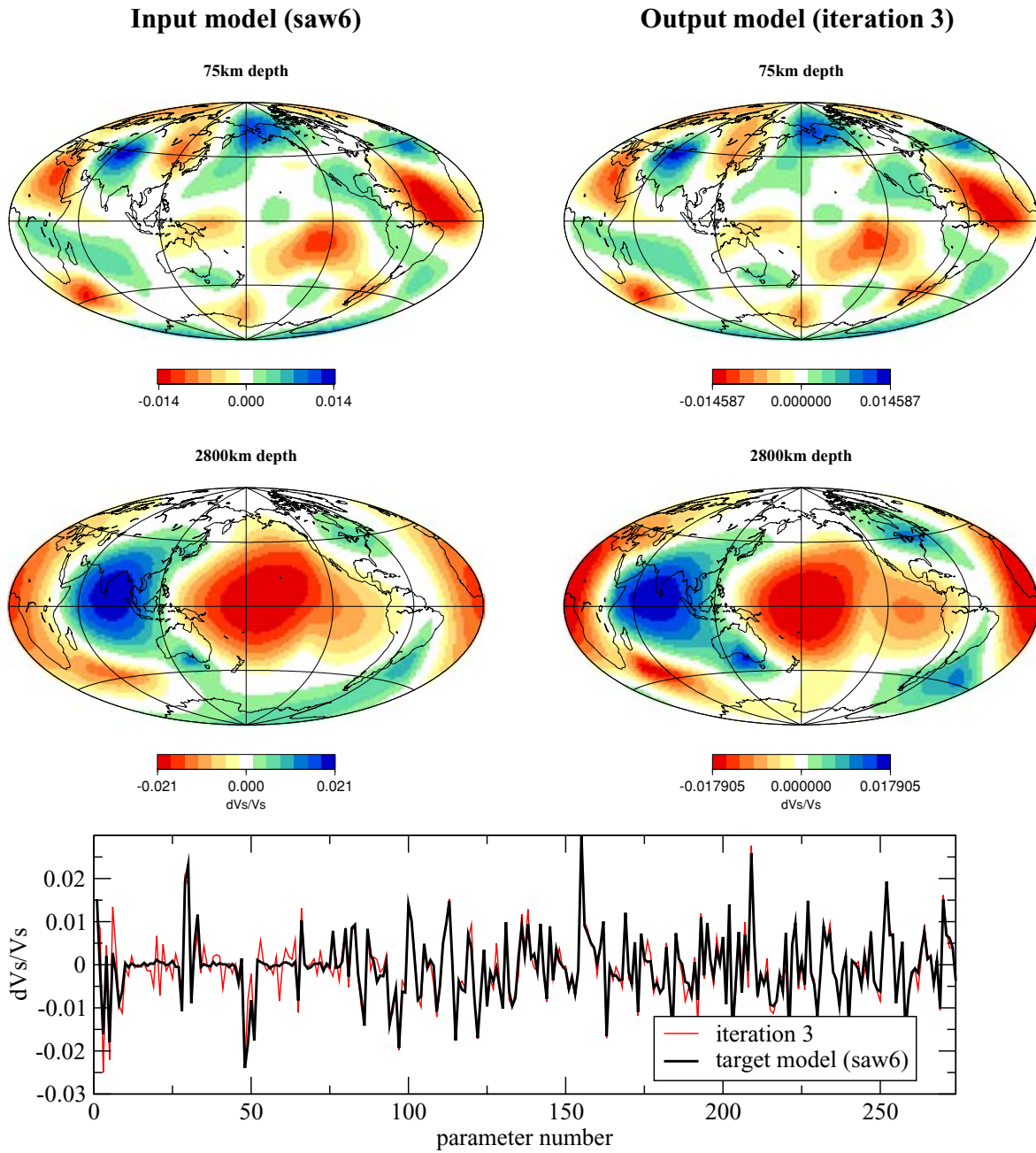


Figure 7. Inversion results for third iteration for data generated in the model SAW6 (Fig. 4) but with noise added to synthetic data. The two left maps show the input model (SAW6) at two different depths and the two right maps show the result of the inversion (output model) at the same depths after three iterations when synthetic noise is added to the synthetic data. The lower plot shows the V_s velocity contrast on the input and output model as a function of the parameter number. Note that the deep parameters are more affected by the noise than the one close to the surface. The model is nevertheless correctly retrieved by the inversion.

When combining the 80 stations, even for large events (magnitude from 6.5 to 7) we end up with about 10–20 per cent of missing data. For our inversion scheme, missing data is a problem as it is easy to generate the sum of all the data in one run but impossible to remove some of them without computing each missing source individually. Since almost all sources are missing at least at one station, removing missing data would require to perform a simulation for each source and it seems we are back at our starting point.

However, the sum of all the data \mathbf{d}_i can be separated into the sum of missing data \mathbf{d}_m and the sum of available data \mathbf{d}_a :

$$\mathbf{d}_i = \mathbf{d}_a + \mathbf{d}_m. \tag{6}$$

The total direct problem can also be separated into missing and available synthetic parts:

$$\mathbf{g}_i(\mathbf{p}) = \mathbf{g}_a(\mathbf{p}) + \mathbf{g}_m(\mathbf{p}). \tag{7}$$

The main difficulty is that there is no way to compute efficiently the partial derivatives matrices of \mathbf{g}_a and \mathbf{g}_m , therefore trying to solve

$$\mathbf{g}_a(\mathbf{p}) = \mathbf{d}_a \tag{8}$$

or

$$\mathbf{g}_i(\mathbf{p}) - \mathbf{g}_m(\mathbf{p}) = \mathbf{d}_a \tag{9}$$

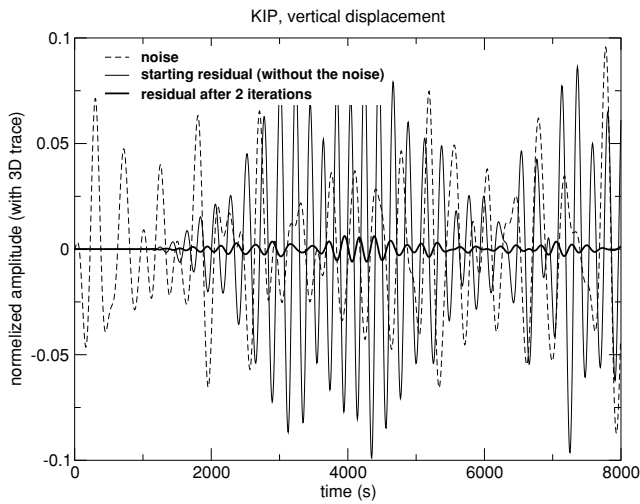


Figure 8. Here is plotted, at KIP station: the noise (dotted line) added to the synthetic vertical component data before inversion; the starting residual signal without the noise (solid line) which is difference between the synthetic data when noise is not added yet and the synthetic seismogram in the reference model (PREM); finally the residual (solid bold) after two iterations that is the difference between the synthetic in the obtained model after inversion and the starting model. We see the scheme is able to go beyond the noise level (the amplitude of the last residual is smaller than the noise level).

is not an option. On the other hand, solving

$$\mathbf{g}_t(\mathbf{p}) = \mathbf{d}_a + \mathbf{g}_m(\mathbf{p}) \quad (10)$$

is possible because the partial derivatives matrix of this last problem only depends on the sum of all the data, the missing and available ones. Nevertheless, since the right hand side of the last equation depends on \mathbf{p} , it requires an iterative scheme. This scheme is only interesting if we do not update the partial derivative matrix at each iteration of the iterative scheme, but since we can expect only a small number of missing data, this should not be a problem. The available solutions for $\mathbf{g}_m(\mathbf{p})$ are:

- (1) 0,
- (2) Synthetics in the spherically symmetric model (\mathbf{g}_m independent of \mathbf{p}) with normal mode summation,

◆ Events (50), ★ Receiver (KIP)

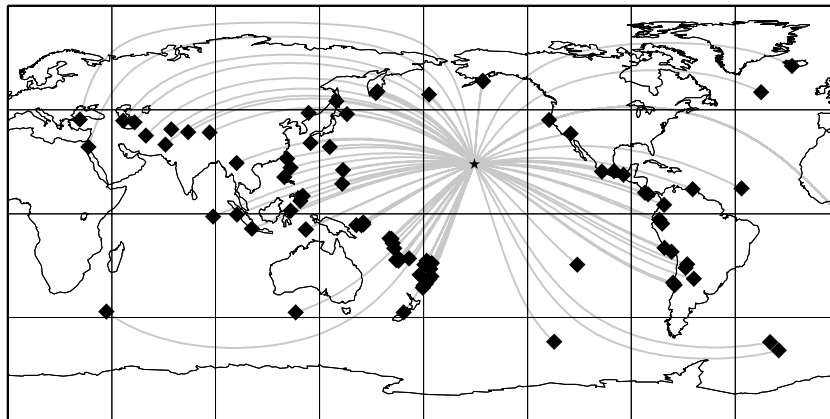


Figure 9. Data coverage used in the single station test. The same number of events (84, plotted as diamonds) as in the other tests is used, but only one station (KIP) is used.

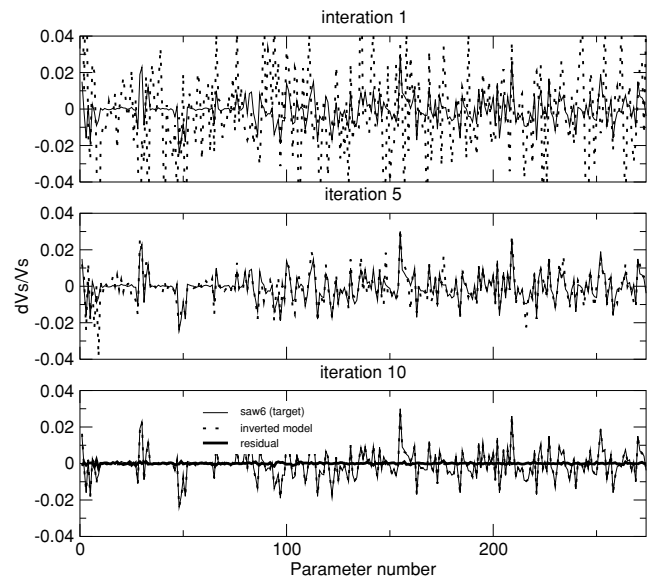


Figure 10. Inversion results with only one station (KIP, see data coverage Fig. 9) for data generated in the model SAW6 (Fig. 4) for iterations 1, 5 and 10. The convergence is slow, but after 10 iterations, the output model match the input model.

(3) Synthetics in \mathbf{p} with normal mode summation first order perturbation and

(4) Synthetics in \mathbf{p} with the spectral element method.

The first two solutions are numerically inexpensive but probably not very good, depending on the amount of missing data. The third one is probably a good compromise between numerical cost and precision and the last one is perfect but expensive. An equally good solution for a finalized model may be to use normal mode perturbation theory during the iterative process of the inversion and spectral elements synthetics at the last iteration.

In order to test this solution, we generate a data set in the model SAW6 (Fig. 4) with missing data. To do so, we first compute synthetic seismograms from each source individually with 84 runs. We then select randomly the missing data among each source receiver pair. The selected data are not used in the construction of the stacked

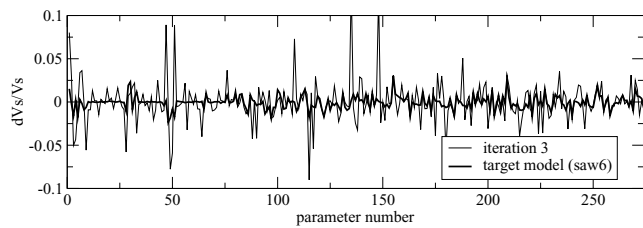


Figure 11. Inversion results for third iteration for data generated in the model SAW6 (Fig. 4, bold line) with error on *a priori* moment tensors. To generate the data, a random coefficient lying between -30 per cent and $+30$ per cent has been applied to each component of each moment tensor. The result of the inversion (without inverting for sources) after three iterations (thin line) doesn't match the input model.

data set. We perform here a rather extreme case with 35 per cent of missing data (our experience with real cases have shown that it is possible to gather data set with 15–20 per cent of missing data when working with 50 sources and 90 vertical component receivers). We then perform three inversions. For the first one, the missing data are replaced with synthetic seismograms computed in the starting spherically symmetric model. This solution is numerically interesting because the scheme is still explicit: the data set completed with the synthetics of missing data does not depend on the obtained model. The drawback is that the final model cannot be accurate as the signature of the starting model will always be present and can not be corrected. The result of such an inversion is given Fig. 13. As expected, the result is noisy, but the main features of the input model are still retrieved. In the second inversion, the missing data are replaced by synthetic seismograms computed in the current model (\mathbf{x}_i) with the Born approximation in the normal modes framework (Capdeville *et al.* 2000; Capdeville 2005). This time the scheme becomes implicit as the data set completed with the synthetics of missing data depends on the current model. The result of such an inversion is given Fig. 14. The result is much better than for the previous test but still not perfect. This is expected as the first order Born approximation is not very accurate especially when time series are long and non-linear effect cannot be neglected anymore. Finally we perform a last test in which the missing data are replaced by synthetics computed in the current model with SEM. This solution is CPU time consuming as it requires to compute each source individually. To perform such a test and to lower the numerical cost we start from the model obtained at the last iteration of the previous test and we perform only one extra iteration. The result is given Fig. 15 and shows a very good result. Some more iterations would be required to obtain the same precision as the one we get when no data are missing, but this result is already accurate enough with respect to noise coming from other effects (noise, error on moment tensors).

6 DISCUSSION AND CONCLUSIONS

In this paper, we have presented a way to perform non-linear full waveform inversion at the global scale, using the spectral element method as a forward modelling tool. This method is based on a non-coherent trace stacking at a common receiver for a common source origin time. The data reduction allows us to simulate the whole data set in a single Spectral Elements run and, therefore, to reduce the number of computations by a factor equal to the number of sources with respect to a classical approach. We have presented preliminary tests which show very promising results. The main advantage of the

approach is that it allows us to investigate full waveform tomography now, without waiting to have the computing power to use a classical approach. There is also a data selection and processing advantage. Phase identification, time picking, phase velocity measurements are not required for waveform inversions which saves a lot of time and also minimizes human error.

Clearly, there are also drawbacks to this approach. One of them is that some information is lost in the data reduction. However all tomographic methods use some data reduction scheme. For example, travel time tomography uses only a limited number of arrival times per trace (often only one or two). Here, we use traces of 12 000-s duration, which represents a lot of information, even with a 160 s corner period. We, therefore, hope that these long traces can overcome the loss of information due to the stacking. Nevertheless it is important to keep in mind that information is lost when that data is stacked. If only a single stacked data set is used, there is a limit on number of sources, after which adding new sources does not bring anything new. This limit is not obvious to address and depends on the corner frequency and the length of the signal that is used. However when this limit is reached, the only way to get more information about the model (e.g. to improve the resolution) will be to use multiple stacked source data sets. Note that the process has the advantage to allow to move gradually toward the classical case (no data stacking) by splitting the data set into two or more data subsets as a function of the computing power available. Through this process, we will end up eventually with the classical case where all the sources are considered individually.

Another drawback is that the process does not allow us to select some time windows on traces to enhance some part of the signal with respect to others, such as separating body wave packages from each other and from surface waves (e.g. Li & Romanowicz 1996). The body waves have small amplitude but contain information about the lower mantle whereas surface waves have a large amplitude but do not contain information about the lower mantle. As surface waves will dominate the stacked signal, there is little chance to recover the lower mantle before the upper mantle is very well explained. We have seen that it is not a problem in our tests, but this is because we exactly know what to invert in order to explain exactly the upper mantle and therefore once the upper mantle is explained, the lower mantle is easy to retrieve. In a real case, it may be much more difficult, since we do not know for sure what elastic parameters are required to explain surface waves well enough, and therefore to be able to access body waves and information about the lower mantle.

This last point leads to another difficulty that we will face in future work. What physical parameters (elastic, anelastic, density, etc.) do we need to invert for and at what resolution to explain our data set correctly? The resolution issue is not obvious: a too low resolution for a given data frequency content will lead to aliasing and a too high resolution may lead to an unstable inversion scheme, as our data set may not have the information to resolve all the parameters, but also to a prohibitive extra numerical cost. The number of physical parameters is also a difficult question. Are V_s fluctuations enough to explain our data set? Probably not. Do we need V_s , V_p , density, anisotropy, 3-D anelasticity, perturbations in source parameters? What is the relative sensitivity of our data set to those parameters? All these questions will need to be addressed in future work.

Finally, it is well known that the choice of the type of least squares inversion can strongly constrain the possible model. This is not really a choice as we cannot afford the numerical cost of a more general inversion scheme based on random exploration of our parameter

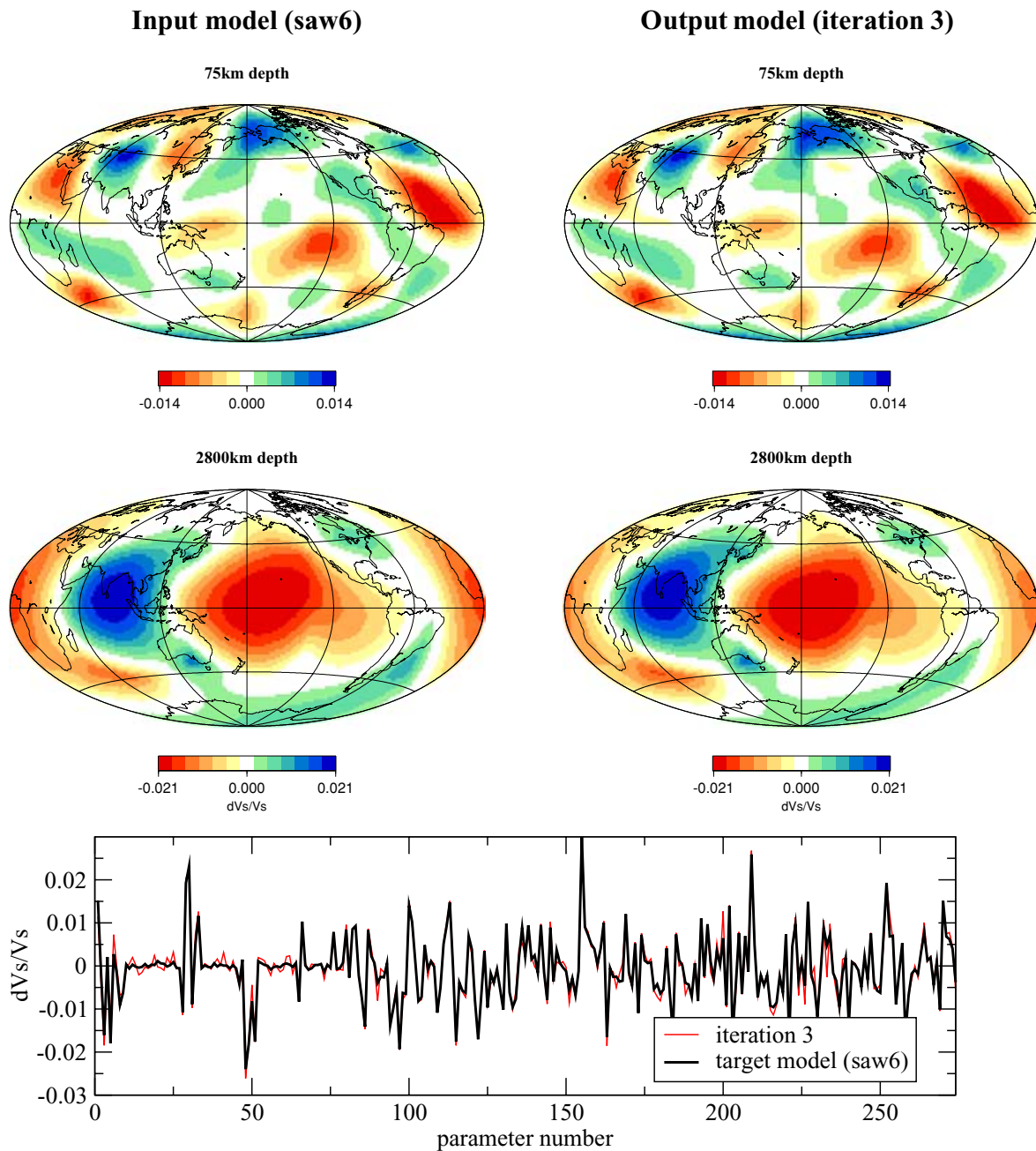


Figure 12. Same test as the one presented Fig. 11 with error on *a priori* moment tensors, but this time a moment tensor inversion joint with the V_s inversion is performed. The two left maps show the model that has been used to generate that data (SAW6, input model) at two different depths and the two right maps show the result of the inversion (output model) at the same depth after three iterations. The lower plot shows the V_s velocity contrast on the input and output model as a function of the parameter number. When moment tensors are inverted as the same time as V_s , the inversion is able to retrieve correctly the input model despite large errors on *a priori* moment tensors.

space. Nevertheless the question of what we may be missing due to this least square inversion scheme, that is, what are the error bars on the obtained model values, will also need to be addressed in future work.

ACKNOWLEDGMENTS

The authors would like to thank to Jeroen Ritsema for helping with the manuscript. Many thanks to V. Maupin, P. Sanchez-Sesma, E. Beucler and many others for helpful discussions. We also thank T. Tanimoto and an anonymous reviewer for very useful comments. It is

BSL contribution #05–10. Computations have been performed on ‘seaborg’, the supercomputer of the NERSC (California), ‘zahir’ the supercomputer of the IDRIS (France), ‘regatta’ the supercomputer of the CINES (France) and at the DMPN (IPGP, France) computing facility.

REFERENCES

Capdeville, Y., 2000. Méthode couplée éléments spectraux—solution modale pour la propagation d’ondes dans la Terre à l’échelle globale, *PhD thesis*, Université Paris 7.

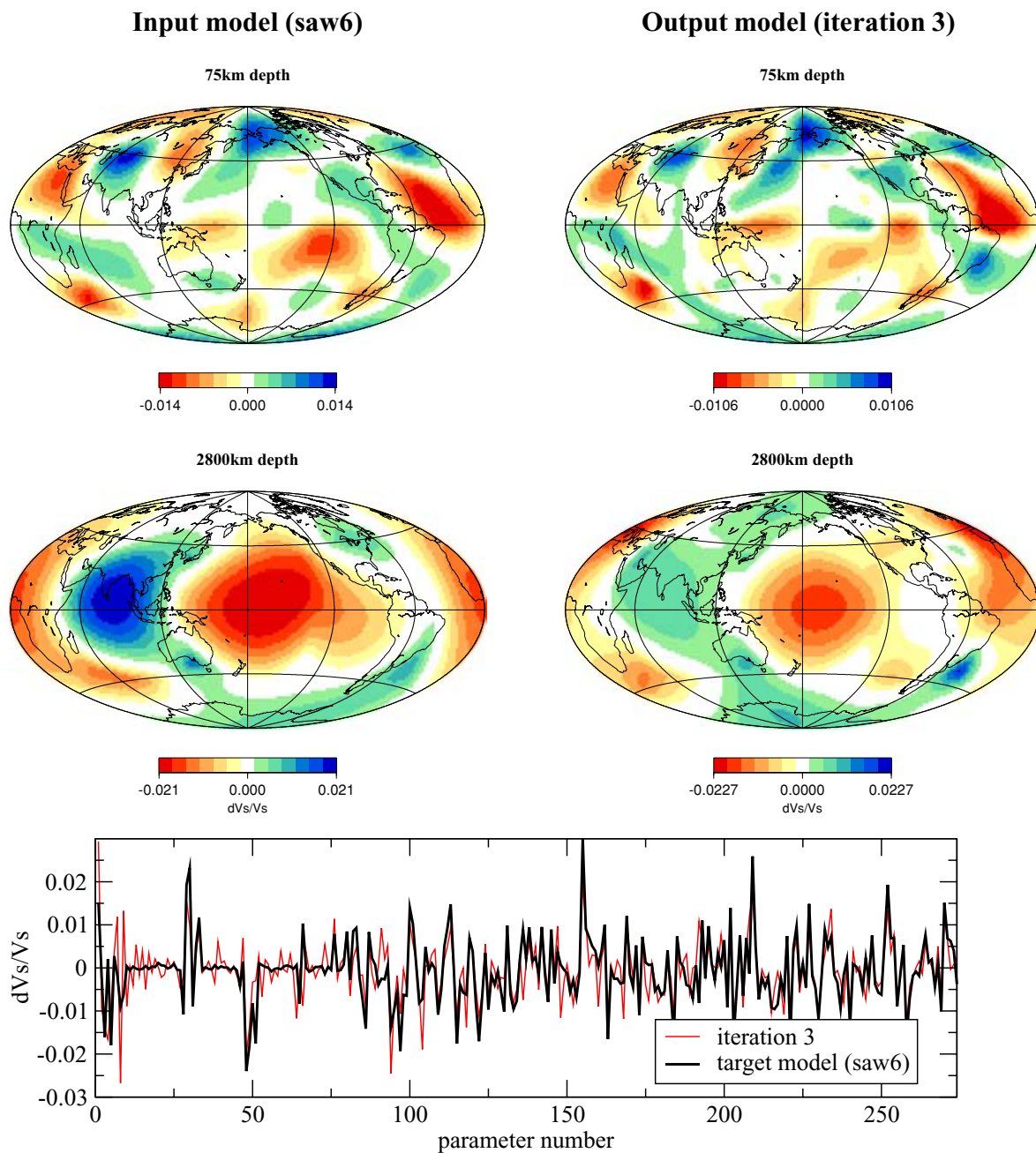


Figure 13. Inversion results for third iteration for data generated in the model *saw6* (Fig. 4, bold line) but with 35 per cent of missing data. The solution adopted to deal with the missing data in this test is to replace the missing data by synthetics computed in the starting model (1-D). They are not updated during the iterative inversion. The result is noisy, which is expected with this solution for missing data. The general features of the input model are nevertheless retrieved.

- Capdeville, 2005. An efficient Born normal mode method to compute sensitivity kernels and synthetic seismograms in the Earth, *Geophys. J. Int.*, **000**, 00–00. submitted.
- Capdeville, Y., Stutzmann, E. & Montagner, J.P., 2000. Effect of a plume on long-period surface waves computed with normal modes coupling, *Phys. Earth planet. Inter.*, **119**, 57–74.
- Capdeville, Y., Gung, Y. & Romanowicz, B., 2002. The Coupled Spectral Element/Normal Mode Method: application to the testing of several approximations based on normal mode theory for the computation of seismograms in a realistic 3-D Earth, in *Eos Trans.*, **83**(47), of Fall Meeting Supplement. AGU. Abstract S51C-10.
- Capdeville, Y., Chaljub, E., Vilotte, J.P. & Montagner, J.P., 2003a. Coupling the spectral element method with a modal solution for elastic

- wave propagation in global earth models, *Geophys. J. Int.*, **152**, 34–66.
- Capdeville, Y., Romanowicz, B. & To, A., 2003b. Coupling spectral elements and modes in a spherical earth: an extension to the sandwich case, *Geophys. J. Int.*, **154**, 44–57.
- Chaljub, E., 2000. *Modélisation numérique de la propagation d'ondes sismiques à l'échelle du globe*. Thèse de doctorat de l'Université Paris 7.
- Chaljub, E., Capdeville, Y. & Vilotte, J., 2003. Solving elastodynamics in a solid heterogeneous 3-Sphere: a spectral element approximation on geometrically non-conforming grids, *J. Comp. Physics*, **183**, 457–491.
- Cummins, P.R., Takeuchi, N. & Geller, R.J., 1997. Computation of complete synthetic seismograms for laterally heterogeneous models using the Direct Solution Method, *Geophys. J. Int.*, **130**, 1–16.

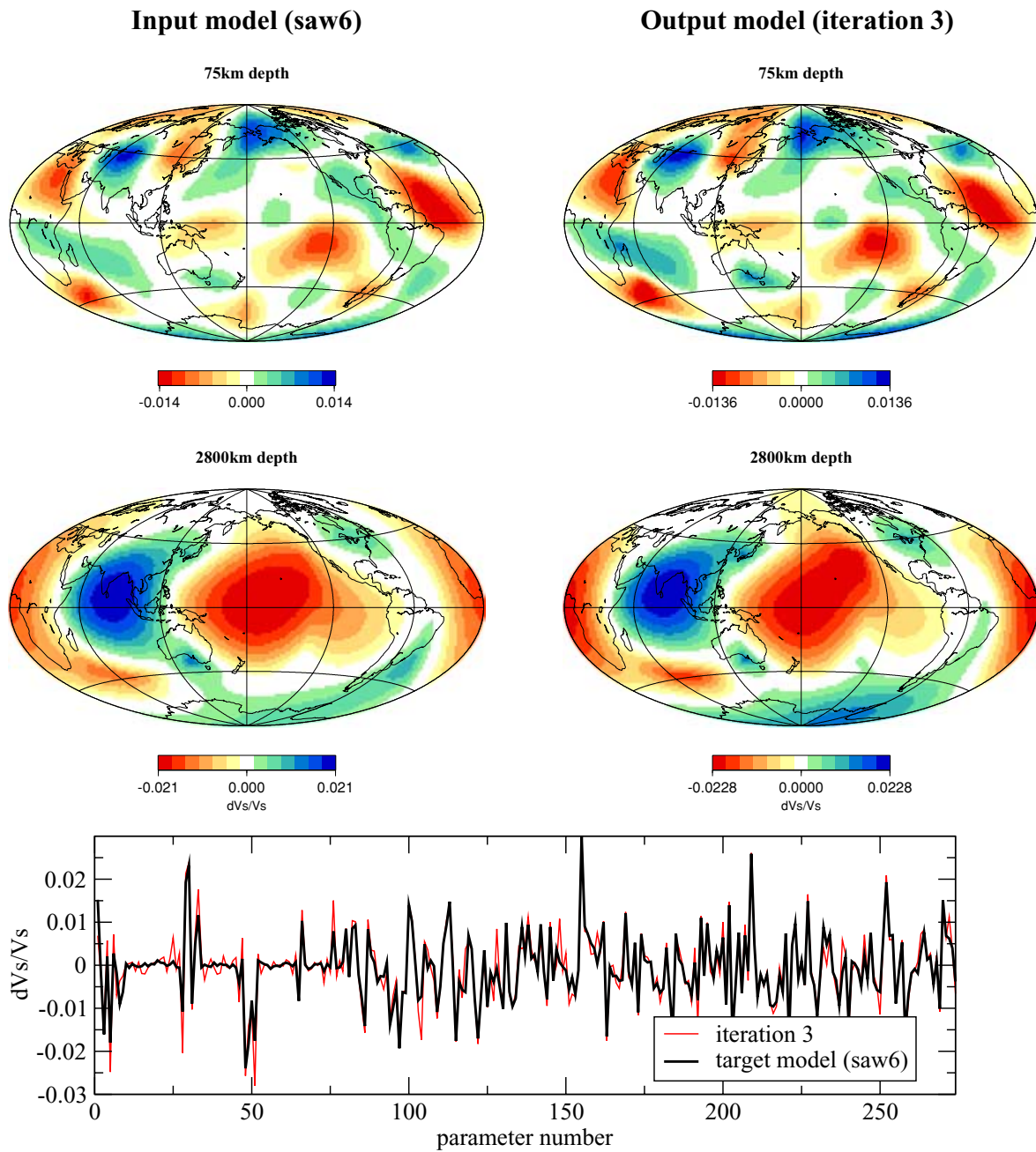


Figure 14. Same as Fig. 13 with 35 per cent of missing data, but solution adopted to deal with the missing data in this test is to replace the missing data by synthetics computed in the current model (3-D) with the born approximation. They are updated during the iterative inversion. The result is good but not perfect, which is expected with this solution for missing data.

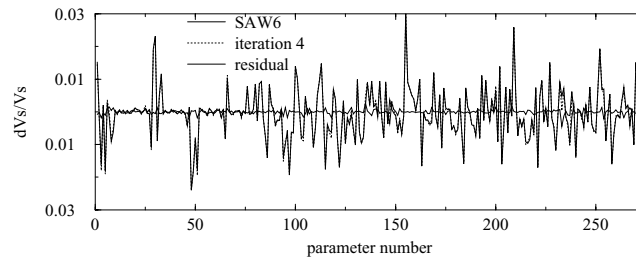


Figure 15. Same as Fig. 13 with 35 per cent of missing data. This time the missing data are replaced by synthetics computed in the current model with SEM at the iteration 4 starting from the iteration 3 of the precedent test (Fig. 14). The result (dotted line) is in a good agreement with the target model.

- Dahlen, F., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes—I. theory, *Geophys. J. Int.*, **141**, 157–174.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**, 297–356.
- Gung, Y.C. & Romanowicz, B., 2004. Q tomography of the upper mantle using three component long-period waveforms, *Geophys. J. Int.*, **157**, 813–830.
- Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for 3-D seismic wave propagation, *Geophys. J. Int.*, **139**, 806–822.
- Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation, part II: 3-D models, oceans, rotation, and gravity, *Geophys. J. Int.*, **150**, 303–318.
- Komatitsch, D. & Vilotte, J.P., 1998. The spectral element method: an effective tool to simulate the seismic response of 2d and 3d geological structures, *Bull. seism. Soc. Am.*, **88**, 368–392.
- Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on inverse scattering: theory and application*, eds Bednar, J., Redner, R., Robinson, E. & Weglein, A., Soc. Industr. appl. Math., Philadelphia, PA.
- Li, X.D. & Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. Int.*, **121**, 695–709.
- Li, X.B. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.*, **101**, 11 245–22 271.
- Li, X.D. & Tanimoto, T., 1993. Waveforms of long-period body waves in slightly aspherical Earth model, *Geophys. J. Int.*, **112**, 92–102.
- Lognonné, P., 1989. *Modélisation des modes propres de vibration dans une Terre anélastique et hétérogène: théorie et application*, Thèse de doctorat, Université Paris VII.
- Lognonné, P. & Romanowicz, B., 1990. Modelling of coupled normal modes of the Earth: the spectral method, *Geophys. J. Int.*, **102**, 365–395.
- Mégnin, C. & Romanowicz, B., 2000. The 3-D shear velocity structure of the mantle from the inversion of body, surface and higher modes wave forms, *Geophys. J. Int.*, **143**, 709–728.
- Millot-Langet, R., Clévéde, E. & Lognonné, P., 2003. Normal modes and long-period seismograms in a 3-D anelastic elliptical rotating Earth, *Geophys. Res. Lett.* **30**(5), 1202, doi: 10.1029/2002GL016257.
- Montelli, R., Nolet, G., Dahlen, F., Masters, G., Engdahl, E. & Hung, S., 2004. Finite-frequency tomography reveals a variety of mantle plumes, *Science*, **303**, 338–343.
- Pratt, R., Shin, C. & Hicks, G., 1998. Gauss–Newton and full newton methods in frequency domain seismic waveform inversion, *Geophys. J. Int.*, **133**, 341–362.
- Romanowicz, B., 1987. Multiplet-multiplet coupling due to lateral heterogeneity: asymptotic effects on the amplitude and frequency of the Earth’s normal modes, *Geophys. J. R. Astron. Soc.*, **90**, 75–100.
- Romanowicz, B., 2003. Global mantle tomography: progress status in the last 10 yr, *Annu. Rev. Geoph. Space Phys.*, **31**, 303–28.
- Ronchi, C., Iacono, R. & Paolucci, P.S., 1996. The ‘Cubed Sphere’: a new method for the solution of partial differential equations in spherical geometry, *J. Comput. Phys.*, **124**, 93–114.
- Sadourny, R., 1972. Conservative finite-difference approximations of the primitive equation on quasi-uniform spherical grids, *Mon. Weather Rev.*, **100**, 136–144.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.
- Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure appl. Geophys.* **128**(1/2), 365–399.
- Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.*, **20**, 219–232.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Woodhouse, J., 1983. The joint inversion of seismic wave forms for lateral variations in Earth structure and earthquake source parameter. In *Physics of the Earth’s interior*, Vol. 85, Amsterdam, pp. 366–397. Int. School of Physics ‘Enrico’ Fermi: North-Holland.
- Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle: Three-dimensional modelling of earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89**, 5953–5986.
- Zhao, L., Jordan, T.H. & Chapman, C.H., 2000. 3-D Fréchet differential kernels for seismic delay times, *Geophys. J. Int.*, **141**, 558–576.
- Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. 3-D sensitivity kernels for surface wave observables, *Geophys. J. Int.*, **158**, 142–168.