

# A flexible, extendable, modular and computationally efficient approach to scattering-integral-based seismic full waveform inversion

this is the accepted version of a paper originally published by Oxford University Press in *Geophysical Journal International* 2016 204 (2): 1100-1119  
doi:10.1093/gji/ggv505

F. Schumacher<sup>1</sup>, W. Friederich<sup>1</sup>, S. Lamara<sup>1</sup>  
<sup>1</sup>Ruhr-Universität Bochum, Germany  
E-mail: florian.schumacher@rub.de

## Abstract

We present a new conceptual approach to scattering-integral-based seismic full waveform inversion that allows a flexible, extendable, modular and both computationally and storage-efficient numerical implementation. To achieve maximum modularity and extendability, interactions between the three fundamental steps carried out sequentially in each iteration of the inversion procedure, namely solving the forward problem, computing waveform sensitivity kernels and deriving a model update, are kept at an absolute minimum and are implemented by dedicated interfaces. To realize storage efficiency and maximum flexibility, the spatial discretization of the inverted earth model is allowed to be completely independent of the spatial discretization employed by the forward solver. For computational efficiency reasons, the inversion is done in the frequency domain.

The benefits of our approach are as follows: (1) Each of the three stages of an iteration is realized by a standalone software program. In this way, we avoid the monolithic, unflexible and hard-to-modify codes that have often been written for solving inverse problems. (2) The solution of the forward problem, required for kernel computation, can be obtained by any wave propagation modeling code giving users maximum flexibility in choosing the forward modeling method. Both time-domain and frequency-domain approaches can be used. (3) Forward solvers typically demand spatial discretizations that are significantly denser than actually desired for the inverted model. Exploiting this fact by pre-integrating the kernels allows a dramatic reduction of disk space and makes kernel storage feasible. No assumptions are made on the spatial discretization scheme employed by the forward solver. (4) In addition, working in the frequency domain effectively reduces the amount of data, the number of kernels to be computed and the number of equations to be solved. (5) Updating the model by solving a large equation system can be done using different mathematical approaches. Since kernels are stored on disk, it can be repeated many times for different regularization parameters without need to solve the forward problem, making the approach accessible to Occam's method. Changes of choice of misfit functional, weighting of data and selection of data subsets are still possible at this stage.

We have coded our approach to FWI into a program package called ASKI (Analysis of Sensitivity and Kernel Inversion) which can be applied to inverse problems at various spatial

scales in both Cartesian and spherical geometries. It is written in modern FORTRAN language using object-oriented concepts which reflect the modular structure of the inversion procedure. We validate our FWI method by a small-scale synthetic study and present first results of its application to high-quality seismological data acquired in the southern Aegean.

## Keywords

Inverse theory; Computational seismology; Wave propagation; Seismic tomography; Full waveform inversion; Sensitivity kernels

## 1. Introduction

Full seismic waveform inversion (FWI) methods aim at explaining the complete information content of measured seismic waveforms by earth structure. They are computationally demanding, but thanks to the increasing availability of computational resources, they have become more and more popular in the last decade. Established FWI methods are formulated as deterministic iterative minimization problems where the desired earth model is found as the minimum of some misfit functional. Currently, FWI is realized in two different variants: Gradient-based and Newton-like or Gauss-Newton methods.

In gradient-based methods, the gradients of the misfit functional with respect to earth model properties are used to iteratively update the earth model. Early applications of gradient-based FWI used steepest descent methods (Bamberger et al., 1982; Gauthier et al., 1986) which straightly follow the gradient direction in the model space and can have slow convergence properties. Later, more elaborate preconditioned conjugate-gradient schemes (Tarantola, 1987) were applied to 3D-elastic full waveform inversion (Mora, 1987; Tromp et al., 2005; Fichtner et al., 2009; Butzer et al., 2013). The gradient of the misfit functional is commonly calculated with the adjoint-wavefield method (Lailly, 1983; Tarantola, 1984). It is inferred from combining the incident wavefield radiated from a seismic source with the adjoint wavefield obtained by backpropagating data residuals from the receiver positions into the medium. Some implementations of the adjoint method store the required wavefield information throughout the medium (Butzer et al., 2013), others memorize only part of the information and reconstruct the complete wavefields again when computing the gradient (Liu and Tromp, 2006). Additional solutions of the forward problem are required in order to determine how far to proceed in the model space along the direction of the (conjugate) gradient.

The most significant advantage of the adjoint approach over other methods is its storage efficiency. However, inversion codes based on it tend to be monolithic because they calculate the misfit gradient on the fly during back-propagation of the data residuals. Hence, at least back-propagation and gradient computation are done in one program, separation of these two steps becomes impossible and gradient calculation is restricted to the one chosen forward solver. Since the gradient, which is proportional to the model update, is computed using the spatial discretization of the forward solver, the latter is forced upon the inverted model, although very different representations of the inverted model may be desirable. Adding damping or smoothing terms to the misfit functional is possible, but varying smoothing and damping within one iteration is cumbersome, as gradient calculation implies running the expensive forward solver. Finally, changing the weighting of data or selecting subsets of the data requires an expensive recalculation of the gradient,

and changing the misfit functional during iterations is entirely impossible.

Alternatives to gradient-based methods are Newton-like (Liu and Nocedal, 1989; Pratt et al., 1998; Fichtner and Trampert, 2011) or Gauss-Newton approaches (Akcelik et al., 2002; Epanomeritakis et al., 2008; Chen et al., 2007) which employ approximations to the Hessian of the misfit functional for updating the earth model. The latter can be obtained e. g. by extending the adjoint wavefield method (Fichtner and Trampert, 2011) or by using the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (“L-BFGS”, e.g. Monteiller et al., 2015; Hanasoge and Tromp, 2014). Advantages of the Newton or Gauss-Newton FWI methods are their enhanced convergence which can be close to quadratic compared to the linear convergence of gradient-based approaches.

A specific variety of the Gauss-Newton approach to FWI is the scattering-integral (SI) method (Chen et al., 2007) which uses Fréchet derivatives of the seismic displacement field with respect to earth model properties to construct an approximation of the Hessian. Traditionally, the Fréchet derivatives are often called Born or waveform sensitivity kernels. For example, Nissen-Meyer et al. (2007), Chen et al. (2007), Zhao and Chevrot (2011) and Fuji et al. (2012) derive and apply such full waveform sensitivity kernels. The SI approach to FWI is very well suited for a strict organizational separation of the 3 basic inversion steps, namely solving the forward problem, computing kernels and deriving a model update. If kernel storage is feasible, it allows sensitivity and resolution analysis based on the kernels before data acquisition. As shown by Chen et al. (2007), in terms of forward computations it is even more efficient than the adjoint-wavefield method except for the case that the number of receivers strongly exceeds the number of sources. Moreover, since the model update is not directly constructed from the misfit gradient but by using a regularized inverse of an approximation to the Hessian, the SI approach does not require preconditioning of the kernels to suppress high sensitivities close to source and receiver positions. Its one severe disadvantage is its enormous need of disk space to store the kernels and the resulting need to perform inefficient input/output operations to read and write wavefields and kernels.

In this paper, we focus on the scattering-integral-based FWI and develop a new conceptual approach which aims at maximizing flexibility, modularity, extendability, storage efficiency and computational efficiency of the SI-based inversion procedure. Our approach works in the frequency domain and is characterized by a very strict organizational separation of the 3 basic steps of the inversion procedure. Necessary interactions are kept at a minimum and are realized by specific interfaces which ensure maximum flexibility and extendability of the implementation. Ultimately a very modular structure of the resulting code results in this way.

The most important new aspect of our approach is that we introduce a spatial discretization of the inverted earth model which is entirely independent of the spatial discretization employed by the forward solver. In this way, storage required for the waveform sensitivity kernels can be minimized by pre-integrating the kernels over the basis functions that represent the inverted model. No assumptions are made with regard to the spatial discretization used by the forward solver.

Benefits of the new approach are, for example, straightforward extendability to any forward solver, change of forward solver between iterations, sensitivity and resolution analysis before data acquisition, flexible playing with regularization without need for solving the forward problem, flexible selection and weighting of data, flexible selection of frequencies for inversion and adjustment of spatial discretization of the forward solver and inverted model to the desired maximum frequency within

each iteration. Even the misfit functional could be changed during iterations.

The paper starts with a short review of scattering-integral-based FWI. Then, a detailed description and discussion of our new approach follows. The final section gives a general description of the software package ASKI – Analysis of Sensitivity and Kernel Inversion (Schumacher, 2015) written according to the new concept and provides some validating synthetic examples of SI-based FWI carried out using ASKI.

## 2. Theoretical aspects of scattering-integral based full seismic waveform inversion

Basic aim of seismic full waveform inversion is to find an earth model that explains observed seismic waveforms in some well-defined sense. We assume here that the observed waveforms are represented by data values  $d_i$  with errors  $\sigma_i$  which may be direct samples of the observed seismogram or the result of some mathematical transformation of the seismograms. Given an earth model, we can compute predictions  $s_i$  of the data which we formally express as the result of functionals  $F_i[\mathbf{m}]$  acting on the earth model  $\mathbf{m}$ :

$$s_i = F_i[\mathbf{m}], \quad i = 1, \dots, N_D. \quad (1)$$

Evaluating  $F_i[\mathbf{m}]$  constitutes the forward problem. The number of data is denoted by  $N_D$ . The model  $\mathbf{m}$  is considered as element of an infinitely dimensional linear vector space  $V$  of vector-valued functions

$$\mathbf{m}(\mathbf{x}) = (m_1(\mathbf{x}), m_2(\mathbf{x}), \dots, m_{N_M}(\mathbf{x}))^T \quad (2)$$

of position  $\mathbf{x} \in \oplus$  inside the earth, where the real-valued vector components  $m_p(\mathbf{x})$  refer to material properties of the earth model. Observed and predicted data are considered as real numbers ( $d_i, s_i \in \mathbb{R}$ ). For any given elastic earth model  $\mathbf{m}(\mathbf{x}) = (C_{klqr}(\mathbf{x}), \rho(\mathbf{x}))^T$ , where  $\rho$  is density and  $C_{klqr}$  are elastic constants, both dependent on spatial position  $\mathbf{x}$ , the predictions  $s_i = F_i[\mathbf{m}]$  are obtained by solving the elastodynamic equation of motion, (cp. Aki and Richards, 1980, eq. 2.17), here given in the frequency domain:

$$-\omega^2 \rho(\mathbf{x}) u_k(\mathbf{x}, \omega) - \partial_\ell (C_{klqr}(\mathbf{x}) \partial_r u_q(\mathbf{x}, \omega)) = f_k(\mathbf{x}, \omega). \quad (3)$$

Spatial partial derivatives are denoted by  $\partial_\ell$  and Einstein summation rules apply. The  $k$ -th component of the seismic displacement field  $\mathbf{u}$  is denoted by  $u_k(\mathbf{x}, \omega)$  and the predictions  $s_i$  are obtained from some mathematical transformation of  $\mathbf{u}$  evaluated at the receiver positions  $\mathbf{r}$  and frequency  $\omega$ . Boundary conditions for  $\mathbf{u}$  are not dealt with here, since scattering from perturbations of boundaries and seismic discontinuities is not discussed in this work. The seismic source is described by the respective body force field  $\mathbf{f}(\mathbf{x}, \omega)$ . Variations of  $\mathbf{u}$  due to perturbations in the source term are not considered here.

Note that the functionals  $F_i[\mathbf{m}]$  are non-linear with respect to earth model material properties in spite of the fact that the elastodynamic equation is a linear partial differential equation and also linear in density and elastic constants. This fact is easily recognized by considering a different model  $2\mathbf{m}$  with twice the density and twice the elastic constants of the original one. For fixed body force and boundary conditions, the displacement field for this model is actually  $\mathbf{u}/2$  instead of  $2\mathbf{u}$  as expected for a linear functional. Hence, the forward problem is non-linear with respect to earth model material properties.

In order to solve the inverse problem, a measure of fitness is required quantifying whether some model  $\mathbf{m}$  satisfactorily

predicts the data within their errors. Such a measure of fit may be defined by the misfit functional

$$\chi^2[\mathbf{m}] = \sum_{i=1}^{N_D} \frac{1}{\sigma_i^2} (d_i - F_i[\mathbf{m}])^2 . \quad (4)$$

In order to find a model explaining the data well, however, it is not desirable to simply minimize  $\chi^2[\mathbf{m}]$ . In this way, uncertainties in the data due to measurement errors or noise would be explained by earth structure. Instead, e.g. a minimization of some squared model norm is sought subject to the condition that misfit  $\chi^2[\mathbf{m}]$  reaches an acceptable level. This problem may be formulated as minimizing the Lagrange functional

$$L[\mathbf{m}] = \|S\mathbf{m}\|_V^2 + \nu\chi^2[\mathbf{m}] , \quad (5)$$

where  $S$  is an operator acting on  $\mathbf{m}$  which may effect smoothing or damping, respectively,  $\|\cdot\|_V$  denotes the norm on model space  $V$  and  $\nu > 0$  is an additional Lagrange multiplier (Parker, 1994, eq. 3.02(26)). The Lagrange parameter  $\nu$  may be interpreted as a “trade-off” parameter, balancing two undesirable properties of a model  $\mathbf{m}$ , namely a large norm and a bad fit of the data.

A model  $\hat{\mathbf{m}}$  minimizing eq. (5) is found in practice by iteratively deriving a sequence of models  $\mathbf{m}^1, \mathbf{m}^2, \dots, \mathbf{m}^n, \dots$  converging to  $\hat{\mathbf{m}}$  which is initiated by a suitable starting model  $\mathbf{m}^0$ . The non-linear functional  $L$  is linearized in each iteration  $n \geq 1$  by linearizing the functionals  $F_i$  within a neighbourhood  $U(\mathbf{m}^{n-1})$  of the previously derived model  $\mathbf{m}^{n-1}$ :

$$F_i[\mathbf{m}^{n-1} + \delta\mathbf{m}^n] \simeq F_i[\mathbf{m}^{n-1}] + \int_{\oplus} \mathbf{k}_i^{n-1}(\mathbf{x}) \cdot \delta\mathbf{m}^n(\mathbf{x}) d^3\mathbf{x}, \quad \delta\mathbf{m}^n \in U(\mathbf{m}^{n-1}) \quad (6)$$

(cp. Parker, 1994, eqs. 2.03(1), 2.02(16)). The vector-valued function  $\mathbf{k}_i^{n-1} \in V$  is dependent on  $\mathbf{m}^{n-1}$  and is identical to the Fréchet derivative of the functional  $F_i$ , quantifying the change of prediction  $F_i[\mathbf{m}^{n-1}]$  due to small changes in the model space. For this reason, it is commonly termed *sensitivity kernel*. In case that the functionals  $F_i$  directly represent samples of the waveform, it is termed *waveform sensitivity kernel*. The model update  $\delta\mathbf{m}^n$  is then obtained by minimizing a linearization of the Lagrange functional  $L$  defined in eq. (5) using the linearization eq. (6) and replacing the norm term  $\|S\mathbf{m}^n\|_V^2$  by  $\|S\delta\mathbf{m}^n\|_V^2$ :

$$L_{\text{lin}}[\delta\mathbf{m}^n] = \|S\delta\mathbf{m}^n\|_V^2 + \nu \sum_{i=1}^{N_D} \frac{1}{\sigma_i^2} \left( d_i - F_i[\mathbf{m}^{n-1}] - \int_{\oplus} \mathbf{k}_i^{n-1}(\mathbf{x}) \cdot \delta\mathbf{m}^n(\mathbf{x}) d^3\mathbf{x} \right)^2 . \quad (7)$$

The update  $\delta\mathbf{m}^n$  then defines the new model after iteration  $n$  as

$$\mathbf{m}^n = \mathbf{m}^{n-1} + \delta\mathbf{m}^n . \quad (8)$$

In practice, eq. (7) is minimized by solving an overdetermined system of equations in a least-squares sense, after choosing some spatial discretization of the model domain. The system consists of an equation for each data sample  $i$  of the form

$$\frac{1}{\sigma_i} d_i - \frac{1}{\sigma_i} s_i^{n-1} = \frac{1}{\sigma_i} \int_{\oplus} \mathbf{k}_i^{n-1}(\mathbf{x}) \cdot \delta\mathbf{m}^n(\mathbf{x}) d^3\mathbf{x} , \quad (9)$$

where  $s_i^{n-1} = F_i[\mathbf{m}^{n-1}]$  denotes predictions of the data based on the preceding model  $\mathbf{m}^{n-1}$ . Additional regularization equations are added to this system which are based on relation

$$\nu^{-\frac{1}{2}} S\delta\mathbf{m}^n = \mathbf{0} . \quad (10)$$

The sensitivity kernels  $\mathbf{k}_i^n(\mathbf{x})$  can be derived using the Born scattering integral

$$\delta u_m(\mathbf{r}, \omega) = \int_{\oplus} \left[ \delta\rho(\mathbf{x}) \omega^2 u_k(\mathbf{x}, \omega) G_{km}(\mathbf{x}, \omega; \mathbf{r}) - \delta C_{k\ell qr}(\mathbf{x}) (\partial_r u_q(\mathbf{x}, \omega)) \partial_\ell G_{km}(\mathbf{x}, \omega; \mathbf{r}) \right] d^3\mathbf{x} , \quad (11)$$

which can be derived in a linearized sense via the Born approximation, additionally applying Betti’s theorem (Aki and Richards, 1980, eq. 2.34) and integration by parts, neglecting any perturbations of internal or outer boundaries (cp. Hudson and Heritage 1981, eq. 3; Chapman 2004, eq. 10.3.42; Schumacher 2014, eq. 3.9). Here,  $\delta u_m(\mathbf{r}, \omega)$  is a perturbation of the  $m$ -th displacement field component  $u_m$  at receiver position  $\mathbf{r}$  and frequency  $\omega$  caused by the model perturbation  $\delta\mathbf{m}^n = (\delta C_{k\ell qr}, \delta\rho)^T$ . The exciting seismic source is not indicated, for notational simplification.  $G_{km}(\mathbf{x}, \omega; \mathbf{r})$  is the elastodynamic Green function, i.e. the solution to eq. (3) for a body force situated at the receiver position  $\mathbf{r}$ ,  $f_k(\mathbf{x}, \omega) = \delta_{km} \delta(\mathbf{x} - \mathbf{r})$ , which is the Fourier transform of  $f_k(\mathbf{x}, t) = \delta_{km} \delta(\mathbf{x} - \mathbf{r}) \delta(t)$  (cp. Aki and Richards, 1980, eq. 2.36). Both wavefields,  $u_k(\mathbf{x}, \omega)$  and  $G_{km}(\mathbf{x}, \omega; \mathbf{r})$  are computed with model  $\mathbf{m}^{n-1}$  which constitutes the unperturbed background model in iteration  $n$ .

The connection of the scattering integral to the waveform sensitivity kernels becomes clear by observing that the displacement perturbation  $\delta u_m(\mathbf{r}, \omega)$  can be interpreted as the difference  $F_i[\mathbf{m}^n] - F_i[\mathbf{m}^{n-1}]$  between the values of the functional  $F_i$  for the perturbed and the unperturbed model, respectively, if the  $F_i$  represent spectral samples of seismic displacement. Then, noting (8), the integral in eq. (6) is identical to the integral in eq. (11) from which the kernels can be read off. If the  $F_i$  represent mathematical transformations of the displacement field, the corresponding kernels can be expressed as linear combinations of the waveform kernels, at least in a linearized sense.

Using eq. (11) allows to compute kernels for general anisotropic elastic earth models. In case of an isotropic elastic earth model  $\mathbf{m}(\mathbf{x}) = (v_p(\mathbf{x}), v_s(\mathbf{x}), \rho(\mathbf{x}))^T$  using seismic velocities and density, one can write eq. (11) as follows:

$$\delta u_m(\mathbf{r}) = \int_{\oplus} \left[ \delta v_p(\mathbf{x}) K_m^{v_p}(\mathbf{x}; \mathbf{r}) + \delta v_s(\mathbf{x}) K_m^{v_s}(\mathbf{x}; \mathbf{r}) + \delta\rho(\mathbf{x}) K_m^\rho(\mathbf{x}; \mathbf{r}) \right] d^3\mathbf{x} . \quad (12)$$

Here, we used a more standard notation for the kernels by setting  $\mathbf{k}_i = (K_m^{v_p}(\mathbf{x}; \mathbf{r}), K_m^{v_s}(\mathbf{x}; \mathbf{r}), K_m^\rho(\mathbf{x}; \mathbf{r}))^T$ . The data index  $i$  stands for the combination of receiver component  $m$ , receiver location  $\mathbf{r}$  and (not indicated) frequency  $\omega$  and seismic source. Explicit formulae of these kernels, as well as kernels for general anisotropic elastic earth models can be found in appendix A. To compute the kernels, one needs pre-calculated wavefield spectra  $u_k(\mathbf{x}, \omega)$  emanating from the source positions and pre-calculated Green function spectra  $G_{km}(\mathbf{x}, \omega; \mathbf{r})$  emanating from the receiver positions, as well as their strains.

All considerations also hold true for subvolumes  $\Omega \subset \oplus$  of the earth, provided there are no model perturbations outside  $\Omega$ . Figure 1 shows waveform kernels for shear wave velocity  $v_s$  and P-wave velocity  $v_p$  at a specific frequency, taken from a synthetic inversion example which is presented in section 4.1. Figure 2 shows a  $v_s$  waveform kernel taken from the spherical application which is presented in section 4.2.

### 3. A new conceptual approach to scattering-integral based FWI

In each iteration of scattering-integral-based FWI as described in section 2 three basic tasks have to be performed: solving the

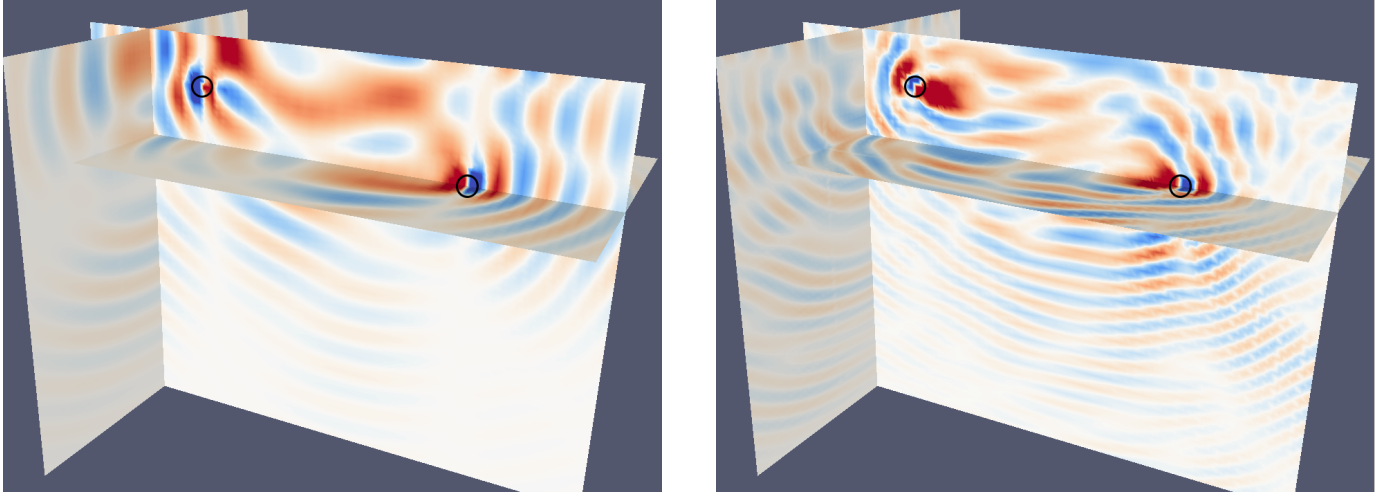


Figure 1: Slices of  $v_p$  sensitivity kernel (*left*) and  $v_s$  sensitivity kernel (*right*) through a medium with free surface at the top, buried seismic source on the left and buried receiver on the right (circles in each image). The kernels are taken for the “left-right” receiver component from the 12<sup>th</sup> iteration step of the cross borehole setting in section 4.1. The background model contains an embedded heterogeneity between source and receiver, as can be seen by the distortion of the kernel structure there. Note that sensitivities are in general highest at and in between source and receiver positions. For both kernels, the real part of the complex kernel is plotted at the same frequency of 76 Hz. The finer structure of  $K^{v_s}$  can be explained by smaller wavelengths of S-waves compared with P-waves. White colors represent zero sensitivity, blue and red colors signify positive and negative sensitivity, respectively. Scales are linear and normalized and slightly saturated.  $K^{v_s}$  has larger absolute values than  $K^{v_p}$  (by a factor of about 3), potentially due to the larger amplitudes of S-waves compared with P-waves.

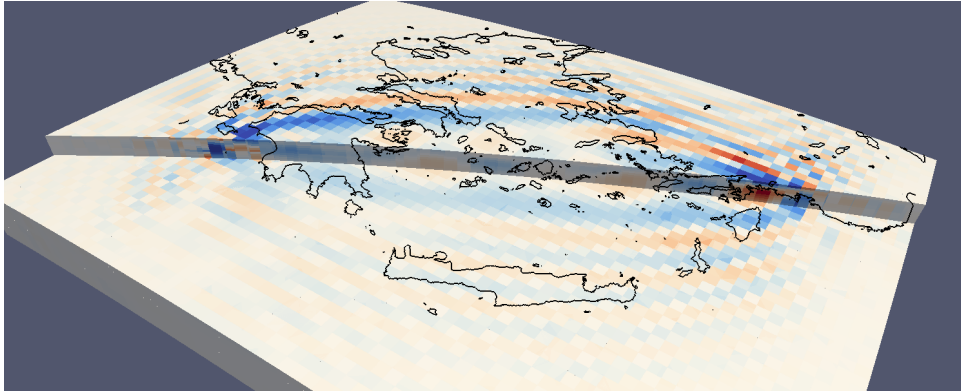


Figure 2: Real part of the pre-integrated  $v_s$  sensitivity kernel taken from the spherical application presented in section 4.2. The kernel for the vertical upwards component of a station in Turkey and an earthquake in the Peloponnese is shown. White colors represent zero sensitivity, blue and red colors signify positive and negative sensitivity, respectively. The color scale is linear and slightly saturated due to high sensitivities around source and receiver (about 1.9 times higher than in the rest of the medium).

forward problem (stage I), computing sensitivity kernels (stage II) and deriving a model update (stage III). A straightforward implementation of these tasks could look as follows:

Choose a forward solver suitable for the problem to be solved. It may work in spherical or Cartesian coordinates, in the frequency or time domain, and it may use finite-difference, finite-volume, finite-element, spectral, pseudo-spectral or discontinuous Galerkin methods to solve the elastodynamic equation. The solver will demand its own constraints on the spatial discretization of the wave propagation domain. Provide a suitable starting model sufficiently close to the real earth and use the solver to compute synthetic wavefields and strains for each seismic source throughout the inversion domain. To later compute sensitivity kernels using the Born scattering integral, also use this solver to compute Green functions and Green strains emanating from the receiver positions. Store these wavefields on disk.

In a second step, read these wavefields from disk and compute sensitivity kernels in frequency or time domain for each source-receiver pair from the corresponding wavefields associated with source and receiver position. Store these kernels on disk.

Finally, read in synthetic wavefields at the receiver positions, convert, if required, data into the frequency domain, form residuals and read in sensitivity kernels to set up the equation system eqs. (9), (10). Also choose regularization parameters and data weights. Solve the system of equation and derive a model update. Add it to the starting model and proceed with the next iteration.

In the following, we will discuss in detail the disadvantages of this straightforward approach and present our modifications to this procedure which make it flexible, extendable, modular as well as efficient with respect to both computation time and disk space. A graphical flow-chart of our inversion scheme is shown in fig. 3.

### 3.1. Introducing an independent discretization of model space

Our first and major criticism of the straightforward procedure sketched above is the fact that it does not work in real-world applications, because storage of sensitivity kernels is not feasible. The basic reason is that the spatial discretization needs of the forward solver are forced upon kernel computation and model update. The spatial discretization of the forward solver must be chosen according to criteria which ensure stable simulation of seismic wave propagation. They are dependent on the numerical method and apply to the entire wave propagation domain. On the other hand, model space discretization is in principle completely independent of these requirements and should rather be chosen in a problem-specific and data-oriented way. Typically, discretization for the forward solver will be much finer than is desirable for the model space. Since the kernels are defined on the model space, taking over the forward solver's discretization leads to massive unnecessary overhead in kernel computation and storage. In addition, it unnecessarily blows up the system of equations to be solved during model update. Hence, the most important change to the simplistic procedure is to introduce a model space discretization which is completely independent of the forward solver's discretization.

This step has consequences: We need an interface which allows us to compute kernels on the model space discretization from the output that is provided by the forward solver. Doing this requires additional work, but offers the opportunity to entirely decouple kernel computation from the specifications made by the forward solver. In fact, we may use any forward solver for kernel computation once we have written a suitable interface for it. In this way, we can benefit from the wide va-

riety of existing wave propagation codes in terms of numerical concepts and physical theory (e.g. able to deal with gravity, rotation, anelasticity, poroelasticity or anisotropy) even if we do not invert for all these physical properties occurring in the underlying theory. Moreover, we may switch among forward solvers between iterations. For example, if the initial model is 1D, we may want to use a fast 1D forward solver in the first iteration and switch to a 3D-solver later.

Currently, our software package ASKI supports the 1D semi-analytical code GEMINI (Friederich and Dalkolmo, 1995) and the 3D spectral element code SPECSEM3D (Tromp et al., 2008) in both, Cartesian and spherical framework. Support for the 3D nodal Discontinuous Galerkin code NEXD by Lambrecht (2015) is being implemented and tested. Support for further forward solvers may be easily added.

Thus, by introducing an independent model space discretization, the inversion code becomes flexible – many different forward solvers can be used –, extendable – new forward solvers can be added easily – and modular as forward problem solving and kernel computation are fully decoupled. The code is not tied to one single type of inverse problem solvable with the chosen forward method and its particular way of model discretization as it would be in the straightforward approach described above but is applicable to a diverse bunch of inverse problems using various geometries and underlying theories. Finally, since the model space discretization is typically much less dense than the discretization demanded by the forward solver, storage requirements for the kernels are massively reduced solving the kernel storage problem and making separation of the three steps practically feasible.

### 3.2. Working in the frequency domain

Numerical forward solvers working in the time domain commonly strongly oversample the synthetic wavefield in order to satisfy the Courant-Friedrichs-Levy (CFL) criterion (Courant et al., 1928). In contrast, the amount of independent informations on earth structure contained in a synthetic seismogram is typically by orders of magnitude smaller than the number of samples. Keeping the original time sampling when storing wavefields would imply a massive waste of disk space. Moreover, also observed seismograms are typically oversampled with respect to their information content, leading to waste of disk space as a kernel must be computed and stored for each individual time sample. In order to compress observed data and synthetic wavefields, we transform them into the frequency domain and store the resulting complex spectra only at the frequencies desired for the current iteration (e.g. Pratt, 1990). Computing sensitivity kernels from the synthetic wavefield spectra then also becomes cheaper, as the convolution operations required in the time domain reduce to multiplications of complex numbers. The downside is that time domain forward solvers need to be extended by a little routine that performs a Fourier transform of the synthetic wavefields on-the-fly during time stepping.

Working in the frequency domain also helps to reduce computation time and storage needs to the minimum required for the current iteration step. The reason is that for stability and convergence reasons, it is advantageous to start the inversion procedure with low frequency data in early iterations and to gradually increase the frequency content in later iterations (Bunks et al., 1995). Thus, for each individual iteration, the discretization of the forward solver may be adjusted to the desired maximum frequency allowing to decrease discretization density and time sampling frequency and thereby reducing computation time. The resulting wavefields are only stored for the selected discrete frequencies of an iteration. Moreover, model space discretization may also be adjusted to the desired maxi-

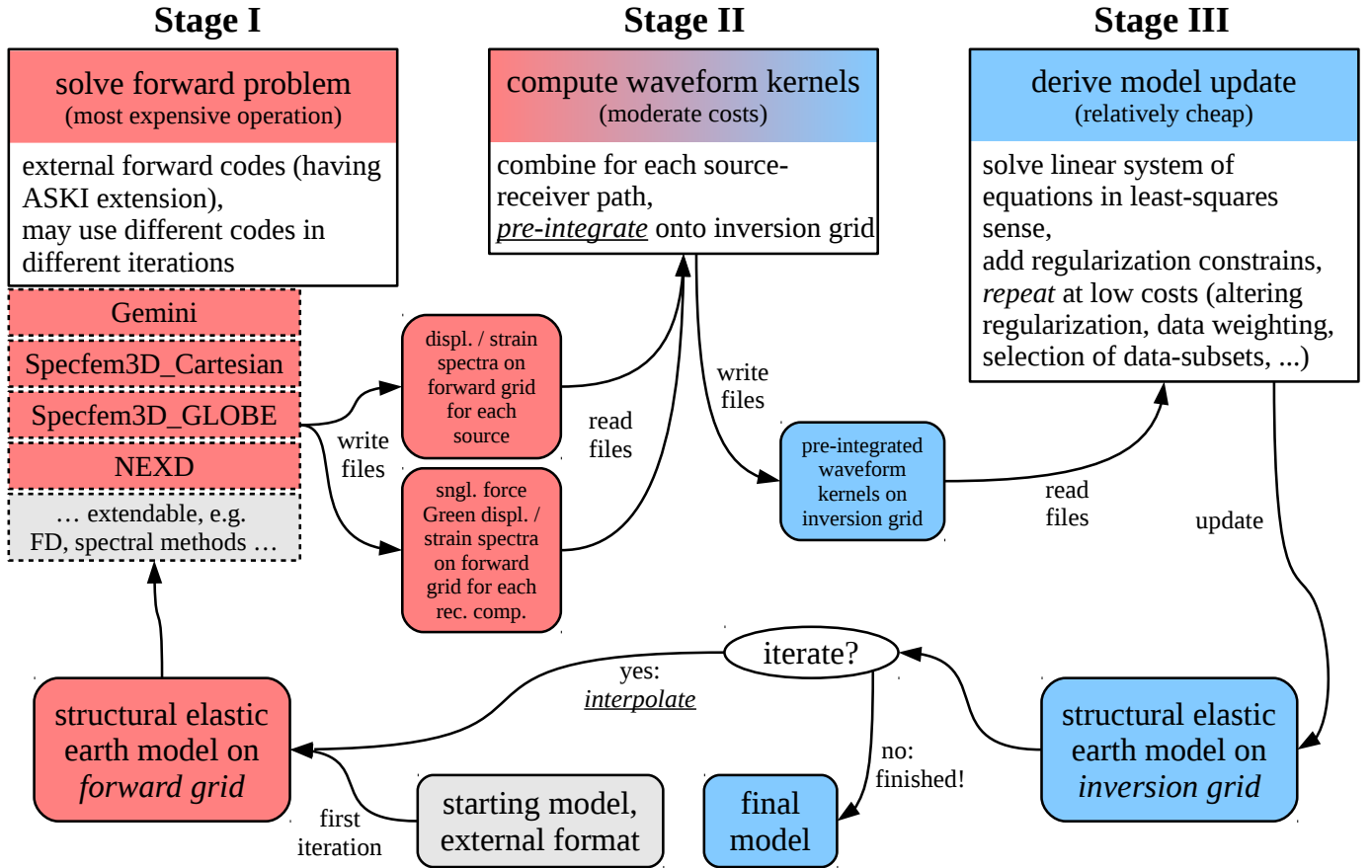


Figure 3: Modularized inversion scheme. *Red*: everything related to external forward solver and forward grid which is defined by the external solver. Note that the forward codes may need to be extended in such a way that they can produce the required waveform output. *Blue*: only dependent on internal software package ASKI, quantities on internal inversion grid (such as pre-integrated kernels or inverted models). *Rounded corners* indicate independent objects which are usually represented by one or more files. Boxes with *sharp corners* indicate operations which are conducted by (a set of) autonomous software programs. Note that the starting model might be given in yet another spatial description which must be transferred to the forward grid of the first iteration.

imum frequency allowing a reduction of model parameters and kernel disk space.

One drawback arising from scattering-integral-based FWI in the frequency domain is that it becomes challenging to invert data for particular seismic phases only. For time-domain forward methods, one could restrict the computation of the wavefield spectra to the time windows containing the phases. However, for each forward wavefield emanating from a seismic source and for each Green function emanating from a receiver position one would have to know the time windows containing these phases at each point inside the medium. This would result in additional computational costs, e.g. applying ray theory methods to determine travel times inside the medium, but more importantly would mean a huge logistical effort in the software which currently is not supported by ASKI and its forward solvers.

### 3.3. Pre-integration of sensitivity kernels

Computation of sensitivity kernels comprises a transition from the discretization of the forward solver to that of the model space. As the two are different and the latter is less dense than the former, a general pre-integration method is required which does not make a-priori assumptions on the specific discretizations used.

Model space discretization can be very generally realized by expressing the model as a linear combination of a set of spatial basis functions  $h_j(\mathbf{x})$  (e.g. Nolet, 2008, ch. 12). A specific model, e.g. comprising a set of material properties  $p = v_p, v_s, \rho$ , is then represented by coefficients  $a_j^p$  such that

$$m_p(\mathbf{x}) = \sum_j a_j^p h_j(\mathbf{x}). \quad (13)$$

The scattering integral in the form of eq. (12) then reduces to

$$\delta u_m(\mathbf{r}) = \sum_p \sum_j \delta a_j^p \int_{\oplus} h_j(\mathbf{x}) K_m^p(\mathbf{x}; \mathbf{r}) d^3 \mathbf{x}. \quad (14)$$

where the  $\delta a_j^p$  are perturbations of the model expansion coefficients. Thus, kernels appropriate for the expansion coefficients  $\delta a_j^p$ , receiver  $\mathbf{r}$  and displacement component  $m$  (as well as seismic source and frequency) are obtained by pre-integration,

$$\kappa_{mj}^p(\mathbf{r}) = \int_{\oplus} h_j(\mathbf{x}) K_m^p(\mathbf{x}; \mathbf{r}) d^3 \mathbf{x}, \quad (15)$$

and eq. (9) can now be written as

$$\frac{1}{\sigma_i} d_i - \frac{1}{\sigma_i} s_i^{n-1} = \frac{1}{\sigma_i} \sum_p \sum_j \kappa_{mj}^p(\mathbf{r}) \delta a_j^p, \quad (16)$$

where the data index  $i$  again represents the combination of receiver location  $\mathbf{r}$  and receiver component  $m$  (as well as seismic source and frequency). This formulation is very general and even admits different basis functions for different material properties  $p$ .

In our current implementation, we choose a discretization of the inversion domain  $\Omega$  into disjoint volumetric cells  $\Omega_j$ ,

$$\Omega \subseteq \oplus \text{ and } \Omega = \dot{\cup}_j \Omega_j, \quad (17)$$

which we will call *inversion grid*. For every cell  $\Omega_j$  we define the basis function  $h_j$  as its indicator function, i.e.

$$h_j(\mathbf{x}) = 1 \text{ if } \mathbf{x} \in \Omega_j \text{ and } h_j(\mathbf{x}) = 0 \text{ otherwise.} \quad (18)$$

This localizing definition comes closest to a nodal model description, since constant values on the inversion grid cells may

be associated with the cells' centers. As a result, expression (15) for the pre-integrated kernels simplifies to

$$\kappa_{mj}^p(\mathbf{r}) = \int_{\Omega_j} K_m^p(\mathbf{x}; \mathbf{r}) d^3 \mathbf{x} \quad (19)$$

and  $\delta a_j^p$  simply represents the average value of  $\delta m_p$  in cell  $\Omega_j$ . The spatial extent of a specific cell  $\Omega_j$  then represents the local model resolution.

To be able to perform the kernel pre-integration, we assume that the forward solver allows to calculate values of the required synthetic wavefields at a set of nodal points which we call the *forward grid*. Some forward solvers (FD, SPEC-FEM, nodal DG) provide an internal point grid which may be used as forward grid. For others, in particular modal and spectral approaches, such a grid must be defined by the user and the forward code must be adjusted to provide the required wavefields on this grid. Based on this information, the pre-integration of the kernels onto the inversion grid cells  $\Omega_j$  must be computed by numerical integration rules in practice. A very general definition of integration rules is required, as assumptions should be made neither on the type of the forward grid nor on the shape of the inversion grid cells which depend on the inversion problem. Such rules form the key in connecting forward and inverse problem.

Suitable integration rules were developed by Levin (1999), according to which we compute integration weights  $w_\ell$  such that

$$\int_{\Omega_j} K_m^p(\mathbf{x}; \mathbf{r}) d^3 \mathbf{x} \simeq \sum_{\substack{\ell, \text{ where} \\ \mathbf{x}_\ell \in \Omega_j}} w_\ell K_m^p(\mathbf{x}_\ell; \mathbf{r}), \quad (20)$$

where the  $\mathbf{x}_\ell$  stand for the nodes of the forward grid. Levin (1999) develops the integration rules from "a general method for near-best approximations to functionals using scattered-data information" (Levin, 1998), which in turn originates from the Backus-Gilbert theory (Backus and Gilbert, 1967, 1968, 1970; Bos and Salkauskas, 1989). The method of computing the integration weights  $w_\ell$  does not assume a specific distribution of the forward grid points  $\mathbf{x}_\ell$  inside the inversion grid cell  $\Omega_j$ . It aims at being stable in the sense that  $\sum_\ell |w_\ell|$  is as small as possible.

Technically, the integration weights are derived by determining a least-squares approximation of the available kernel values  $K_m^p(\mathbf{x}_\ell; \mathbf{r})$ ,  $\mathbf{x}_\ell \in \Omega_j$  by a polynomial of a given maximum degree  $L$  on a set of disjoint subvolumes  $E_k$  of  $\Omega_j$ . To do this, basis polynomials are chosen and their integrals over the subvolumes  $E_k$  evaluated. Integration weights  $w_\ell^k$  for each subvolume  $E_k$  are then calculated as linear combinations of the integrals of the polynomial basis functions over  $E_k$ . The associated expansion coefficients depend in a complicated way not detailed here on the values of the basis polynomials at the forward grid points. The final weights are obtained as  $w_\ell = \sum_k w_\ell^k$ . Our implementation currently supports computation of integration weights for any distorted tetrahedral or hexahedral shapes of cells  $\Omega_j$ .

The higher the polynomial degree is chosen, the more accurate the integration rules become. For the computation of the weights, few symmetric linear systems must be solved for each inversion grid cell  $\Omega_j$  which have a rather small number of equations  $N_e$  equal to the dimension of the polynomial space (e.g.  $N_e = 20$  for  $L = 3$ ). However, in order to assure solvability of these linear systems, the number of forward grid points  $\mathbf{x}_\ell \in \Omega_j$  must not be smaller than  $N_e$ . In practice, weight calculation tries to start with an initially chosen maximum polynomial degree (say 3) and systematically decreases the degree by one until the condition is fulfilled. In general, the absolute values of the integration weights are small in regions of high point density

and larger where the points  $\mathbf{x}_\ell$  are more sparsely distributed inside  $\Omega_j$ .

The software package ASKI provides different types of inversion grids for spherical and Cartesian applications and takes care of localizing the given forward grid points  $\mathbf{x}_\ell$  inside the inversion grid cells  $\Omega_j$  before computing the integration weights  $w_\ell$  for every cell  $\Omega_j$ , as described above. Starting from piecewise regular distributions of inversion grid cells, ASKI offers to subdivide particular inversion grid cells as often as required in order to achieve a locally refined inversion grid whose resolution is in accordance with the expected resolution power of the inverted dataset. The modular and extendable structure of ASKI enables users to add code for their own types of inversion grids.

### 3.4. Flexible data handling and playing with regularization during model update

Since, thanks to the combined effect of working in the frequency domain and introducing an independent discretization of model space, it becomes feasible to store sensitivity kernels on disk, the model update can be considered as a truly independent step of our inversion procedure only linked to the preceding one (kernel computation) by the requirement to read in the sensitivity kernels. This aspect offers new opportunities with regard to regularization and data weighting.

The model update is determined by solving the equation system (16) enlarged by regularization equations in a least-squares sense. Since in our discretization the model parameters  $\delta a_j^p$  are identical to average values of  $\delta m_p$  in inversion grid cell  $\Omega_j$  we denote them in the following by  $\delta m_{pj}^n$ , where the superscript  $n$  indicates the current iteration. Since the values  $\delta m_{pj}^n$  must be real-valued, we formulate each complex-valued equation of (16) by two real-valued equations taking its real and imaginary part, respectively.

We solve the system either by a serial or parallel QR-factorization method using LAPACK or ScaLAPACK libraries (Anderson et al., 1999; Blackford et al., 1997), or by a parallelized conjugate-gradient method (Björck et al., 1998, algorithm 3.1). Yet other mathematical approaches to solve the system could be added to the ASKI software package. In order to compute new absolute model values  $m_{pj}^n$ , first the background model values  $m_p^{n-1}(\mathbf{x}_\ell)$  given on the forward grid must be interpolated onto the inversion grid. This is done using the integration weights  $w_\ell$  introduced above:

$$m_{pj}^{n-1} = \sum_{\substack{\ell, \text{ where} \\ \mathbf{x}_\ell \in \Omega_j}} \frac{w_\ell}{V(\Omega_j)} m_p^{n-1}(\mathbf{x}_\ell),$$

$$\text{where } V(\Omega_j) = \sum_{\substack{\ell, \text{ where} \\ \mathbf{x}_\ell \in \Omega_j}} w_\ell \quad (21)$$

Thereafter, values  $m_{pj}^n = m_{pj}^{n-1} + \delta m_{pj}^n$  can be computed, yielding new values after iteration  $n$  for each material property  $p$  on the inversion grid. The model update code is written in a generic way. It only knows about indices of data and indices of model parameters. In this way, it is applicable to any choice of model space discretization.

The computational expenses of solving the linear system (stage III) are negligible compared with the solution of the forward problem (stage I) and the computation of sensitivity kernels (stage II). Hence, stage III of each iteration can be repeated at relatively low costs without need to solve the forward problem again. This fact allows playing with varying intensities of smoothing and damping or even choosing different data subsets (i.e. some subset of the rows of the equation system (16)). Moreover, it offers the opportunity of discarding or

down-weighting certain data values (e.g. which have the largest remaining residuals or have a large influence on the inverted model due to event clustering) and compare different models before choosing the one for the next iteration of FWI. For the adjoint-wavefield method, by contrast, such a specific weighting of particular data samples is not possible at the stage of deriving a model update, since for each source a combined misfit kernel is computed comprising all data related to this source. Any weighting would have to be applied already at the stage of computing the misfit kernels.

Even changes of choice of the misfit functional during iterations are possible. For example, we may want to explain the spectral phases only instead of the complex spectral amplitudes. Then, the  $d_i$  would represent these phases. Their corresponding sensitivity kernels can be conveniently computed by a linear combination of the available waveform sensitivity kernels thus requiring only minor changes in the computations of stage III.

The ability to repeat the model update with different smoothing and damping makes our inversion procedure suitable for Occam's method (Constable et al., 1987; Aster et al., 2013, ch. 10.2). This regularization method for non-linear inverse problems involving erroneous data aims at finding a model of minimal norm while the misfit reaches a certain predefined level. In each iteration, the effect of varying smoothing and damping intensities on the non-linear misfit are observed. The ones associated with the lowest misfit are used to construct the starting model for the next iteration. While our inversion procedure greatly alleviates to obtain model updates for varying smoothing and damping, it should be kept in mind that additional expensive forward computations are required to evaluate the non-linear misfit. Applying Occam's method, thus adds additional computational costs to each iteration.

### 3.5. Connecting iterations: unstructured 3D interpolation from inversion grid to forward grid

The separation of stages III and I furthermore demands to transfer the inverted model values  $m_{pj}^n$  derived on the inversion grid of iteration  $n$  to the model description as required by the forward code in the next iteration  $n + 1$ . For mesh-based nodal forward codes, doing this amounts to an interpolation capable to connect arbitrary 3D-unstructured grids as no general assumptions should be made on the structure of forward and inversion grid. For forward solvers SPEC-FEM3D (both, GLOBE and Cartesian) and NEXD, we apply a suitable interpolation method based on Shepard (1968) which is founded on inverse-distance weighting and accounts for issues of nearby points, direction and slope. Other interpolation methods based on Shepard (1968) or modifications of it may be found in the literature (e.g. Franke and Nielson, 1980; Łukaszuk, 2004; Masjukov and Masjukov, 2005).

## 4. Applications of seismic full waveform inversion by software package ASKI

The full waveform inversion procedure described above was implemented in the software package ASKI – Analysis of Sensitivity and Kernel Inversion (Schumacher, 2015). It is implemented in the fashion of an object-oriented software library written in modern FORTRAN and Python and stands in contrast to the monolithic, unflexible and hard-to-extend codes often written for solving inverse problems. Object orientation is motivated by the modular and extendable character of our new approach to SI-based FWI. ASKI consists of numerous autonomous software programs and scripts which communicate via file interfaces and follow the modularized inversion



procedure. ASKI is well documented and freely available at <http://www.rub.de/aski> under the terms of the GNU general public license (<http://www.gnu.org>).

The use of external forward modelling codes is technically realized by specific routines in ASKI that serve as interfaces to the forward code. File formats used to communicate wavefield data, model values, point geometries and auxiliary information from the forward code to ASKI can be defined by the forward code itself. For example, structured grids and simple earth models can be communicated by very few metadata, whereas unstructured grids and 3D models usually need to be communicated by explicit point-wise information. Thereby, parameter and model files used with the forward code can also be used with ASKI, provided the knowledge of how to read and interpret these files is implemented in the specific ASKI interface routines dedicated to that forward code. This concept allows ASKI to extend to very different kinds of forward methods in a very flexible way.

By its implementation, ASKI can be easily extended to any kind of sensitivity analysis tools operating on the sensitivity matrix. New code that extends the ASKI software package will be based on the existing software modules and, thus, will benefit from all the existing functionality of ASKI. For instance, new serial or parallelized equation solvers or data processing and visualization tools can be added and used easily. Due to the object-oriented character of ASKI, also the software library itself can be easily extended to further functionality, e.g. supporting other kinds of inversion grids or implementing different methods for kernel pre-integration.

#### 4.1. Synthetic validation of ASKI

In order to validate our new approach to scattering-integral-based FWI, as well as its implementation, we invert noise-free synthetic datasets for known model anomalies, which we present in the following. A near-surface scale Cartesian setting is chosen, using the spectral element code SPECSEM3D Cartesian 2.1 as the forward solver for ASKI. The true elastic earth model is a homogeneous isotropic half space with a free surface into which a block of mixed slow and fast velocity anomalies is placed. Density is not perturbed and is not inverted for. For comparison, we try to resolve the anomaly using twelve single force sources and twelve three-component receivers in different acquisition layouts: cross borehole (sources and receivers in boreholes with the anomaly in between), VSP – vertical seismic profile (sources on the free surface above the anomaly, receivers in a borehole through the anomaly) and shallow surface wave inversion (both, sources and receivers on the free surface above the anomaly). These three layouts, as well as the velocity anomaly, are shown in figure 4 (*top*). The cross borehole inversion is presented in detail in Schumacher (2014, ch. 5.2).

The three datasets to be inverted are synthetically computed for the true earth model with SPECSEM3D using for each shot a record length of 0.25 s and a Ricker wavelet source time function with center frequency of  $f_c = 50$  Hz. All sources excite a force of 1 kN and are always oriented in direction towards the anomaly in each acquisition layout. As a frequency discretization of the datasets, we choose the 21 discrete frequencies 36, 38, 40, ..., 76 Hz, which constitute the main frequency content. The chosen frequency step of  $2 \text{ Hz} = \frac{1}{2} \frac{1}{0.25 \text{ s}}$  is the maximum possible step for which a stable inversion of time series of length 0.25 s can be expected in theory (cf. Schumacher, 2014, ch. 3.3.2). However, we observed that the cross-borehole inversion yields the very same result when using a frequency step of  $4 \text{ Hz} = \frac{1}{0.25 \text{ s}}$ , i.e. only using the 11 discrete frequencies 36, 40, 44, ..., 76 Hz. This finding indicates that the frequency domain data also contain redundant information.

Although for a real-data application one would probably choose different inversion strategies, the full waveform inversion procedures for the three different datasets are performed in a very similar fashion in order to assure comparability of the inversion results. Inverting jointly for updates of P-wave and S-wave velocities, 12 iterations are conducted using the homogeneous halfspace as a starting model. Since the synthetic data is free of noise, we do not account for data uncertainties. Some specifications of the iterations identical for all three inversions are listed in table 1. The subset of the data inverted in each iteration varies by frequency content only, i.e. both real and imaginary part of the displacement spectra from all sources and three components of all receivers are used. An exception to this is made for the surface wave setting, where four source-receiver pairs with identical source and receiver position are omitted (in the corners of the acquisition array), resulting in 140 instead of 144 source-receiver pairs. For reasons of computational and storage efficiency, the frequency windows do not contain all previous frequencies, but a moving window of five discrete frequencies is used in an overlapping fashion. The resolution of both, the forward grid and the inversion grid are gradually increased in each iteration step according to the frequency content of the respective inverted data subset (note the increasing numbers  $N_{\text{forw}}^{\text{grid}}$  and  $N_{\text{inv}}^{\text{grid}}$  in table 1). We choose hexahedral inversion grid cells with approximately cubic shape. The general size of the inversion grid cells in each iteration was determined through heuristic testing of small variations in cell size and comparing the resulting model updates.

As a time domain forward solver, we added code to SPECSEM3D that performs a Fourier transform of the synthetic wavefields on-the-fly during time stepping. With Fourier transform, the computation times for solving the forward problem in the waveform inversions presented here are approx. 47% higher than they are for regular SPECSEM3D simulations without Fourier transform. These additional costs depend on the number of frequencies at which the Fourier transform is performed. They may also be dependent on the number of CPU cores, when choosing a simulation domain larger than the actual inversion domain (as we do, see appendix B). In this case, the perfect load balancing of SPECSEM3D is disturbed by our implementation of Fourier transform. Since we apply an explicit discrete summation of the Fourier integral, performance can be enhanced by applying for instance Goertzel's algorithm or variations of it (e.g. Goertzel, 1958; Sorensen et al., 1988), which significantly reduce the required number of floating point operations. Furthermore, computation times for solving the forward problem can be reduced by a factor of about 2.5 by using the latest SPECSEM3D Cartesian version 3.0, since a time step 2.5-times larger compared with version 2.1 can be chosen for stable simulations. Also note, that for experimental reasons the forward grid of the first iterations was resolved a little higher than actually necessary. Therefore, computation times and storage might additionally be reduced in the first iteration steps.

For regularization we apply smoothing conditions *only*, no damping is used. We do not apply Occam's method but, knowing the original model, rather try to choose suitable smoothing intensities manually. We apply Laplacian smoothing, requesting that the value of the model perturbation in each cell be equal to its mean taken over directly neighbouring cells. The smoothing equations are scaled separately for each material property  $p = v_p, v_s$  by the maximum of the absolute  $p$ -kernel values. This is done in order to account for different possible orders of magnitude of kernel values for the different model parameters. Thereafter, all smoothing equations are multiplied with a constant smoothing intensity factor which is chosen anew in every iteration step in order to control the smoothness of the

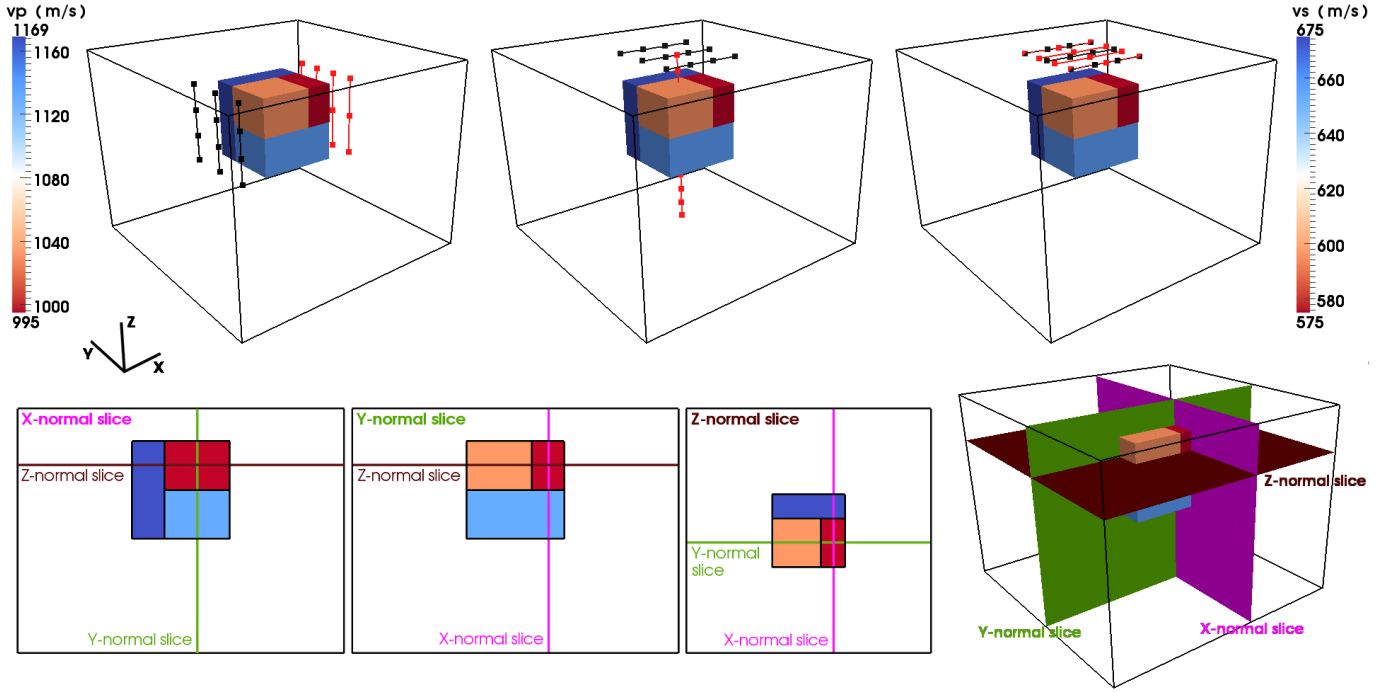


Figure 4: *Top*: Acquisition layouts for cross borehole (*top left*), VSP – vertical seismic profile (*top center*) and surface wave inversion (*top right*). Positions of single force sources are plotted in black and are always oriented along the coordinate axes in direction towards the anomaly. Receiver positions are plotted in red, their 3 orthogonal components are in X, Y, and Z-direction. Also shown is the velocity anomaly, placed into a homogeneous isotropic elastic halfspace of constant material properties  $\rho = 1800 \frac{\text{kg}}{\text{m}^3}$ ,  $v_p = 1082 \frac{\text{m}}{\text{s}}$  and  $v_s = 625 \frac{\text{m}}{\text{s}}$  (white colors). Colors refer to both,  $v_p$  and  $v_s$  (velocity values according to the respective colorbars). The anomalous cube has an edge length of 30m, is placed 10m below the free surface and consists of  $\pm 5\%$  and  $\pm 8\%$  velocity perturbations. The inversion domain is indicated by the surrounding box which is  $100 \text{ m} \times 100 \text{ m}$  wide and 75m deep.

*Bottom*: Definition of the slices which are shown in figures 5, 6, B1. Slices here show the true anomaly model. Coloured lines in the three *bottom-left* panels indicate intersections with other slices.

resulting model update. More details on the specifications of simulation grid, inversion grid and smoothing in each iteration are given in appendix B.

#### 4.1.1. Inversion results and fit of data

As expected from the acquisition geometry, the surface wave setting represents the least favorable inverse problem and indeed the inversion results show that only some of the  $v_s$  anomaly is resolved laterally in the upper part of the model. Since the data contain mostly surface waves, there is not much for the inversion to work on. For this reason we present detailed inversion results in appendix B for interested readers and in the following only discuss in detail the results of the cross borehole and VSP settings.

Figures 5–6 show the inverted  $v_p$  (*left*) and  $v_s$  (*right*) models after the first (*top*), the sixth (*center*) and the last (*bottom*) iteration. The images show slices through the inversion domain as defined by figure 4 (*bottom*). Figures 7–8 demonstrate the fitting of the waveforms showing exemplary seismograms as well as the normalized misfit  $\bar{\chi}^2$  of the frequency-domain data defined by

$$\bar{\chi}^2 = \frac{\sum_{i \in S} |d_i - s_i|^2}{\sum_{i \in S} |d_i|^2}, \quad (22)$$

where  $S \subseteq \{1, \dots, N_D\}$  represents some data subset. As data subsets for the plots below, different discrete frequencies are chosen as well as the complete frequency range 36, 38, 40,  $\dots$ , 76 Hz (thick blue line in figs. 7–8).

It can be clearly seen that in the cross borehole setting the velocity anomalies are resolved very well by the full waveform inversion in spite of some of the anomalous structure being smaller than the wavelengths at the dominating frequency of 50 Hz, e.g. the  $-8\%$  P-velocity perturbation of size  $20 \text{ m} \times 10 \text{ m} \times 15 \text{ m}$  (dominating wavelength 21.6 m). Nevertheless, the P-velocity model exhibits some artefacts which do not occur in the S-velocity model. Exemplary seismograms in figure 7 show that the waveforms are fitted equally well. Signal is produced correctly at times where there was no signal before, phases are fitted as well as amplitudes and even completely distorted wavelets are reconstructed well. A misfit reduction of about 95.4% demonstrates that most of the information contained in the data is extracted.

In the VSP setting the velocity anomalies are not as well reproduced as in the crosshole experiment. This finding is not surprising as the VSP experiment illuminates the anomaly less well than the crosshole experiment does. By comparison, the VSP model is smoother. This is due to the approach of choosing the smoothing intensities in such a way, that the model updates stay in value ranges which we found to be reasonably stable in the cross borehole inversion. In addition, we tried to avoid artefacts outside the anomaly block. This forced us to apply rather strong smoothing, indicating that for better convergence additionally damping conditions should have been applied in order to stabilize the inversion process. Figure 8 shows exemplary seismograms for the VSP setting. It can be observed, that the inversion starts with an overall smaller waveform misfit, compared with the cross borehole example. This is due to the acquisition and the fact that most of the ray paths cross only parts of the anomaly. Nevertheless it can be clearly seen that this geometry is capable to resolve the anomaly, which is also indicated by the misfit reduction of about 86.5%. For inversion methods which require preconditioning, the receivers inside the anomaly would be counterproductive for inversion.

#### 4.1.2. Discussion and implications

Approaching each inversion example by an individual strategy of choosing frequencies, inversion grid resolutions and starting model could help to improve the inversion results. For the sake of comparability, however, the VSP and surface wave inversion were conducted in a similar manner as the cross borehole inversion. This makes the VSP and surface wave inversion even more demanding. In the context of what can be expected from a theoretical point of view, all inversion examples yield very satisfying results, which in principle validates the applied inversion method and its implementation by software package ASKI.

Note that in iteration 3 we invert data with the same frequency content (36Hz, 38Hz, 40Hz) as in iteration 2 (cf. table 1). Also the inversion grid is only slightly refined in iteration 3. In the cross borehole and the VSP inversion, we observe that the model does not significantly change in the third iteration (cp. horizontal misfit curves between iterations 2 and 3 in figs. 7 and 8, *bottom right*). Therefore, it is reasonable to assume that the complete information content of the data at frequencies 36Hz–40Hz is already accounted for in iteration 2. For the cross borehole and the VSP inversion, this result supports the expectation that the frequency windows of the particular iterations are chosen appropriately in order to account for the complete information content of the data.

The inversion results might be improved and stabilized by additionally applying damping constraints and a complete decoupling of regularization constraints for different model parameters. Even different inversion grid resolutions for  $v_p$  and  $v_s$  could be used, dependent on the inverse problem. This could help to stabilize the inversions in accord with the actual resolving power of the dataset. In principle, doing this is possible due to the modularized implementation by ASKI, but would be more complicated to handle in practice and could be computationally more demanding, especially when applying Occam’s method for a larger number of regularization parameters.

When inverting noisy data in a real-data application, we can expect the resolving power to be smaller than in the noise-free synthetic inversion examples. Data uncertainties increase the influence of the regularization conditions since the data equations are down-weighted. Additional independent data are needed to mitigate this effect.

#### 4.2. First application of ASKI to real data in spherical geometry

An example of an application of our FWI approach and of ASKI to real data in spherical geometry is given by Lamara (2015). He studies the S-wave velocity structure in the Hellenic subduction zone down to a depth of 100 km inverting a high-quality dataset collected with a very dense and homogeneous amphibian seismographic network deployed within the framework of the EGEADOS project (Friederich and Meier, 2008). We do not give a full account of this work here but instead concentrate on demonstrating how the flexibility of ASKI was harnessed to tackle such a problem.

##### 4.2.1. Path-specific approach to ASKI FWI

Due to the complex geology in the Aegean and the rather high frequencies of the inverted data, available 1D models constructed for the Aegean proved to be inadequate for starting a FWI. Since a 3D initial model was not available, a path-specific approach was adopted following Friederich (2003) to solve the inverse problem. In the *first* iteration of the 3D FWI, multiple path-specific starting models are used instead of a 3D global initial model which properly reflect the average structural prop-

$n$	inverted data subset		forward problem				pre-integrated kernels				inversion step		
	$f$ [Hz]	$N_D^n$	$N_{\text{forw}}^{\text{grid}}$	$\Delta x_{\text{forw}}$	$S_{\text{forw}}$	$T_{\text{forw}}^{48c}$	$N_{\text{inv}}^{\text{grid}}$	$\Delta c_{\text{inv}}$	$S_{\text{kern}}$	$T_{\text{kern}}^{1c}$	$N_{\text{row}}$	$N_{\text{col}}$	$T_{\text{inv}}^{1c}$
1	36	864	59168	1.75	260.01	41.6	384	12.5	2.53	< 0.5	1632	768	< 2
2	36,38,40	2592	77315	1.625	849.41	59.2	567	10.98	11.21	$\vdots$	3726	1134	$\vdots$
3	36,38,40	2592	77315	1.625	849.41	61.6	700	10.24	13.84	$\vdots$	3992	1400	$\vdots$
4	36,38,...,44	4320	91238	1.5	1603.80	72.0	968	9.19	31.90	0.5	6256	1936	$\vdots$
5	40,42,...,48	4320	91238	1.5	1603.80	78.4	1296	8.33	42.71	0.5	6912	2592	$\vdots$
6	44,46,...,52	4320	162000	1.25	2847.66	124.8	1960	7.26	64.59	0.9	8240	3920	2
7	48,50,...,56	4320	162000	1.25	2847.66	131.2	2475	6.72	81.57	0.9	9270	4950	$\vdots$
8	52,54,...,60	4320	217800	1.125	3828.52	174.4	3072	6.25	101.25	1.25	10464	6144	$\vdots$
9	56,58,...,64	4320	217800	1.125	3828.52	185.3	3757	5.84	123.82	1.25	11834	7514	6
10	60,62,...,68	4320	329232	1.0	5787.29	283.2	4536	5.49	149.50	2	13392	9072	8
11	64,66,...,72	4320	329232	1.0	5787.29	286.4	5054	5.29	166.57	2	14428	10108	11
12	68,70,...,76	4320	329232	1.0	5787.29	354.4	6000	5.0	197.75	2	16320	12000	14

Table 1: Some statistics about the iteration steps of the three waveform inversions conducted in section 4.1:  $n$  = iteration step;  $f$  = frequencies [Hz] occurring in data subset;  $N_D^n$  = number of data samples in data subset;  $N_{\text{forw}}^{\text{grid}}$  = number of points in forward grid at which the wavefields are stored;  $\Delta x_{\text{forw}}$  = mean distance [m] between these points;  $S_{\text{forw}}$  = total amount of storage [MiB] of forward wavefields;  $T_{\text{forw}}^{48c}$  = total runtime [min] to solve forward problem (parallel on 48 CPU cores);  $N_{\text{inv}}^{\text{grid}}$  = number of inversion grid cells;  $\Delta c_{\text{inv}}$  = mean edge length [m] of inversion grid cells;  $S_{\text{kern}}$  = total amount of storage [MiB] of pre-integrated kernels;  $T_{\text{kern}}^{1c}$  = approximate runtime [min] of kernel computation (serial on 1 CPU core);  $N_{\text{row}}$  = number of rows of the linear system (including smoothing conditions, but no damping conditions here);  $N_{\text{col}}$  = number of columns of the linear system;  $T_{\text{inv}}^{1c}$  = approximate runtime [min] of LSE solver (serial QR-factorization method on 1 CPU core using LAPACK).

Note that  $N_{\text{col}} = 2 N_{\text{inv}}^{\text{grid}}$  is the number of unknowns ( $v_p$  and  $v_s$  value at each inversion grid cell) and  $N_{\text{row}} = N_D^n + 2 N_{\text{inv}}^{\text{grid}}$  ( $N_D^n$  data equations plus one smoothing equation for each of the  $2 N_{\text{inv}}^{\text{grid}}$  unknowns, no damping equations applied here).

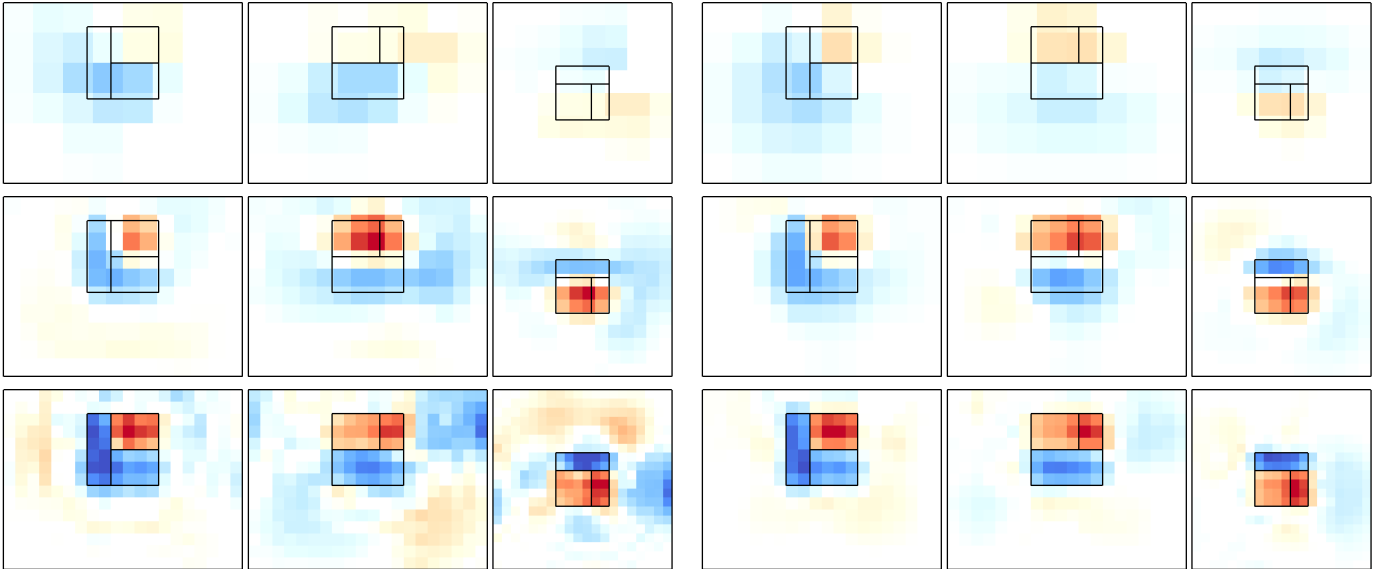


Figure 5: Inverted model from cross borehole inversion. P-wave speed (*left*), S-wave speed (*right*); colors as defined by colorbars in fig. 4. From *top* to *bottom*: model after iteration 1, iteration 6, iteration 12; slices as shown in fig. 4. The velocity anomalies are resolved very well, especially in the  $v_s$  model. Due to the larger wavelengths of the P-waves and their smaller resolving power, some artefacts are produced in the  $v_p$  model. This might be better controlled by completely independent smoothing intensities or even different inversion grid resolutions for  $v_p$  and  $v_s$ . Note the refinement of the inversion grid throughout the iterations.

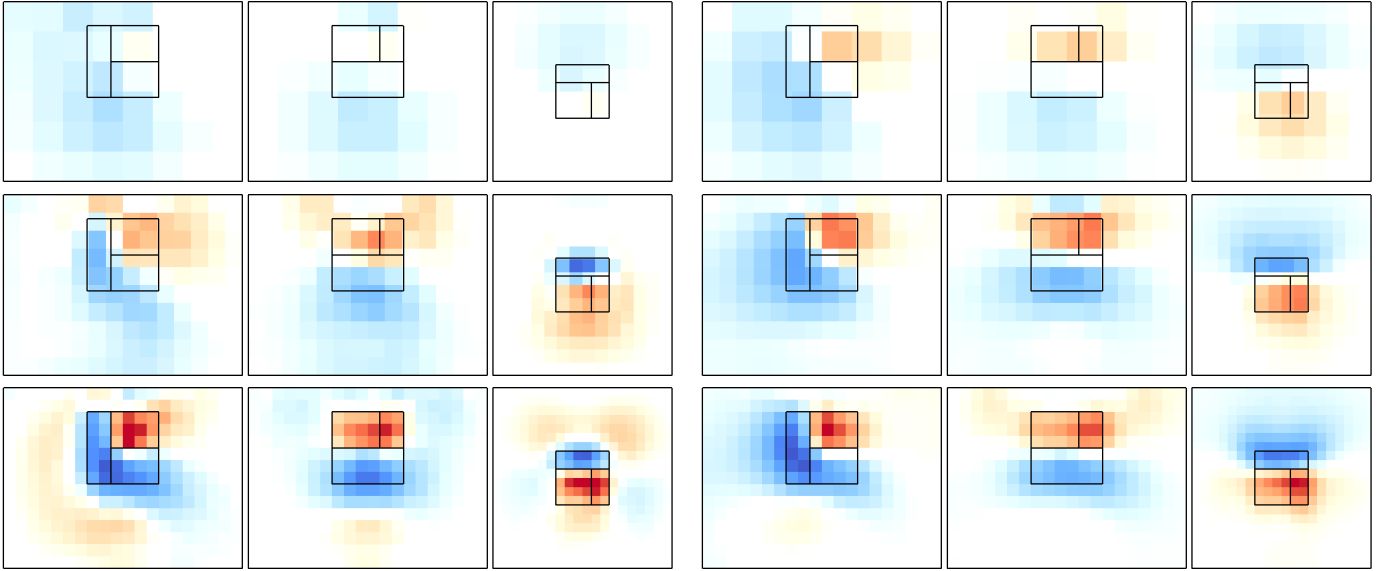


Figure 6: Inverted model from VSP inversion. P-wave speed (*left*), S-wave speed (*right*); colors as defined by colorbars in fig. 4. From *top* to *bottom*: model after iteration 1, iteration 6, iteration 12; slices as shown in fig. 4. The anomalies are generally resolved, however not as good as in the cross borehole setting. The smeared and smooth reconstruction results from a stronger smoothing which had to be applied in order to avoid stronger artefacts in the surrounding of the anomaly. Additional damping conditions could be suitable to better stabilize the inversion. Note the refinement of the inversion grid throughout the iterations.

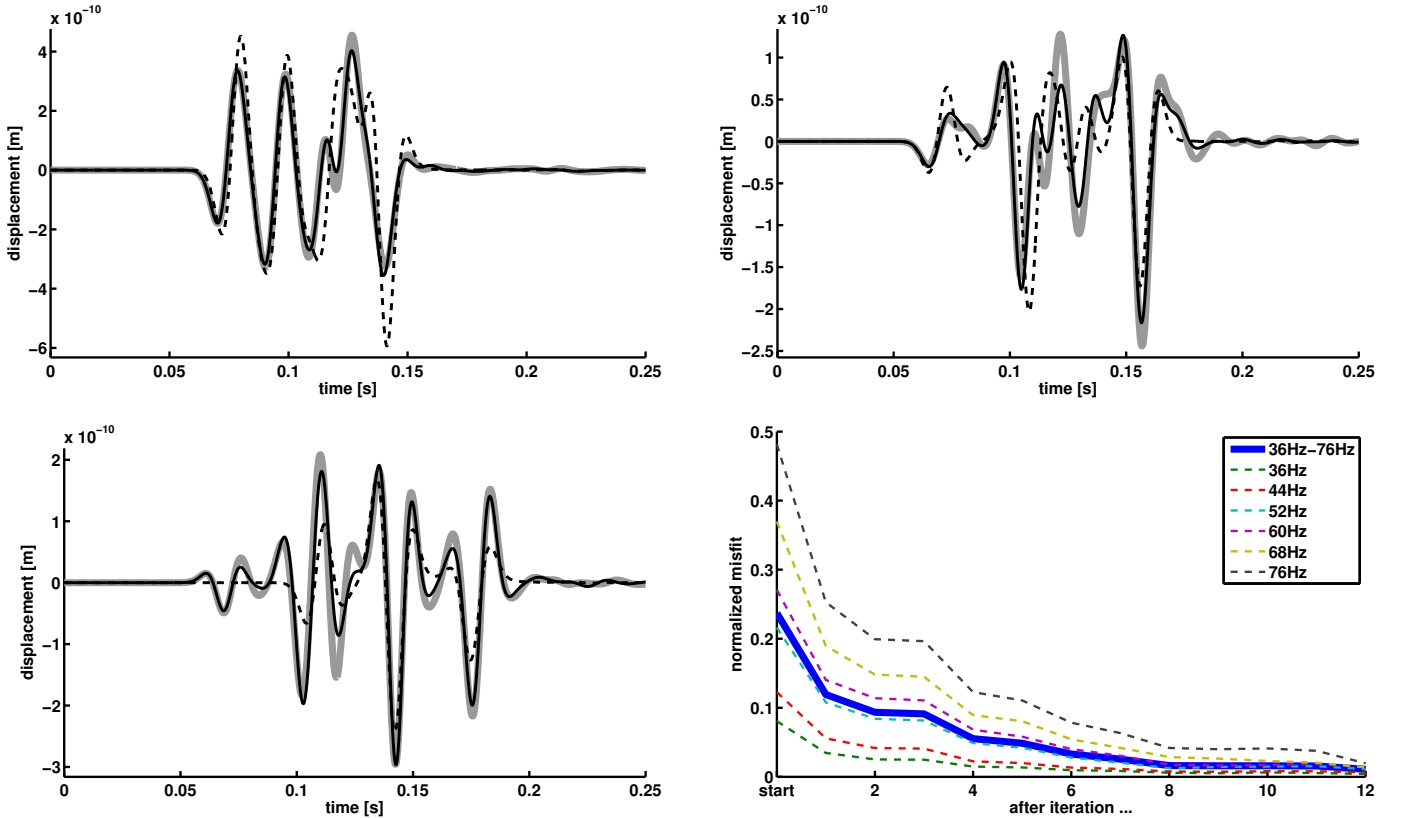


Figure 7: Three exemplary seismogram plots from cross borehole inversion, comparing data w.r.t. true earth model (thick grey line) with synthetic waveforms for the homogeneous starting model (dashed black line) and the waveform after the last iteration (solid black line). The *bottom-right* diagram shows the development of the normalized misfit (22) for different data subsets which vary with frequency content only. The waveforms are fitted very well and the total misfit reduction for the complete dataset (thick blue line) is 95.4 %.

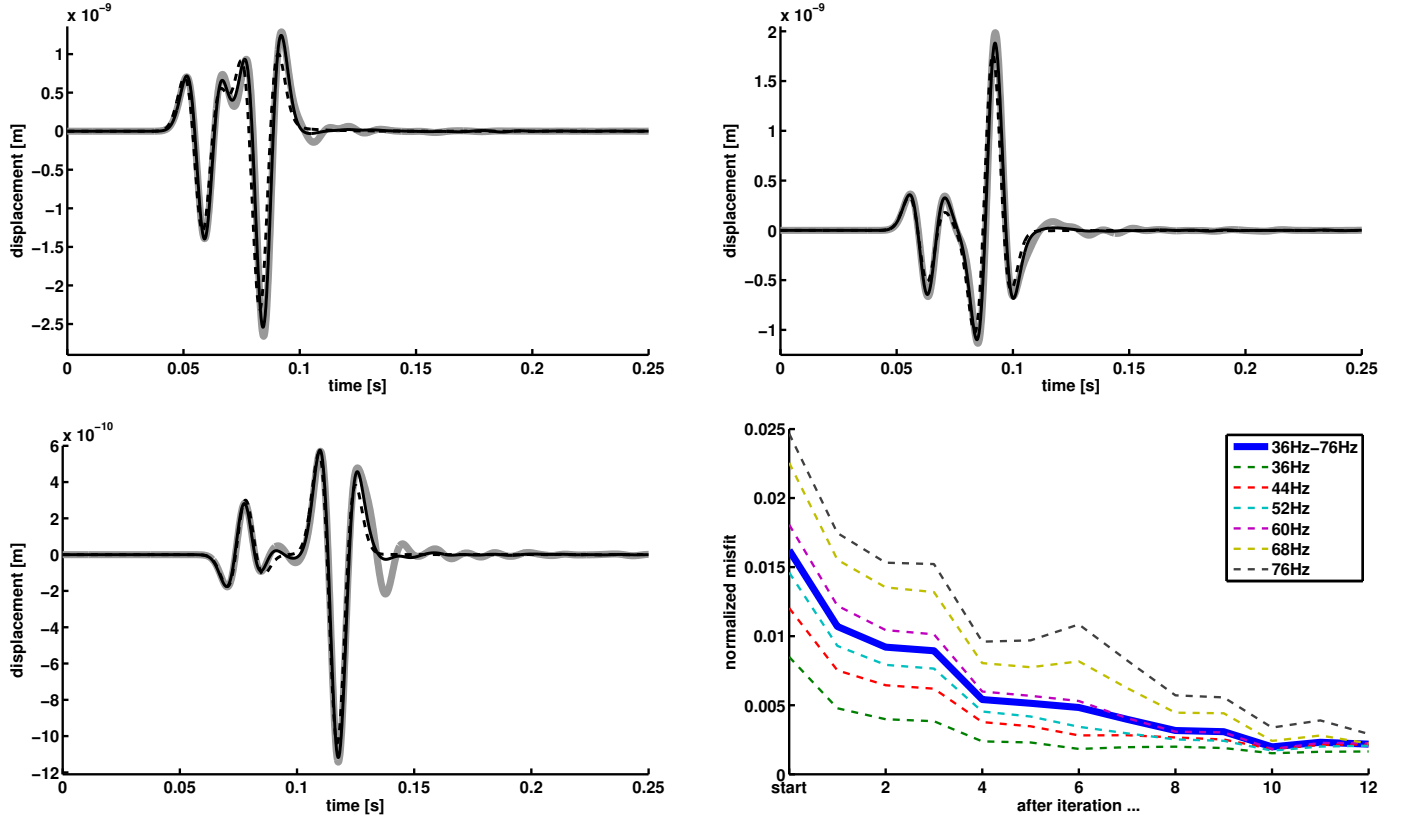


Figure 8: Three exemplary seismogram plots from VSP inversion, comparing data w.r.t. true earth model (thick grey line) with synthetic waveforms for the homogeneous starting model (dashed black line) and the waveform after the last iteration (solid black line). The *bottom-right* diagram shows the development of the normalized misfit (22) for different data subsets which vary with frequency content only. The VSP inversion starts with an overall smaller waveform misfit, compared with the cross borehole example, due to most of the wave paths crossing only parts of the anomaly. Nevertheless, the waveform misfit is significantly reduced by about 86.5 %, which is also indicated by the seismogram plots.

erties of each event-station path and help to ensure the applicability and convergence of the linearized inversion scheme. Unfortunately, the path-specific approach requires wavefield computations for each source-receiver pair. That is, denoting the number of sources by  $N_s$  and the number of receivers by  $N_r$ , the forward problem must be solved  $N_s \times N_r$  times instead of just  $N_s + N_r$  times when a global initial model is available. Hence it is only feasible in practice because a comparably fast 1D solver can be used for the forward problem. Lamara actually used GEMINI in the first iteration which is one of the forward solvers currently supported by ASKI.

The path-specific models are obtained by first conducting a grid search for crustal thickness and crustal average shear wave velocities followed by a 1D waveform inversion. The 1D path-specific models thus obtained, are used afterwards as reference models  $\mathbf{m}_{\text{pth}}^0$  for the kernel computation in the first iteration of the 3D inversion. By this approach, eq. (9) is now dependent on a specific event-station path and reads (multiplied by  $\sigma_i$ )

$$d_i - s_i^0[\mathbf{m}_{\text{pth}}^0] = \int_{\oplus} \mathbf{k}_i^0[\mathbf{m}_{\text{pth}}^0] \cdot (\mathbf{m}_{3\text{D}}^1 - \mathbf{m}_{\text{pth}}^0) d^3\mathbf{x}. \quad (23)$$

Here,  $[\mathbf{m}_{\text{pth}}^0]$  indicates the dependence of the synthetic data and the kernels on the path-specific starting model and subscript “3D” is added to the notation of the inverted model of the first iteration. Note that such a path-specific approach is only sensible to be used in the first iteration  $n = 1$  in order to initiate the iterative 3D inversion.

Initiating the waveform inversion process with a collection of different models, it is not straightforward how to invert for a single 3D model. Therefore, we introduce an auxiliary global 1D reference model  $\mathbf{m}_{\text{ref}}^0$ . Adding a zero to eq. (23) we obtain:

$$\begin{aligned} d_i - s_i^0[\mathbf{m}_{\text{pth}}^0] &= \\ &\int_{\oplus} \mathbf{k}_i^0[\mathbf{m}_{\text{pth}}^0] \cdot (\mathbf{m}_{3\text{D}}^1 - \mathbf{m}_{\text{ref}}^0 + \mathbf{m}_{\text{ref}}^0 - \mathbf{m}_{\text{pth}}^0) d^3\mathbf{x} \\ &= \int_{\oplus} \mathbf{k}_i^0[\mathbf{m}_{\text{pth}}^0] \cdot (\mathbf{m}_{3\text{D}}^1 - \mathbf{m}_{\text{ref}}^0) d^3\mathbf{x} \\ &+ \int_{\oplus} \mathbf{k}_i^0[\mathbf{m}_{\text{pth}}^0] \cdot (\mathbf{m}_{\text{ref}}^0 - \mathbf{m}_{\text{pth}}^0) d^3\mathbf{x}. \end{aligned} \quad (24)$$

Redefining the model update of the first iteration as

$$\delta\mathbf{m}^1 = \mathbf{m}_{3\text{D}}^1 - \mathbf{m}_{\text{ref}}^0 \quad (25)$$

and introducing a correction term accounting for the difference between the global 1D model and the path-specific one,

$$c_i^0 = \int_{\oplus} \mathbf{k}_i^0[\mathbf{m}_{\text{pth}}^0] \cdot (\mathbf{m}_{\text{ref}}^0 - \mathbf{m}_{\text{pth}}^0) d^3\mathbf{x}, \quad (26)$$

we obtain equations

$$d_i - s_i^0[\mathbf{m}_{\text{pth}}^0] - c_i^0 = \int_{\oplus} \mathbf{k}_i^0[\mathbf{m}_{\text{pth}}^0] \cdot \delta\mathbf{m}^1 d^3\mathbf{x} \quad (27)$$

for which the unknown model update  $\delta\mathbf{m}^1$  is no longer path dependent. Hence, a linear system consisting of equations (27) (along with additional regularization conditions) is suitable to derive the new 3D model  $\mathbf{m}_{3\text{D}}^1$  and can simply be substituted in our new inversion procedure, replacing eq. (9) after division by  $\sigma_i$ . As our new approach allows to change forward solvers between iterations and also permits data modifications when updating the model, the path-specific approach can be easily realized using ASKI.

The global reference model  $\mathbf{m}_{\text{ref}}^0$  should be chosen close to all path-specific starting models  $\mathbf{m}_{\text{pth}}^0$  in order to keep the correction terms small. Besides that it has no physical implications and is not restricted to 1D models. Lamara (2015) calculates

$\mathbf{m}_{\text{ref}}^0$  as the average of all path-specific 1D models at each depth. As such, it is not a realistic earth model but perfectly suitable for fitting the path-specific approach into the general FWI inversion scheme. In addition, it has a smooth transition from crust to mantle facilitating the graphical display of model perturbations.

#### 4.2.2. Dataset, model parameterization and computational requirements

Lamara (2015) selected high-quality seismic data from 72 events with magnitudes ranging from 3.7 to 5.7 which occurred between October 2005 and March 2007. He used in total 2695 vertical component seismograms of 360 s length recorded at 65 stations (see fig. 9). The data was filtered in the frequency range of 0.03 Hz to 0.1 Hz in order to avoid noise owing to microseisms. No further processing of the seismograms was applied, except a Fourier transform at the 26 discrete frequencies  $f = \frac{1}{360\text{s}}\ell$ ,  $11 \leq \ell \leq 36$ , which is the frequency discretization of the data we chose for ASKI. Accounting for real and imaginary part of the complex-valued spectral data, the total number of real-valued data samples  $d_i$  is 140 140.

The inversion grid was chosen according to the model resolution in crust and upper mantle that can be expected from this frequency range of data: Laterally the inversion grid covers  $7 \times 9$  degrees (as apparent in figs. 9, 10) with a uniform cells size of 15.2 km, whereas the depth coverage of 100 km is subdivided into cells of 3 km thickness in the crust (up to 30 km depth), 5 km thickness in the range of 30-60 km depth and 10 km thickness down to 100 km depth. This results in a total of 67 320 inversion grid cells for which  $v_s$  is determined only.

The path-specific approach is computationally very demanding and actually contradicts the computational efficiency of the SI approach of ASKI, since forward wavefields and Green functions must be independently computed for every path and not only for source and receiver positions, respectively. However, by the general modularization of the software package ASKI, we can realize this variation of the SI method nonetheless. The computation of the spectral wavefields and Green functions by GEMINI for one source-receiver pair for a total of 1.99 million forward grid points takes approx. 75 min in parallel on 16 CPU cores. The required storage for the wavefield and strains for one seismic source as well as Green tensor and Green strains for one receiver sums to 13.89 GiB. Computing the respective pre-integrated sensitivity kernel afterwards for one receiver component and one elastic parameter takes only a few minutes on a single CPU core and requires storage of only 13.35 MiB.

From the second iteration onwards, the inversion follows the procedure as described in section 3, i.e. all wavefields from the 72 seismic sources and the Green functions for the 65 receiver components are computed separately based on the 3D background model of the iteration and are combined for kernel computation of all source-receiver pairs. This results in very different computational requirements compared with the path-specific first iteration. We are conducting the second iteration using SPECSEM3D GLOBE as a forward solver, using a similar number of almost 2 million forward grid points (all inner points of spectral elements). This yields storage requirements for wavefield and strains of 3.46 GiB for each of the 72 seismic sources as well as each of the 65 Green functions and Green strains (accounting for vertical receiver components only). In total 474.54 GiB of wavefield data must be stored before pre-integrated kernels are computed which require the very same amount of storage as in the first iteration (13.35 MiB for one path, 35.15 GiB in total), since the inversion grid resolution is not changed. Regarding computation time, each of the  $72 + 65 = 137$  SPECSEM3D GLOBE simulations takes approx.

94 minutes on 80 CPU cores. In total, this equals 17 171 CPU hours, i.e. total computation time relative to a single core. For comparison, the total computation time for *all* source-receiver pairs of the 1D forward solver GEMINI in the path-specific first iteration equals 53 900 CPU hours, even though GEMINI is more than 12 times faster in terms of CPU hours than SPECFEM3D doing a single simulation. This reflects the very high computational demands of the path-specific approach.

#### 4.2.3. Short summary of inversion results of first iteration

The first iteration of the waveform inversion following the path-specific approach was conducted with joint application of damping and depth-depending smoothing conditions. Already after the first iteration, the 3D model shows in great detail the strong variations of shear wave velocity and Moho depth in the Aegean and highlights subtle features of the forearc of the Hellenic subduction zone (fig. 10). Velocity variations at this depth can be mainly explained by variations of crustal thickness featuring mantle velocities in some regions and crustal velocities in others. The normalized misfit (22) for the path-specific starting models is very small, by construction of the  $\mathbf{m}_{\text{pth}}^0$  (fig. 11, *top left*). Since a particular model  $\mathbf{m}_{\text{pth}}^0$ , however, needs to explain a single seismogram only and the path-specific models are even allowed to contradict each other, a comparison with the fit for 3D model  $\mathbf{m}_{3D}^1$  is not fair. The overall fit for the global reference model  $\mathbf{m}_{\text{ref}}^0$ , on the other hand, is relatively bad, since as a 1D global model it cannot explain the complex geological structure in the Aegean (fig. 11, *top right*). Looking at the general fit for the 3D inverted model  $\mathbf{m}_{3D}^0$  (fig. 11, *bottom*), we observe an improvement compared to  $\mathbf{m}_{\text{ref}}^0$ , but a worse overall fit compared with the path-specific starting models. This is a fair indication that the inversion goes in the right direction, even though  $\mathbf{m}_{\text{ref}}^0$  is the average of all path-specific 1D models and, hence, has limited physical meaning. Further iterations of this inversion are necessary to evaluate the convergence behaviour and finally derive a reliable model. More details on this work and results of further 3D iterations using SPECFEM3D GLOBE as a forward solver for ASKI will be presented elsewhere.

## 5. Conclusions

In spite of its obvious advantages with regard to computational efficiency and convergence over gradient-based full waveform inversion methods, wide-spread use of scattering-integral-based seismic full waveform inversion (SI-FWI) has been hampered in the past by its enormous demand of disk space for storing sensitivity kernels. With a new conceptual approach to SI-FWI, we make kernel storing feasible in practice and open access to up to now unnoticed or unfeasible opportunities offered by SI-FWI.

The basic new feature of our approach is a complete decoupling of model space discretization from wave propagation space discretization. The latter is prescribed by forward solvers and must ensure stability of the wave propagation simulation. The former rather depends on available data, expected resolution and available a priori information. Model resolution is controlled by the dominating wavelength while propagation space discretization should resolve a fraction of the smallest occurring wavelength. Hence, pre-integration of sensitivity kernels onto the model space discretization massively reduces required computer storage and ultimately makes kernel storage possible in practice. We have implemented pre-integration routines that support point-cloud wavefield data from the forward solver side and an adaptive subdivision into subvolumes on the model space side.

Another feature of our approach which greatly helps in keeping computation time and storage requirements at a minimum is

to perform the inversion in the frequency domain. Wave propagation solvers in the time domain are extended by a small routine that performs a Fourier transform on-the-fly during time stepping. Frequency-domain inversion is particular useful with inversion strategies that start the iteration process using low-frequency data and increase the frequency content in later iterations. With our approach, wave propagation discretization and model space discretization can be adapted to the frequency range of the current iteration and computational effort and storage can be reduced to the necessary minimum.

Storing of kernels as well as decoupling of model space discretization has several desirable consequences when applying SI-FWI. It essentially allows to completely decouple the forward problem from the inverse problem and to split kernel computation and model update. Synthetic wavefields produced by any seismic wave propagation code are sampled on a point grid (the forward grid) and used by a separate kernel computation program to path-wisely calculate pre-integrated sensitivity kernels for the independently chosen model space discretization which are written to disk. A third program reads the kernels needed for the data to be inverted and sets up a system of linear equations enhanced by damping and smoothing constraints. This system is solved in a least-squares sense to produce a model update. This last step is relatively cheap and can be repeated many times for different regularization constraints without renewed forward or kernel computations. Modifications to the data are still possible in this last step. The code for the model update is written in a very generic way and is applicable to any choice of model space discretization.

Organizational separation into three separate stages enables a flexible, modular and extendable implementation of our inversion approach. Result is our software package ASKI – Analysis of Sensitivity and Kernel Inversion – written as an object-oriented library of modern FORTRAN and Python. It is freely available at <http://www.rub.de/aski> under the terms of the GNU General Public License (<http://www.gnu.org>).

## Acknowledgments

This work was supported by the German Federal Ministry of Education and Research through grant 03G0752C, as well as by the German Research Foundation (DFG) through the Collaborative Research Centre 526 “Rheology of the Earth – from the upper crust to the subduction zone”. We are grateful for the helpful advice given by two anonymous reviewers.

## References

- V. Akcelik, G. Biros, and O. Ghattas. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation. In *Supercomputing, ACM/IEEE 2002 Conference*, pages 41–41. IEEE, 2002.
- K. Aki and P. G. Richards. *Quantitative seismology, theory and methods*. W. H. Freeman and Company, 1980.
- E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users’ Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999. ISBN 0-89871-447-8 (paperback).
- R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter estimation and inverse problems*. Academic Press, 2013.
- G. Backus and F. Gilbert. The resolving power of gross earth data. *Geophysical Journal International*, 16(2):169–205, 1968.



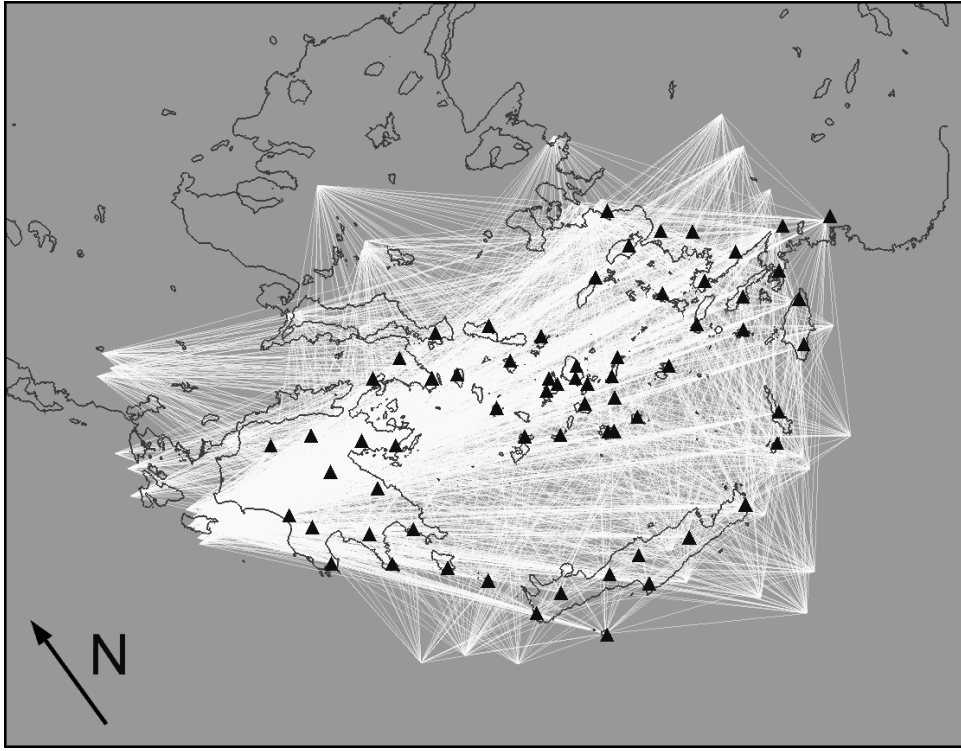


Figure 9: Path coverage of the data set described in section 4.2.2. Triangles indicate the 65 receiver locations.

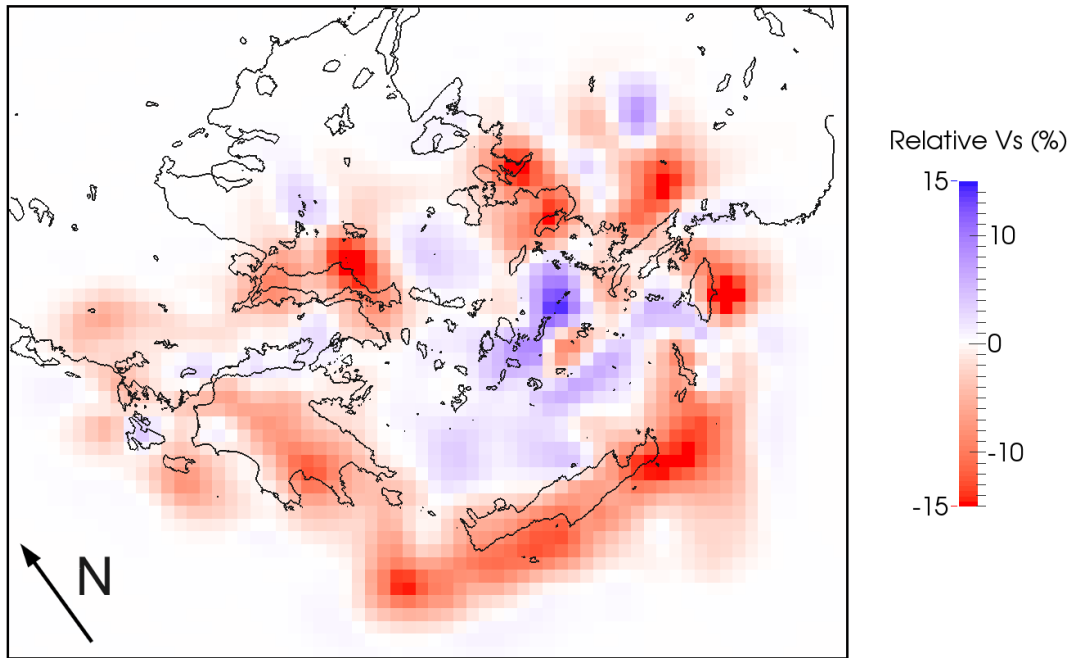


Figure 10: 3D  $v_s$  model after the first iteration derived by Lamara (2015) at 30 km depth (average Moho depth in the Aegean). Variations of shear wave velocity relative to the reference model  $\mathbf{m}_{\text{ref}}^0$  are shown (reference velocity at this depth is  $3.77 \frac{\text{km}}{\text{s}}$ ). The 3D model resolves in high details the strong variation of shear wave velocity and Moho depth in the Aegean and highlights subtle features of the forearc. For the first time in a tomographic study, the extension and crustal thinning beneath the Gulf of Corinth are imaged and the volcanic arc is clearly delineated, especially in its eastern part. The inversion domain extends from the surface to 100 km depth. For more details we refer to Lamara (2015).

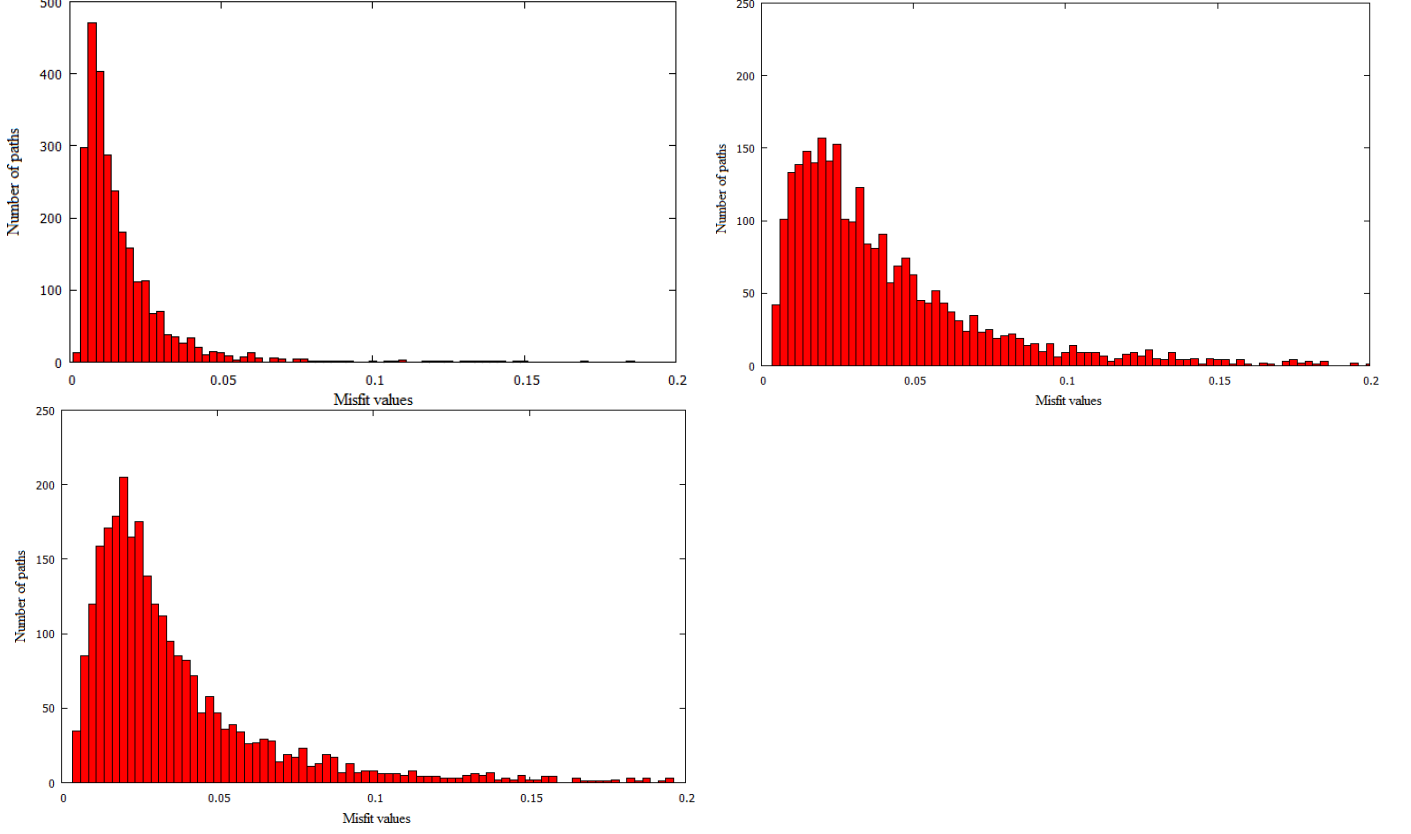


Figure 11: Histograms of normalized misfit (22) for all paths. Misfits based on synthetics computed w.r.t. path-specific starting models  $\mathbf{m}_{\text{pth}}^0$  (*top left*), global reference model  $\mathbf{m}_{\text{ref}}^0$  (*top right*) and inverted 3D model  $\mathbf{m}_{3\text{D}}^1$  (*bottom*). The general fit for the path-specific starting models is very good, by construction of the  $\mathbf{m}_{\text{pth}}^0$ . A comparison with the fit for 3D model  $\mathbf{m}_{3\text{D}}^1$ , however, is not fair. The overall fit for the global reference model  $\mathbf{m}_{\text{ref}}^0$  is relatively bad, since as a 1D global model it cannot explain the complex geological structure in the Aegean. Looking at the general fit for the 3D inverted model  $\mathbf{m}_{3\text{D}}^1$ , we observe an improvement compared to  $\mathbf{m}_{\text{ref}}^0$ , but a worse overall fit compared with the path-specific starting models. This is a fair indication that the inversion goes in the right direction, even though  $\mathbf{m}_{\text{ref}}^0$  is the average of all path-specific 1D models and, hence, has limited physical meaning. Further iterations will reveal the convergence behaviour of this inversion more clearly.

- G. Backus and F. Gilbert. Uniqueness in the inversion of inaccurate gross earth data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 266(1173):123–192, 1970.
- G. E. Backus and J. Gilbert. Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal International*, 13(1-3):247–276, 1967.
- A. Bamberger, G. Chavent, C. Hemon, and P. Lailly. Inversion of normal incidence seismograms. *Geophysics*, 47(5):757–770, 1982.
- Å. Björck, T. Elfving, and Z. Strakoš. Stability of conjugate gradient and Lanczos methods for linear least squares problems. *SIAM Journal on Matrix Analysis and Applications*, 19(3):720–736, 1998.
- L. S. Blackford, J. Choi, A. Cleary, E. D’Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users’ Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997. ISBN 0-89871-397-8 (paperback).
- L. Bos and K. Salkauskas. Moving least-squares are Backus-Gilbert optimal. *Journal of approximation theory*, 59(3):267–275, 1989.
- C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent. Multi-scale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, 1995.
- S. Butzer, A. Kurzmann, and T. Bohlen. 3D elastic full-waveform inversion of small-scale heterogeneities in transmission geometry. *Geophysical Prospecting*, 61(6):1238–1251, 2013.
- C. Chapman. *Fundamentals of seismic wave propagation*. Cambridge University Press, 2004.
- P. Chen, T. H. Jordan, and L. Zhao. Full three-dimensional tomography: A comparison between the scattering-integral and adjoint-wavefield methods. *Geophysical Journal International*, 170(1):175–181, 2007.
- S. C. Constable, R. L. Parker, and C. G. Constable. Occams inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *GEOPHYSICS*, 52(3):289–300, 1987. doi: 10.1190/1.1442303.
- R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 100(1):32–74, 1928.
- I. Epanomeritakis, V. Akelik, O. Ghattas, and J. Bielak. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. *Inverse Problems*, 24(3):034015, 2008.
- A. Fichtner and J. Trampert. Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, 185(2):775–798, 2011.
- A. Fichtner, B. L. Kennett, H. Igel, and H.-P. Bunge. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725, 2009.
- R. Franke and G. Nielson. Smooth interpolation of large sets of scattered data. *International Journal for Numerical Methods in Engineering*, 15(11):1691–1704, 1980. ISSN 1097-0207. doi: 10.1002/nme.1620151110. URL <http://dx.doi.org/10.1002/nme.1620151110>.
- W. Friederich. The S-velocity structure of the East Asian mantle from inversion of seismic shear and surface waveforms. *Geophys. J. Int.*, 153:88–102, 2003.
- W. Friederich and J. Dalkolmo. Complete synthetic seismograms for a spherically symmetric earth by a numerical computation of Green’s function in the frequency domain. *Geophys. J. Int.*, 122:537–550, 1995.
- W. Friederich and T. Meier. Temporary seismic broadband network acquired data on Hellenic subduction zone. *EOS Transactions AGU*, 89(40):378, 2008.
- N. Fuji, S. Chevrot, L. Zhao, R. J. Geller, and K. Kawai. Finite-frequency structural sensitivities of short-period compressional body waves. *Geophysical Journal International*, 190(1):522–540, 2012.
- O. Gauthier, J. Virieux, and A. Tarantola. Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics*, 51(7):1387–1403, 1986.
- G. Goertzel. An algorithm for the evaluation of finite trigonometric series. *American mathematical monthly*, pages 34–35, 1958.
- S. M. Hanasoge and J. Tromp. Full waveform inversion for time-distance helioseismology. *The Astrophysical Journal*, 784(1):69, 2014.
- J. Hudson and J. Heritage. The use of the born approximation in seismic scattering problems. *Geophysical Journal International*, 66(1):221–240, 1981.
- P. Lailly. The seismic inverse problem as a sequence of before stack migrations. In *Conference on inverse scattering: theory and application*, pages 206–220. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983.
- S. Lamara. *3D waveform tomography of the Hellenic subduction zone*. doctoral dissertation, Institute of Geology, Mineralogy and Geophysics, Ruhr-Universität Bochum, 2015. urn:nbn:de:hbz:294-44153.
- L. Lambrecht. *Forward and inverse modeling of seismic waves for reconnaissance in mechanized tunneling*. doctoral dissertation, Institute of Geology, Mineralogy and Geophysics, Ruhr-Universität Bochum, 2015. urn:nbn:de:hbz:294-44164.
- D. Levin. The approximation power of moving least-squares. *Mathematics of Computation of the American Mathematical Society*, 67(224):1517–1531, 1998.
- D. Levin. Stable integration rules with scattered integration points. *Journal of Computational and Applied Mathematics*, 112:181–187, 1999.
- D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.
- Q. Liu and J. Tromp. Finite-frequency kernels based on adjoint methods. *Bull. Seismol. Soc. Am.*, 96(6):2383–2397, 2006. doi: 10.1785/0120060041.

S. Łukaszuk. A new concept of probability metric and its applications in approximation of scattered datasets. *Computational Mechanics*, 33(4):299–304, 2004. ISSN 0178-7675. doi: 10.1007/s00466-003-0532-2. URL <http://dx.doi.org/10.1007/s00466-003-0532-2>.

A. V. Masjukov and V. V. Masjukov. Multiscale modification of Shepard’s method for multivariate interpolation of scattered data. In *Proceedings: 10th International Conference “Mathematical Modelling and Analysis 2005” and 2nd International Conference “Computational Methods in Applied Mathematics”*, pages 467–472. Technika, Vilnius, 2005.

V. Monteiller, S. Chevrot, D. Komatitsch, and Y. Wang. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method. *Geophysical Journal International*, 202(2):811–827, 2015.

P. Mora. Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52(9):1211–1228, 1987.

T. Nissen-Meyer, F. Dahlen, and A. Fournier. Spherical-earth Fréchet sensitivity kernels. *Geophysical Journal International*, 168(3):1051–1066, 2007.

G. Nolet. *A breviary of seismic tomography: Imaging the interior of the earth and sun*. Cambridge University Press, 2008. ISBN 9780521882446.

R. L. Parker. *Geophysical inverse theory*. Princeton University Press, 1994.

G. Pratt, C. Shin, and Hicks. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362, 1998.

R. G. Pratt. Inverse theory applied to multi-source cross-hole tomography. *Geophysical Prospecting*, 38(3):311–329, 1990.

F. Schumacher. *Modularized iterative full seismic waveform inversion for 3D-heterogeneous media based on waveform sensitivity kernels*. doctoral dissertation, Institute of Geology, Mineralogy and Geophysics, Ruhr-Universität Bochum, 2014. urn:nbn:de:hbz:294-40511.

F. Schumacher. *ASKI–Analysis of Sensitivity and Kernel Inversion, user manual*, 2015. <http://www.rub.de/aski>.

D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, ACM ’68, pages 517–524, New York, NY, USA, 1968. ACM. doi: 10.1145/800186.810616. URL <http://dx.doi.org/10.1145/800186.810616>.

H. V. Sorensen, C. Burrus, and D. L. Jones. A new efficient algorithm for computing a few DFT points. In *Circuits and Systems, 1988., IEEE International Symposium on*, pages 1915–1918. IEEE, 1988.

A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984.

A. Tarantola. *Inverse problem theory: Methods for data fitting and model parameter estimation*. Elsevier, 1987.

J. Tromp, C. Tape, and Q. Liu. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1):195–216, 2005. ISSN 1365-246X. doi: 10.1111/j.1365-246X.2004.02453.x. URL <http://dx.doi.org/10.1111/j.1365-246X.2004.02453.x>.

J. Tromp, D. Komatitsch, and Q. Liu. Spectral-element and adjoint methods in seismology. *Communications in Computational Physics*, 3(1):1–32, 2008.

L. Zhao and S. Chevrot. An efficient and flexible approach to the calculation of three-dimensional full-wave Fréchet kernels for seismic tomography–I. Theory. *Geophysical Journal International*, 185(2):922–938, 2011.

## A. Kernel formulae for elastic (an)isotropy

In the following, we will derive expressions  $K_m^p$  of waveform sensitivity kernels by which eq. (11) rewrites as

$$\delta u_m(\mathbf{r}) = \int_{\oplus} \left[ \sum_p \delta m_p(\mathbf{x}) K_m^p(\mathbf{x}; \mathbf{r}) \right] d^3\mathbf{x} . \quad (\text{A.1})$$

Here,  $\mathbf{m}(\mathbf{x}) = (m_1(\mathbf{x}), \dots, m_p(\mathbf{x}), \dots, m_{N_M}(\mathbf{x}))^T$  represents the vector field of material properties (depending on the elastic symmetry and including density  $\rho$ ) and  $p$  indicates a specific material property, which occurs as superscript in the notation of the kernel. We simplify notation here by omitting any dependence on angular frequency  $\omega$ . Remember that the dependence on the seismic source is not indicated, either. A more detailed derivation, including illustrations and examples, can be found in Schumacher (2014, ch. 3).

### A.1. Sensitivity kernels for general anisotropy

In general, the kernels  $K_m^p(\mathbf{x}; \mathbf{r})$  for elastic properties  $p$  can be expressed in terms of the unperturbed strain fields

$$\begin{aligned} e_{k\ell} &= e_{\ell k} = \frac{1}{2} (\partial_\ell u_k(\mathbf{x}) + \partial_k u_\ell(\mathbf{x})) , \\ \gamma_{k\ell} &= \gamma_{\ell k} = \frac{1}{2} (\partial_\ell G_{km}(\mathbf{x}; \mathbf{r}) + \partial_k G_{\ell m}(\mathbf{x}; \mathbf{r})) , \end{aligned} \quad (\text{A.2})$$

where the wavefield  $u_k(\mathbf{x})$  and and Green tensor  $G_{km}(\mathbf{x}; \mathbf{r})$  are computed for the unperturbed background model. Using this strain nomenclature, as well as the Voigt notation and the well-known symmetries of the elasticity tensor  $C$ , the term  $-\delta C_{k\ell qr}(\mathbf{x}) (\partial_r u_q(\mathbf{x})) \partial_\ell G_{km}(\mathbf{x}; \mathbf{r})$  in eq. (11) can be rewritten as

$$\begin{aligned} & -\delta C_{k\ell qr}(\mathbf{x}) (\partial_r u_q(\mathbf{x})) \partial_\ell G_{km}(\mathbf{x}; \mathbf{r}) = \\ & -\delta C_{11} (\gamma_{11} e_{11}) - \delta C_{12} (\gamma_{11} e_{22} + \gamma_{22} e_{11}) \\ & -\delta C_{13} (\gamma_{11} e_{33} + \gamma_{33} e_{11}) - \delta C_{14} (2\gamma_{11} e_{23} + 2\gamma_{23} e_{11}) \\ & -\delta C_{15} (2\gamma_{11} e_{13} + 2\gamma_{13} e_{11}) - \delta C_{16} (2\gamma_{11} e_{12} + 2\gamma_{12} e_{11}) \\ & -\delta C_{22} (\gamma_{22} e_{22}) - \delta C_{23} (\gamma_{22} e_{33} + \gamma_{33} e_{22}) \\ & -\delta C_{24} (2\gamma_{22} e_{23} + 2\gamma_{23} e_{22}) - \delta C_{25} (2\gamma_{22} e_{13} + 2\gamma_{13} e_{22}) \\ & -\delta C_{26} (2\gamma_{22} e_{12} + 2\gamma_{12} e_{22}) \\ & -\delta C_{33} (\gamma_{33} e_{33}) - \delta C_{34} (2\gamma_{33} e_{23} + 2\gamma_{23} e_{33}) \\ & -\delta C_{35} (2\gamma_{33} e_{13} + 2\gamma_{13} e_{33}) - \delta C_{36} (2\gamma_{33} e_{12} + 2\gamma_{12} e_{33}) \\ & -\delta C_{44} (4\gamma_{23} e_{23}) - \delta C_{45} (4\gamma_{23} e_{13} + 4\gamma_{13} e_{23}) \\ & -\delta C_{46} (4\gamma_{23} e_{12} + 4\gamma_{12} e_{23}) \\ & -\delta C_{55} (4\gamma_{13} e_{13}) - \delta C_{56} (4\gamma_{13} e_{12} + 4\gamma_{12} e_{13}) \\ & -\delta C_{66} (4\gamma_{12} e_{12}) . \end{aligned} \quad (\text{A.3})$$

Hence, by eq. (11) and for full anisotropy, the kernels for the elastic moduli  $C_{\alpha\beta}$  equal their respective coefficient in eq. (A.3), e.g.  $K_m^{C_{45}}(\mathbf{x}; \mathbf{r}) = -(4\gamma_{23} e_{13} + 4\gamma_{13} e_{23})$ . The density kernel reads  $K_m^\rho(\mathbf{x}; \mathbf{r}) = \omega^2 u_k(\mathbf{x}) G_{km}(\mathbf{x}; \mathbf{r})$ .

## A.2. Sensitivity kernels for isotropic elastic earth models

Applying formula (A.3), kernels for any elastic symmetry can be derived. In case of an isotropic medium, the elasticity tensor  $C$  has only two independent parameters and, using Lamé moduli  $\lambda$  and  $\mu$ , may be written as

$$C_{k\ell qr} = \lambda \delta_{k\ell} \delta_{qr} + \mu (\delta_{kq} \delta_{\ell r} + \delta_{kr} \delta_{\ell q}) . \quad (\text{A.4})$$

Inserting this expression into (A.3), formulae for waveform sensitivity kernels for the set of material properties  $\mathbf{m}(\mathbf{x}) = (\lambda(\mathbf{x}), \mu(\mathbf{x}), \rho(\mathbf{x}))^T$  are obtained as:

$$\begin{aligned} K_m^\lambda(\mathbf{x}; \mathbf{r}) &= -(\gamma_{11} + \gamma_{22} + \gamma_{33})(e_{11} + e_{22} + e_{33}) \\ K_m^\mu(\mathbf{x}; \mathbf{r}) &= -2(\gamma_{11}e_{11} + \gamma_{22}e_{22} + \gamma_{33}e_{33}) \\ &\quad -4(\gamma_{23}e_{23} + \gamma_{13}e_{13} + \gamma_{12}e_{12}) \\ K_m^\rho(\mathbf{x}; \mathbf{r}) &= \omega^2 u_k(\mathbf{x}) G_{km}(\mathbf{x}; \mathbf{r}) . \end{aligned} \quad (\text{A.5})$$

In seismic applications, the compressional and shear wave speeds,  $v_p$  and  $v_s$ , are of interest. In order to derive waveform sensitivity kernels for an isotropic model represented by the set of material properties  $\mathbf{m}(\mathbf{x}) = (v_p(\mathbf{x}), v_s(\mathbf{x}), \rho(\mathbf{x}))^T$  we can make use of formulae (A.5) and relate  $(v_p, v_s, \rho)^T$  to  $(\lambda, \mu, \rho)^T$  in a linearized sense by total derivatives (cf. Schumacher, 2014, ch. 3.2.1). Denoting the unperturbed background model by  $(v_p^0, v_s^0, \rho^0)^T$ , we find

$$\begin{aligned} K_m^{v_p}(\mathbf{x}; \mathbf{r}) &= 2\rho^0 v_p^0 K_m^\lambda(\mathbf{x}; \mathbf{r}) , \\ K_m^{v_s}(\mathbf{x}; \mathbf{r}) &= 2\rho^0 v_s^0 K_m^\mu(\mathbf{x}; \mathbf{r}) - 4\rho^0 v_s^0 K_m^\lambda(\mathbf{x}; \mathbf{r}) , \\ \tilde{K}_m^\rho(\mathbf{x}; \mathbf{r}) &= \left( (v_p^0)^2 - 2(v_s^0)^2 \right) K_m^\lambda(\mathbf{x}; \mathbf{r}) \\ &\quad + (v_s^0)^2 K_m^\mu(\mathbf{x}; \mathbf{r}) + K_m^\rho(\mathbf{x}; \mathbf{r}) . \end{aligned} \quad (\text{A.6})$$

Note that although density  $\rho$  is a material property in both sets,  $(\lambda, \mu, \rho)^T$  and  $(v_p, v_s, \rho)^T$ , the respective sensitivity kernels are different. Hence, when deriving sensitivity kernels for some specific material properties, always the complete set of properties must be taken into account. For this reason, the density kernels are denoted by different symbols  $K^\rho$  and  $\tilde{K}^\rho$ , respectively.

## B. Synthetic validation of ASKI – more details on application in Section 4.1

Since the standard absorbing boundaries of SPEC-FEM3D Cartesian 2.1 are not perfectly absorbing, we used a simulation domain for solving the forward problem which is twice as large in every direction of space as the actual inversion domain (fig. 4) in order to avoid unwanted reflections to occur in the wavefields. The approximate size of the uniformly distributed spectral elements in each iteration is displayed in table B1. We choose the number of grid points inside the spectral elements to be  $5 \times 5 \times 5 = 125$ , i.e. using polynomial degree 4 of the spectral element ansatz functions for each direction of space. As the “forward grid” for ASKI, we only use the inner  $3 \times 3 \times 3 = 27$  points of the spectral elements in order to avoid having redundant wavefield information at overlapping points on the element boundaries. Since in ASKI every forward method defines its own forward grid points at which the spectral wavefields are provided for kernel computation, it is in particular possible to choose only a subgrid of the actual simulation point grid.

Table B1 furthermore gives the numbers of uniform inversion grid cells in X,Y and Z-direction. Note that any underestimation of optimal cell sizes should be compensated by the inversion method through the smoothing conditions.

The discretization of simulation grid and inversion grid were chosen in the same way for all three datasets (cross borehole, VSP, shallow surface wave inversion). The smoothing factors, however, which are additionally applied to the smoothing equations (after scaling them separately for each material property  $p = v_p, v_s$  by the maximum of the absolute  $p$ -kernel values) must be chosen differently for each setting, since they strongly depend on the maximum kernel values which themselves depend on the particular radiation of the seismic waves. As a result, the smoothing factors, as shown in table B1, are not comparable across the three datasets and were chosen in such a way that model updates stay within reasonable ranges for which the non-linear inversion process turned out to be stable. Since in the surface wave inversion especially P-waves have a weak influence on the result, we gave priority to perturbation of  $v_s$  when choosing the smoothing factors for the surface wave case. As a result, the  $v_p$  model in the first iteration is unrealistic, which however does not prevent the method from yielding sensible results (discussed in more detail below).

Instead of everywhere applying standard average smoothing conditions of the form

$$\delta m_{pj}^n = \frac{1}{\text{no. of neighb. of cell } j} \sum_{\substack{\ell, \text{ neighb.} \\ \text{of cell } j}} \delta m_{p\ell}^n , \quad (\text{B.1})$$

it turned out to be more stable to apply what we call *zero-boundary smoothing conditions* for cells  $j$  on the outer boundaries of the inversion domain:

$$\delta m_{pj}^n = \frac{1}{6} \sum_{\substack{\ell, \text{ neighb.} \\ \text{of cell } j}} \delta m_{p\ell}^n , \quad (\text{B.2})$$

inspite of the number of neighbours of those cells  $j$  being strictly smaller than 6. This is equivalent to assuming that missing neighbouring cells beyond the boundary do exist but have values  $\delta m_{p\ell}^n = 0$ . Thereby, the additional assumption that the model update is zero beyond the outer boundaries of the inversion grid enters the inversion process. While eq. (B.2) represents zero-boundary smoothing conditions, the regular average of eq. (B.1) can be thought of as a *continuous-boundary smoothing condition*: It is equivalent to assuming the existence of missing neighbours  $\ell$  beyond the boundary having the *same* value  $\delta m_{p\ell}^n = \delta m_{pj}^n$  as cell  $j$  (cp. Schumacher, 2014, ch. 4.2.3).

In the following, the inversion results of the surface wave setting are described. Figure B1 shows the inverted  $v_p$  (*left*) and  $v_s$  (*right*) models after the first (*top*), the sixth (*center*) and the last (*bottom*) iteration. The images show slices through the inversion domain as defined by figure 4 (*bottom*). Figure B2 demonstrates the fitting of the waveforms showing an exemplary seismogram as well as the normalized misfit  $\chi^2$  of the frequency-domain data defined by eq. (22).

As expected from the acquisition geometry, the surface wave setting represents the most difficult inverse problem. Since in this setting surface wavetrains have a very strong influence, we focussed on values of  $v_s$  when choosing the smoothing intensities, allowing  $v_p$  to take unrealistic values, if necessary. By just applying smoothing conditions, we could not in general get rid of the strange  $v_p$  anomaly produced in the first iteration. Therefore, we chose smoothing conditions in a way to produce a sensible  $v_s$  model in the first iteration. The strong  $v_p$  anomaly produces a very large misfit after iteration 1 (see fig. B2, *bottom left*) but vanishes immediately after the second iteration. Thereafter, the  $v_p$  model stays close to the starting model for the rest of the inversion. The  $v_s$  anomaly is resolved in principle laterally in the upper part of the model, as could be expected from the acquisition setup. This inversion example has the least initial misfit (cf. fig. B2, *bottom right*) and the

$n$	forward grid edge length	inversion grid			smoothing factors		
		$N_x$	$N_y$	$N_z$	cross-BH	VSP	surface
1	7.0	8	8	6	7.5	2.0	16.0
2	6.5	9	9	7	11.0	2.5	3.0
3	6.5	10	10	7	11.0	3.0	3.0
4	6.0	11	11	8	11.0	1.5	2.0
5	6.0	12	12	9	9.0	2.0	2.0
6	5.0	14	14	10	13.0	1.3	3.0
7	5.0	15	15	11	9.0	1.2	4.0
8	4.5	16	16	12	5.0	1.4	3.0
9	4.5	17	17	13	5.0	1.4	2.0
10	4.0	18	18	14	9.0	1.2	1.2
11	4.0	19	19	14	5.0	1.4	2.0
12	4.0	20	20	15	5.0	2.0	2.0

Table B1: Some more statistics about the iteration steps of the three waveform inversions conducted in section 4.1:  $n$  = iteration step; edge length [m] of (approx. cubic) spectral elements – note that for experimental reasons, edge lengths of the first iterations were chosen a little smaller than necessary;  $N_x$ ,  $N_y$  = number of uniform inversion grid cells in X,Y-direction (equally spaced over 100 m);  $N_z$  = number of uniform inversion grid cells in Z-direction (equally spaced over 75 m); The most-right 3 columns give the additional smoothing factors for each iteration of the cross borehole, VSP and surface wave inversion, respectively. Note that these are not comparable across the three datasets.

total misfit reduction for the complete dataset is only about 45.1%. In the exemplary seismogram in figure B2 (*top*) very little misfit can be observed, most significantly occurring in the surface wave phase which comprise the major parts of the seismograms. This explains why mostly the  $v_s$  anomaly is resolved laterally in the upper part of the model, since surface waves are most sensitive to this part of the earth model.

The surface wave inversion tends to instability in the first iterations (fig. B2, *bottom*) indicating that the information content of the waveform data is not sufficient to constrain both  $v_p$  and  $v_s$  equally well, in particular in the lower parts of the anomaly.

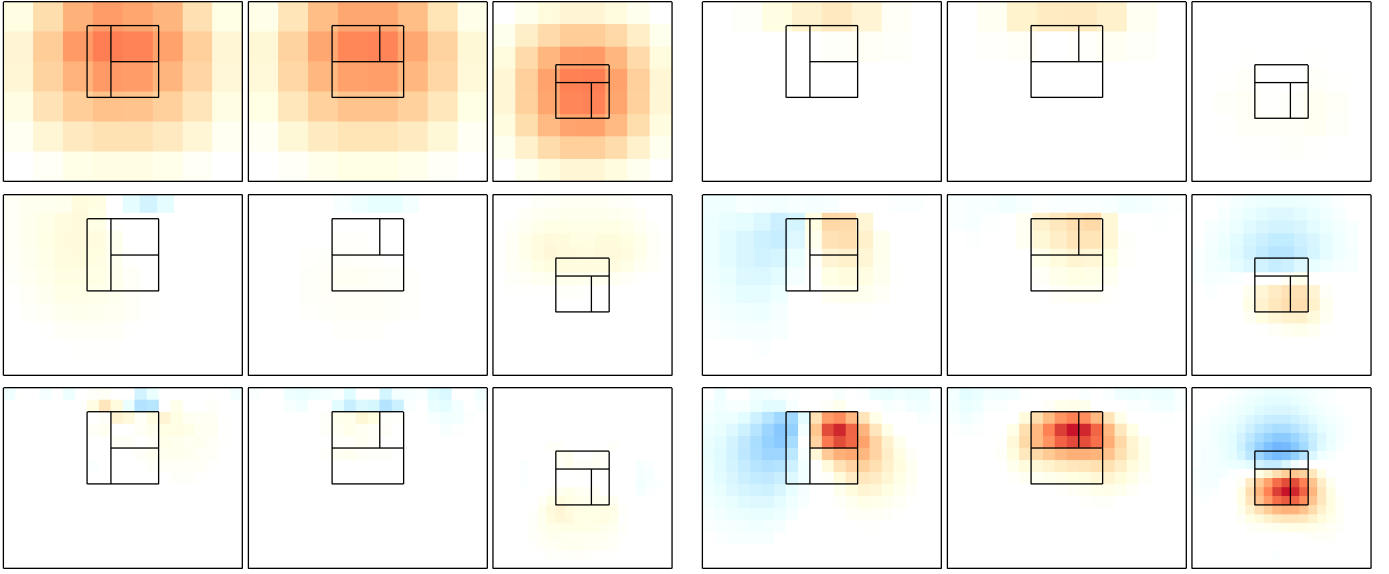


Figure B1: Inverted model from surface wave inversion. P-wave speed (*left*), S-wave speed (*right*); colors as defined by colorbars in fig. 4. From *top* to *bottom*: model after iteration 1, iteration 6, iteration 12; slices as shown in fig. 4. The  $v_s$  anomaly is resolved in principle laterally in the upper part of the model, the  $v_p$  model stays close to the starting model after the unrealistic model of the first iteration, which we could not get rid of by smoothing conditions only. The inversion might be stabilized by taking higher frequencies into account in the first iteration and applying additional damping. For the sake of comparability, however, we followed the same configurations as in the other inversions. Note the refinement of the inversion grid throughout the iterations.

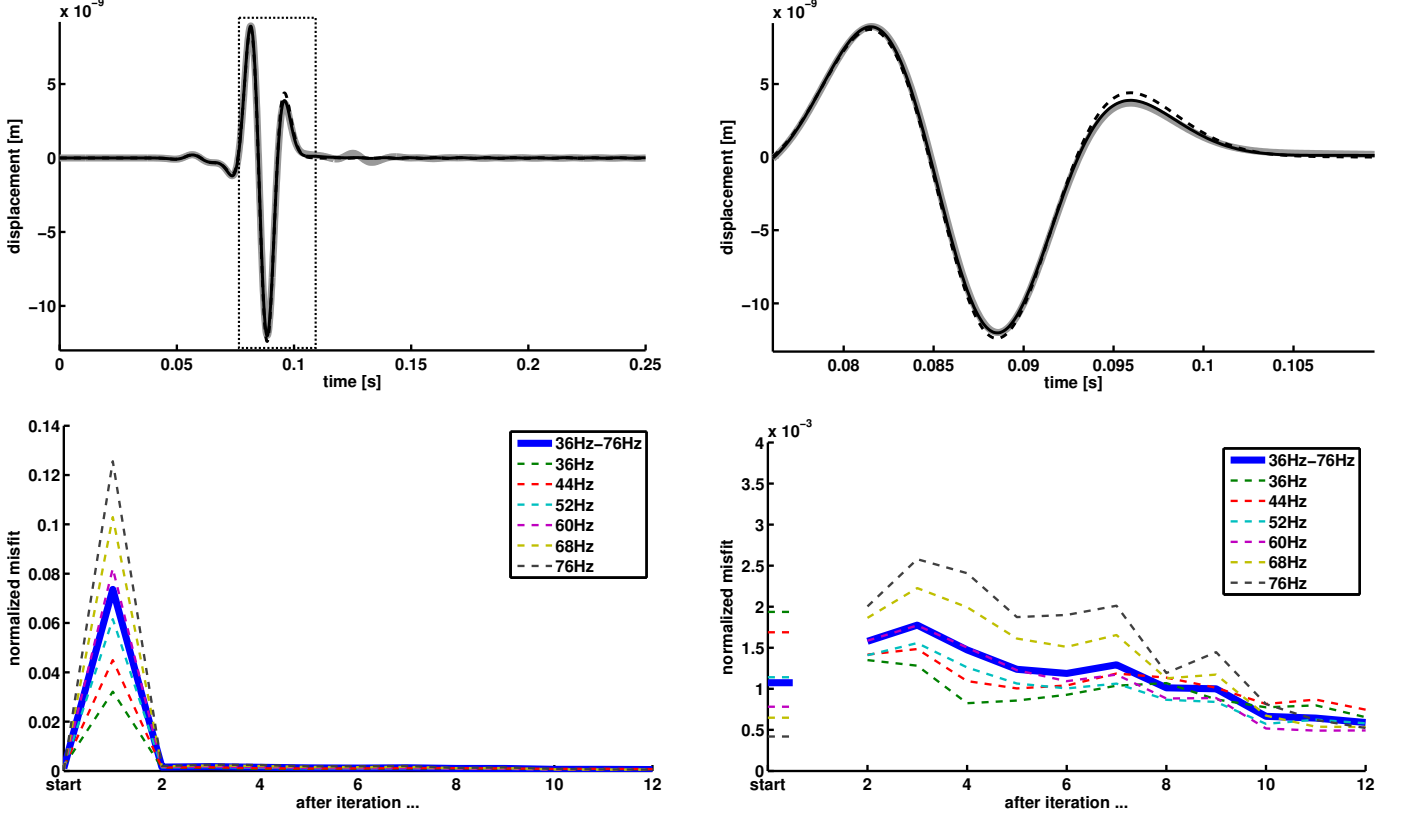


Figure B2: *Top*: Exemplary seismogram from the surface wave inversion, comparing data w.r.t. true earth model (thick grey line) with synthetic waveforms for the homogeneous starting model (dashed black line) and the waveform after the last iteration (solid black line). The *top-right* panel shows a zoomed time window of the surface wave phase, indicated by the box in the *top-left* panel.

*Bottom left*: normalized misfit (22) for different data subsets which vary with frequency content only. *Bottom right*: same as bottom-left panel, but leaving out iteration 1 for better inspection. There is very little misfit for the inversion to work on, mostly occurring in the surface waves. This explains why primarily the  $v_s$  anomaly is resolved laterally in the upper part of the model (cp. fig. B1 *right*), since surface waves are most sensitive to  $v_s$  structure there. The total misfit reduction for the complete data set is only about 45.1 %. Since high-frequency surface waves do not penetrate into the anomaly, high frequencies have the smallest misfit for the starting model.