Ⓔ

Adjoint Tomography for Predicting Earthquake Ground Motion: Methodology and a Blind Test

by Filip Kubina, Filip Michlík, Peter Moczo,[*] Jozef Kristek,[*] and Svetlana Stripajová

Abstract In recent international exercises on numerical prediction of earthquake ground motion (EGM) in local surface sedimentary structures (LSSS), teams with the most advanced numerical-modeling methods reached a very good level of agreement among different methods. The synthetics, however, were not sufficiently close to earthquake records. It was concluded that the structural model must be improved. Here, we apply adjoint tomography to 2D LSSS, aiming to find a model for which EGM characteristics will be sufficiently close to those determined from records. This is an important difference compared to traditional structural inversions. The methodology developed in the exploration, regional, and global scales cannot be directly applied, due to a relatively small amount of data, a relatively large initial waveform misfit, and low frequencies with respect to the size of the structure. We elaborated an inversion procedure specific for the local structures. We present a verification blind test that is closer to real-data inversion than the standard synthetic inversions. A third party provided (a) seismograms numerically simulated for an undisclosed true structure, (b) source parameters, and (c) a homogeneous half-space as the initial model. We demonstrate the quality of the inverted model up to the 4.5 Hz target frequency, using seismograms, waveform misfits, waveform goodness of fit (GOF), and mainly GOF for important EGM characteristics. The development of the 2D procedure, requiring much less computational load compared to the 3D procedure, is the first step. We assume that the procedure can be, in principle, applied to 3D structures after refinements, due to a 3D spatial distribution of sources and receivers.

Electronic Supplement: Overview of the inversion procedure and algorithm, technical details on the implementation of the inversion procedure, and evolution of the medium parameters in the blind test.

Introduction

Prediction of Earthquake Ground Motion

Whether or not seismologists can or cannot predict earthquake occurrence in a timely manner, it is extremely important to predict earthquake ground motion (EGM) during potential future earthquakes in densely populated areas and at sites of special importance. Prediction of what can or will happen during a future earthquake is vital for land-use planning, designing new buildings, and reinforcing existing ones. It is also extremely important for undertaking actions that could help mitigate losses during future earthquakes.

Proximity to a seismogenic fault obviously poses an earthquake threat. A local surface sedimentary structure

(LSSS) can also considerably increase the earthquake hazard. This is because seismic-wave interference and resonant phenomena in sediment-filled basins and valleys can produce anomalously large earthquake motion at the Earth's surface.

Prediction of the EGM for a given area or site might be based on an empirical approach if sufficient earthquake recordings at the site, or physically relevant for the site, are available. In the regions with low-to-moderate earthquake activity, however, seismologists more often face a severe lack of data. Theory and numerical simulations have to be applied.

In recent international numerical exercises on numerical prediction of EGM in LSSS, Third International Symposium on the Effects of Surface Geology on Seismic Motion (ESG) 2006 for Grenoble Valley, France (e.g., Chaljub *et al.*, 2010), and Euroseistest Verification and Validation Project (E2VP) for the Mygdonian basin near Thessaloniki, Greece (e.g., Chaljub *et al.*, 2015; Maufroy *et al.*, 2015, 2016), teams with

*Also at Earth Science Institute, Slovak Academy of Sciences, Dúbravská cesta 9, 84005 Bratislava, Slovakia.

the most advanced numerical-modeling methods reached a very good level of agreement among different methods (finite-difference, spectral element, discontinuous Galerkin, and pseudospectral). The synthetics, however, were not sufficiently close to records of real earthquakes in terms of time–frequency (TF) misfits and misfits in selected ground-motion characteristics (Maufroy *et al.*, 2015). It was concluded that improvement of the available structural model is necessary.

It is natural to think of utilizing the misfit between numerically simulated motion and true motion for improving model of the LSSS.

Full Waveform Inversion

An inversion is a process of obtaining seismic parameters (sources and/or model of the medium) using recorded seismic motion. Full waveform inversion indicates (Fichtner, 2011) the exploitation of as much information as is physically reasonable.

Adjoint tomography is a seismic structural inversion that uses seismic records, the numerical solution of the equation of motion, and a mathematical tool (the adjoint method). The adjoint method gives a recipe for an efficient computation of a kernel, a volumetric density of the Fréchet derivative (gradient) of the misfit, with respect to material parameters. The kernel can be used for iterative improvement of the model. The adjoint tomography has been successfully applied in the regional and global Earth scale and in exploration. An excellent introduction, as well as an overview of the historical development, can be found in the book by Fichtner (2011). Here, we just mention examples of important contributions after 2011. Important recent applications of the 3D full waveform tomography using real data in structural seismology include, for example, the crustal structures in southern California Lee *et al.* (2014), Lee and Chen (2016); in Europe Zhu *et al.* (2012, 2013), Zhu and Tromp (2013), and Fichtner *et al.* (2013); in the North Atlantic Rickers *et al.* (2013); in East Asia Chen *et al.* (2015); and in Japan Simute *et al.* (2016).

Although the resolution of the main features is convincing, smaller details can be still ambiguous, due to the inherent nonuniqueness of the inverse methods (see, e.g., Hosseini and Pezeshk, 2015). Anisotropy poses a challenge, mostly due to its trade-off with isotropic heterogeneity (Liu and Gu, 2012). For many methodological details on full waveform inversion, we refer to a monograph by Chen and Lee (2015).

Multiparameter studies in exploration, for example, Operto *et al.* (2013), Yuan and Simons (2014), and Yuan *et al.* (2015), provide great insight into small subsurface structures. Articles by Sirgue *et al.* (2009, 2010) or Stopin *et al.* (2014) are examples of 3D inversion used in exploration.

In this article, we first explain the relation between an inversion of an LSSS and inversions in the exploration, regional, and global scales. Then, we present a brief overview of basic concepts and principles of the adjoint inversion necessary for the subsequent presentation of the methodol-

ogy of the full waveform adjoint inversion in an LSSS. Eventually, we present a verification of our approach, developed using a set of canonical models, in a blind test with relatively poor waveform data and the simplest possible assumption of an initial structural model.

The potential inaccuracies in the source position, mechanism, and source time function complicate the inversion. In this article, we assume well-determined sources to focus on structural-model inversion.

Full Waveform Adjoint Tomography in a Local Surface Sedimentary Structure: Methodology

Preliminary Considerations

The waveform inversion of LSSS has specific aspects compared with the waveform inversion in the exploration, regional-scale, or global-scale models.

Modeling and Inversion Domain. LSSS is typically several kilometers wide and hundreds of meters deep. The horizontal dimensions are close to the typical dimensions in exploration. The vertical dimensions are often smaller. The true medium is usually highly complex. Velocity contrasts between sediments and bedrock at or close to the free surface commonly reach a value of 10 or more. Sediments can be stratified or with a velocity gradient. Values of elastic parameters range over several orders of magnitude. The latter fact is the main reason why linearized inversion methods do not work well in such a case.

Initial Structural Model. An available initial model of LSSS is usually only poorly determined. The *S* wavespeeds in the initial and true models may differ even by one order of magnitude. Geometry of the sediment–basement interface is easier to know at the free surface but more difficult at depth. Distinct layers within sediments considerably complicate the situation. A large difference between the initial and true models, and a complex wavefield imply significant complications with misfit and potential model improvement: the concept of misfit is well founded only for relatively similar waveforms, and required large model changes lead to nonlinear waveform changes.

Sources. In many cases, records of a very small number of local weak earthquakes are the only usable waveform data. Records of distant earthquakes are much more difficult to use because the computational domain would be much larger, and frequencies important for inversion in LSSS may be attenuated. Moreover, a sufficiently accurate regional model would be necessary.

Source–Receiver Configuration. Typically, the small numbers of recorded events and receivers give only tens of source–receiver pairs, in contrast with the exploration, regional, and global seismology with typically hundreds of sources and receivers with tens of thousands of source–

receiver pairs. Moreover, sources and receivers are spatially concentrated in a few groups. This is especially different from the exploration seismology in which relatively large numbers of sources and receivers are often placed at suitable positions.

Seismic Records. There are two distinct aspects: a relatively small number of available records and the complexity of waveforms due to interference and often a resonant nature of the EGM. Seismic phases in LSSS are not well separated. Absolute values of maximal frequencies are higher than in the regional and global scales and similar or lower than in exploration, but the ratio of a characteristic wavelength to the model dimension is, in fact, much larger in LSSS. Exploration records often miss the lower-frequency content. This makes it impossible to start multiscale inversion at sufficiently low frequencies and consequently contributes to the cycle-skipping problem.

Introductory Concepts and Notes

Misfit. In general, an adjoint tomography is based on an iterative minimization of the difference between calculated and recorded waveforms. The difference is quantified by a positive real-valued misfit. A zero value of a misfit means the identity of the waveforms according to a chosen criterion. The misfit is a function of waveforms, thus implicitly depending on a model. By summing all misfits calculated for a given event and for all events, we obtain event and aggregate misfits, respectively. We minimize the aggregate misfit, which guarantees the misfit convergence. Minimizing misfit calculated only for a subset of events can speed up the process (see, e.g., Warner *et al.*, 2013).

Kernel. A kernel (more accurately "a misfit sensitivity kernel") is defined as a volumetric density of a Fréchet derivative of the misfit with respect to material parameters; that is, it is a vector function of spatial coordinates. The kernel is, in general, computed as a time integral of a product of a regular field quantity and an adjoint field quantity. Particular formulas for kernel components depend on the choice of a model parameterization. Let λ and μ be Lamé elastic parameters and ρ be density. For the (ρ, λ, μ) parameterization, the kernel can be computed using the following equations:

$$K_\rho(\mathbf{r}) = -\int_T \partial_t \mathbf{u}^\dagger(\mathbf{r}, t) \cdot \partial_t \mathbf{u}(\mathbf{r}, t) dt, \qquad (1)$$

$$K_\lambda(\mathbf{r}) = \int_T \mathrm{tr}\boldsymbol{\varepsilon}^\dagger(\mathbf{r}, t) \mathrm{tr}\boldsymbol{\varepsilon}(\mathbf{r}, t) dt, \qquad (2)$$

$$K_\mu(\mathbf{r}) = \int_T 2\boldsymbol{\varepsilon}^\dagger(\mathbf{r}, t) : \boldsymbol{\varepsilon}(\mathbf{r}, t) dt, \qquad (3)$$

(see e.g., Fichtner, 2011), in which $\mathbf{u}$ is a displacement vector, $\dagger$ denotes adjoint field quantities, dot denotes the dot

product, $\mathrm{tr}\boldsymbol{\varepsilon}$ is a trace of matrix $\boldsymbol{\varepsilon}$, and a colon denotes a summation of products over all repeating indices. An overview of algorithms for efficient kernel computation can be found, for example, in Yang *et al.* (2016).

Model Update. After the kernel is preconditioned (see the Appendix), the model $\mathbf{M}$ is updated in the opposite kernel direction, the direction of the fastest descent of misfit, according to

$$\mathbf{M}_{i+1}(\mathbf{r}) = \mathbf{M}_i(\mathbf{r}) - \gamma_i^{\mathrm{opt}} \mathbf{K}_i(\mathbf{r}). \qquad (4)$$

To find the step length corresponding to the misfit minimum, that is, the optimal step length γ_i^{opt}, different trial step lengths γ_i^j are used. For each trial step length, a misfit is evaluated in a forward simulation. Based on the trial steps and corresponding misfits, γ_i^{opt} is determined. Using γ_i^{opt}, the model can be iteratively updated in each iteration i.

Multiscale Approach. Complicated seismograms cause the inversion often to end at some local misfit minimum close to the initial model. To overcome the problem of local minima, it is favorable to start inversion at the lowest meaningful frequencies and include higher frequencies only after the misfit at lower frequencies is sufficiently small (Bunks *et al.*, 1995). This is because the misfit (as a function of a model) at lower frequencies is relatively smooth, and the global minimum is more visible. Moreover, low-frequency computations are relatively efficient. Inversions using only high frequencies often lead to models that are far from a true structure (e.g., Gauthier *et al.*, 1986; Fichtner, 2011).

A used frequency range defines the scale of the computation. An inversion procedure using different subsequent scales is called a multiscale procedure. Examples of the multiscale approach can be found in Yuan and Simons (2014) and Yuan *et al.* (2015), whose approach is based on the decomposition of seismograms using sets of wavelets and in Bozdağ *et al.* (2016). Later in this article, we present an algorithm for determining frequency ranges suitable for inversions of LSSS.

Waveform Misfits

Because seismic phases are difficult or impossible to separate in LSSS, we cannot determine corresponding arrival-time misfits. It is necessary to use misfits based on the entire waveforms. In this article, we use the L2-norm misfit between waveforms in the process of inversion. The reason for L2 is that it involves information both on phase and amplitude. For evaluating results of inversion, we also use the envelope misfit (EM) and phase misfit (PM), as defined by Kristeková *et al.* (2006, 2009). EM and PM utilize information from the TF representation of a signal. An appropriate tool for obtaining the TF representation of transient signals is the continuous wavelet transform (CWT).

A choice of an appropriate waveform misfit determines the quality of the inversion (see, e.g., Modrak and Tromp, 2016).

However, different models may lead to the same value of misfit, the problem of nonuniqueness (see, e.g., Cercato, 2009).

Mask

We assume that we know material parameters of the bedrock. However, the kernel can also be nonzero in the bedrock. Considering the whole kernel would lead to model updates, even in parts of the model in which the inversion problem is ill-conditioned (mostly in deep parts). Large CPU time requirements would be an additional problem. It is not feasible to simulate full waveform propagation from source to receivers for all iterations. Computational load can be reduced using the so-called excitation box (e.g., Moczo et al., 2014) around LSSS. The total wavefield is simulated only within the excitation box. This means that the kernel can be computed only inside the excitation box.

Without further adjustments, the excitation box, by artificially cutting the kernel at the box sides, introduces an artificial discontinuity in the inverted model. Therefore, we need to apply a spatial function, a mask, making the kernel smoothly vanishing at the excitation box sides.

We may, however, choose a mask that restricts the inversion to a certain part of the model inside the excitation box. In an ideal case, the mask could have a shape of LSSS. Because this is usually *a priori* unknown, it is reasonable to use a simple analytical mask nulling kernels only near the excitation box sides. The use of a wrong mask (e.g., cutting off a piece of LSSS) would lead to error in a model which cannot be removed by the inversion process.

Misfit Minimization

Optimal-Step Selection. The parabolic-approximation algorithm for finding the optimal step is widely used in the exploration, regional, and global inversions. In the case of LSSS, the situation is more complicated because the initial model is often very different from the true LSSS. Consequently, the misfits are often highly nonparabolic functions of step length. It is important to use a robust and efficient algorithm that will find the approximation of the optimal step with a minimal number of (computationally expensive) trials.

One-Parameter Inversion. The numerical simulations show that the wavefield is much more sensitive to parameter μ than to λ, due to dominant S and surface waves in records. This means that an inverted model can be considerably less accurate in λ compared to μ. In a typical situation in LSSS, there may not be enough data to successfully constrain λ-values. On the other hand, because of relative insensitivity, inaccuracy in λ does not pose a problem. Therefore, we invert only μ and calculate λ analytically from an approximate empirical relation (e.g., Brocher, 2005). In this case, a one-component kernel $\tilde{K}_\mu$ equals

$$\tilde{K}_\mu = K_\mu + K_\lambda \frac{d}{d\mu} f_\lambda(\mu), \qquad (5)$$

in which $f_\lambda(\mu)$ is the chosen relation and K_λ and K_μ are defined according to equations (2) and (3), respectively.

Relationships between V_P, V_S, and ρ have been used by, for example, by Marone et al. (2007), Kustowski et al. (2008), and Nettles and Dziewonski (2008). The approach has several advantages, the computation of only the μ component of the kernel and a reduction of the number of free parameters in the inversion process. If needed, a separate λ inversion can be added after a reasonable model of μ is obtained. See Prieux et al. (2013a,b), Vigh, Cheng, et al. (2014), and Vigh, Jiao, et al. (2014) for a similar approach. For a joint multiparameter inversion, see, for example, Operto et al. (2013), Métivier et al. (2015), and Blom et al. (2017).

Adaptive Multiscale Approach

Extending Frequency Ranges. After the inverted model sufficiently explains lower-frequency data, it is possible to extend the used range by adding higher frequencies or by increasing both the minimal and maximal frequencies. Numerical tests suggest the first possibility that keeps the low frequencies. This can be explained: If the model changes due to a misfit at higher frequencies, the model change has an effect on the misfit at all frequencies. An absence of the lower frequencies in further updates of the model could deteriorate the waveform misfit at those frequencies.

Determination of Frequency Ranges. By adding higher frequencies, we are increasing energy and thus considerably changing the wavefield. After extending the frequency range, the inversion continues until the waveform misfit is small enough. Then, we can further extend the frequency range. We may call a set of iterations for a frequency range a scale cycle. Although extending the frequency range very slowly would be desirable, it would lead to many scale cycles and consequently to huge CPU cost. Therefore, some compromise is necessary.

We determine frequency ranges based on energy distribution in the available records. The CWT (e.g., Daubechies, 1992) of signal $y(t)$ is defined as

$$\text{CWT}\{y\}(t, a) = \frac{1}{\sqrt{|a|}} \int_T y(\tau) \psi * \left(\frac{t - \tau}{a}\right) d\tau, \qquad (6)$$

in which a is the scale parameter of CWT, ψ is the analyzing wavelet, and $*$ denotes the complex conjugation. To obtain a TF representation, the central frequency f of the wavelet is assigned to the scale parameter, an approximate relation being $f \approx f_{W1}/a$, in which f_{W1} is the central frequency of the wavelet for the scale parameter $a = 1$. We define energy in the seismogram component i from source s at receiver r, say $u_i^{sr}(t)$, as

$$E_i^{sr} = \int_T \int_F |\text{CWT}\{u_i^{sr}\}(t, f)|^2 df\, dt. \qquad (7)$$

An aggregate energy E_a^n up to frequency f_n is a sum of energies in all seismograms and all components $E_a^n = \sum_{i,s,r} E_i^{sr}(f_n)$, filtered by a low-pass filter with the cutoff frequency f_n. A sequence of the cutoff frequencies f_n for an energy increase ratio e_{RI} is defined by condition

$$f_n : E_a^{n+1} = e_{RI}^{-1} E_a^n; \qquad f_n = f_0, ..., f_N. \qquad (8)$$

Here, f_N and f_0 are the smallest and highest cutoff frequencies.

The determination of the frequency ranges based on energy increase may be referred to as an adaptive multiscale approach.

Full Waveform Adjoint Inversion in a Local Surface Sedimentary Structure: Procedure

Selection of a Frequency Range of Interest

There is a natural upper limit given by the frequency content of available seismic records. Usually, a more restricting upper limit is imposed by computational costs. The lower-frequency limit is imposed by the size of the computational domain. The results of inversions are not very sensitive to the choice of the upper bound of the lowest-frequency range f_N, as long as the frequency is sufficiently small; that is, the aggregate misfit for the starting model is insignificant in this frequency range.

Recommended Parameter Ranges

Several inversion parameters can be reasonably predetermined. Some others are difficult to determine because the correct (optimal) values may differ case to case:

• sequence of frequency ranges (scales),
• the number of iterations at one scale,
• the mask,
• smoothing parameters.

The selection of scales is determined by the relative amount of energy according to equation (8). The subsequent scale (frequency range) has higher frequencies and more energy. Numerical tests have shown that the increase in energy should be between 15% and 40%. Less than 15% does not lead to reasonable improvement in the inversion, and, moreover, considerably increases the computational load. For an energy increase larger than 40%, the wavefield changed too much, and there was often a problem with local minima. f_N in our inversions was around one-tenth of the first peak frequency determined from the records.

The number of iterations at a scale is highly variable. It often takes about 25 iterations until the model reaches the minimum at the particular scale, and the additional model changes are negligible. Sometimes, however, no significant change can be made to the model to decrease misfit.

The mask must be at least as deep as the LSSS, for which depth is *a priori* often unknown. However, our numerical tests show that our method leads to satisfying results for a wide range of estimated depths different from the real one. Therefore, the precise depth does not need to be known.

We obtained the best results by smoothing the kernel with a characteristic smoothing length approximately one-tenth of the shortest expected wavelength.

Inversion Scenarios

In the Recommended Parameter Ranges section, we recommended only parameter ranges for several inversion parameters, not specific exact values. This is because, for those parameters it is not possible to find single values universally suitable for all types of LSSSs. Additionally, numerical tests have shown that for particular structures there has been significant variability of the final misfits for a set of inversions differing only by a small change in the values of the inversion parameters. Therefore, instead of trying to find the best set of values of the inversion parameters (i.e., one best value of each inversion parameter), it is more efficient to try multiple different inversions with different sets of parameter values.

Formally, a complete multiscale inversion using one set of values of the inversion parameters will be called a scenario. Different scenarios can be compared using the aggregate misfits. The inverted model from the scenario with the lowest misfit can be selected as the best inverted model. If no further procedure is applied, the best inverted model is considered a final inverted model.

Because the completion of all scenarios would be computationally very demanding, it is practically necessary at some moment to select the best scenario. This leads us to the idea of termination of inaccurate and inefficient scenarios in the course of inversions. We start a multiscale procedure at the lowest frequencies with multiple scenarios. Computational cost of the individual forward computation rises with increasing frequencies, that is, when we move to a lower scale. Therefore, we eliminate scenarios with large values of the aggregate L2 misfit to keep computational costs of the whole set of scenarios at the new scale similar to that at the previous scale. The inversion is completed (at the highest frequencies) only for the best scenarios.

Repetitive Multiscale Inversion

After all scenarios finish, a set of models is obtained. Some of them are much better (leading to a much lower aggregate misfit) than the starting model. It is reasonable to use the best inverted model(s) as a starting model for another set of scenarios. This procedure may be repeated several times, as long as the misfit reasonably decreases. We call this procedure the repetitive multiscale inversion. With better starting models, we may expect even better resulting models. The best model after the last repetition is called the final inverted model. The inversion procedure is illustrated in Figure 1.
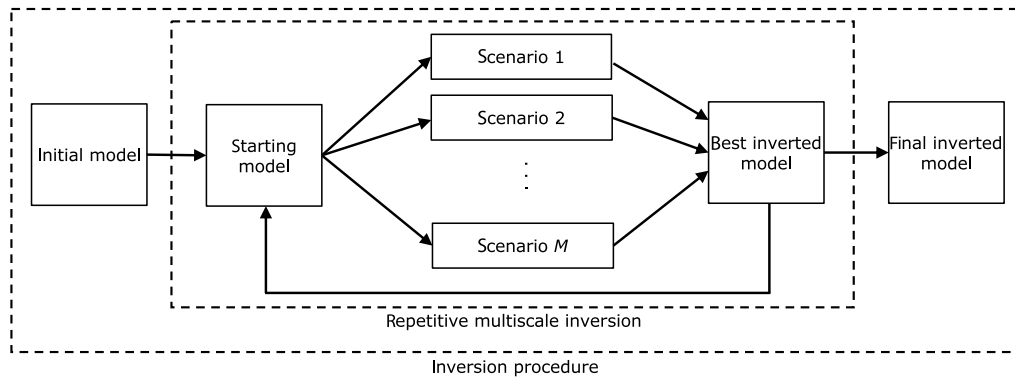
Figure 1. Workflow of inversion procedure. Note that each scenario is an independent full waveform multiscale inversion for a given set of inversion parameters.

Numerical Example: A 2D *P–SV* Blind Test

Here, we present a blind test. A third party was asked to choose a model of LSSS and source–receiver configuration and to numerically simulate synthetic records. We obtained the records, the material parameters of the bedrock, the positions of eight receivers, and a full description of two-point double-couple sources. Such a limited amount of data well represents a possible situation with a site of interest (considering the dimensionality of the problem). Using the data, we apply our inversion procedure to find a model of LSSS. Eventually, we verify the quality of the inverted model up to the target 4.5 Hz frequency, using a direct comparison of seismograms, waveform misfits, waveform goodness of fit (GOF), and mainly GOF for selected EGM characteristics.

Here, we should point out two important aspects. We have to clearly distinguish what the criterion of the quality of the inverted model is and what the goal of the inversion is. The most important criterion of the quality of the inverted

model is a reasonable level of GOF between EGM characteristics determined from the records and characteristics determined from synthetics for the inverted model. Why? Clearly, we would not mind to obtain a model which would be sufficiently close to the true structure, in terms of structural fidelity. However, data available for the inversion are not sufficient for such a straight, rigorous criterion. Then, the question is whether the inversion should be dismissed or not. We think it should not be because the main goal or purpose of the inversion is the possibility of predicting EGM at any site atop the underground local structure for different sources. For this, it is reasonable to consider the quality of the inverted model in terms of EGM characteristics.

We apply the second-order accurate in time and fourth-order accurate in the space displacement–velocity–stress staggered-grid finite-difference scheme. The methodology is presented in the book by Moczo *et al.* (2014). The method is sufficiently accurate and computationally efficient for the inversion.

Note on the color convention: Where applicable, we use the following color convention: a black color relates to the true model, blue to the initial homogeneous model, green to the best inverted model after the first multiscale inversion, that is, the intermediate inverted model, and red to the final inverted model.

An overview of the inversion procedure and algorithm, the technical details on the implementation of the inversion procedure, and evolution of the medium parameters in the blind test are available in the Ⓔ electronic supplement to this article.

Waveform Data

Waveform data consist of 16 records (synthetics simulated by the third party) in the frequency range up to 4.5 Hz. The records are shown in Figure 2.
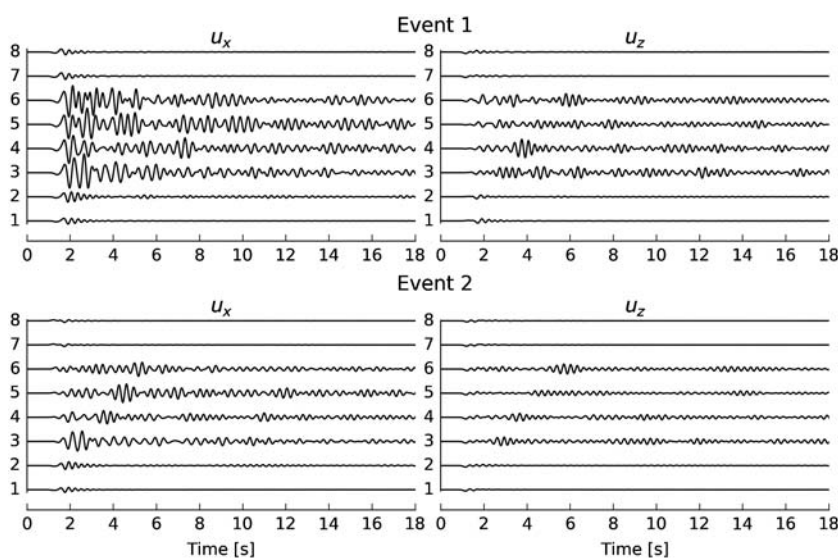

Figure 2. Synthetic records at eight receivers due to two sources. u_x and u_z indicate the horizontal and vertical components of the displacement vector. (The horizontal x axis is positively oriented from left to right; the vertical z axis is positively oriented downward.)
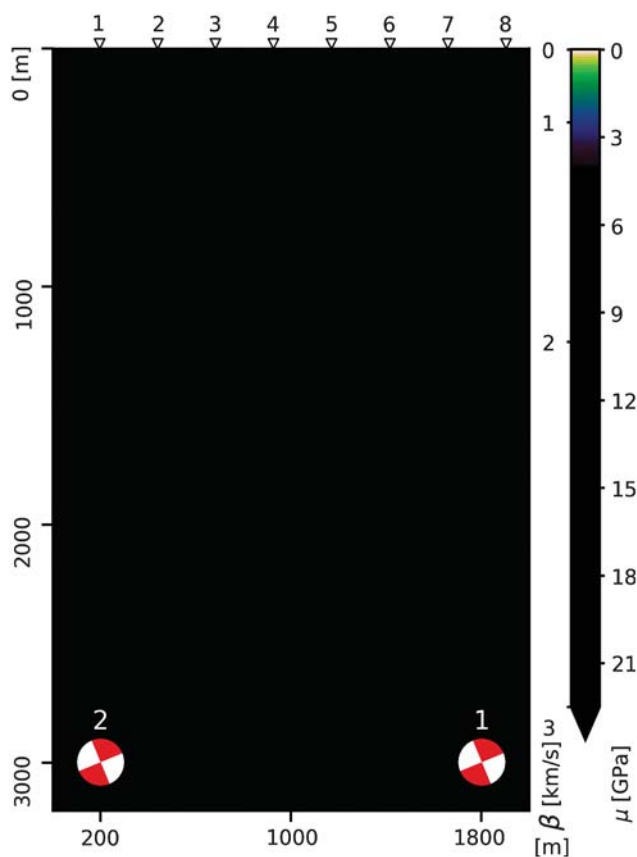
Figure 3. Initial homogeneous model and the source–receiver configuration. The model is illustrated using S wavespeed β and shear modulus μ. Their values in the initial models are 3000 m/s and 22.5 GPa. Two sources are indicated by the focal mechanism plots, eight receivers at the free surface by white triangles.
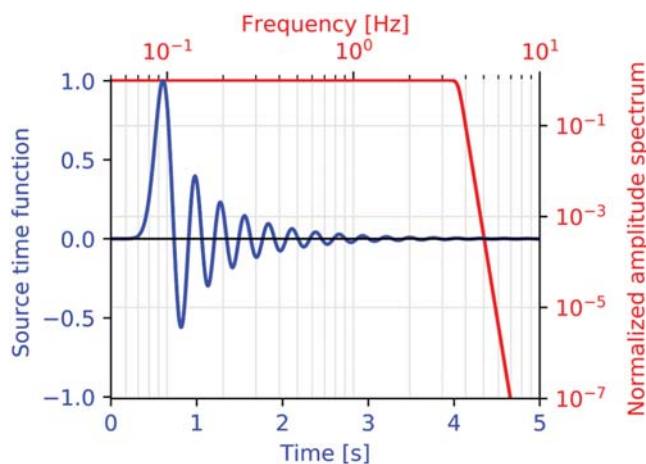

Figure 4. Normalized source-time function and its normalized amplitude spectrum.

Seismograms for Inverted Models

The best inverted model after the first multiscale inversion was already very good in terms of the misfit decrease. We denote this model as intermediate. We continued with more repetitions of the inversions. Each of them improved the model, but the rate of improvement was decreasing. The inversion process was stopped after seven multiscale inversions. We call the obtained model the final inverted model.

Figures 6 and 7 compare the seismograms for the intermediate and final inverted models, respectively, with the records in the frequency range up to 4.5 Hz. The visual impression from the comparison is surprisingly good. The level of agreement, however, has to be quantified. This is possible, for example, using the envelope and phase GOF based on the TF decomposition of the records and seismograms simulated for the inverted models (Kristeková *et al.*,
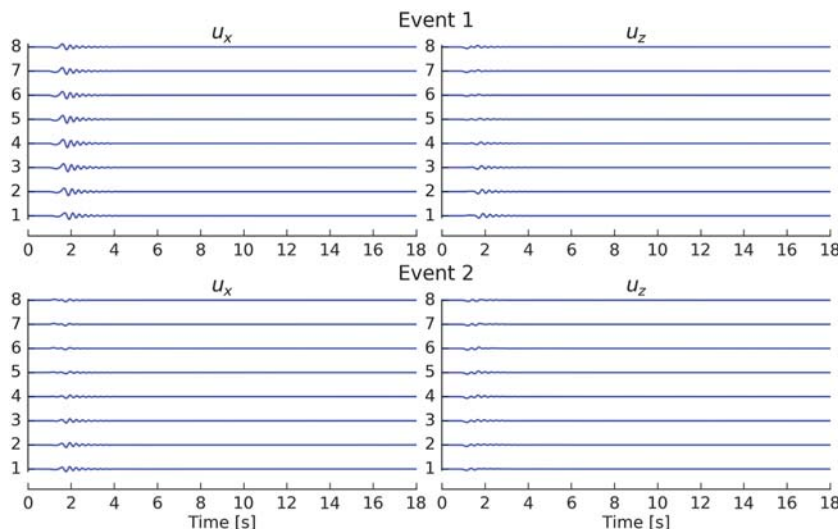
Source–Receiver Configuration, Initial Model, and Initial Waveforms

Given the values of material parameters in bedrock as the only known parameters, we assume the corresponding homogeneous half-space as the initial model. Obviously, no simpler initial model can be assumed. The model is indicated in Figure 3. The figure also shows the source–receiver configurations for two considered events. Sources, both at 3 km depth, are indicated by focal mechanism plots labeled 1 and 2. The uniformly distributed receivers at the free surface are indicated by triangles labeled 1–8. Figure 4 shows the normalized source time function, used for both sources, and its normalized amplitude spectrum. Synthetic seismograms for the initial model are shown in Figure 5.

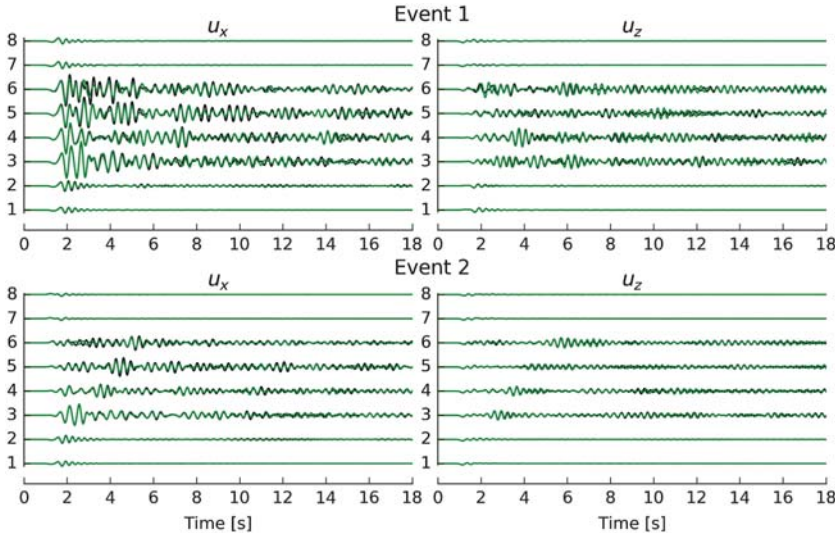**Figure 5.** Synthetic seismograms for the initial model.

Figure 6. Seismograms at receivers 1–8 simulated for the intermediate inverted model (in green) compared with records (in black). Event 1, seismograms simulated for source 1; event 2, seismograms simulated for source 2.
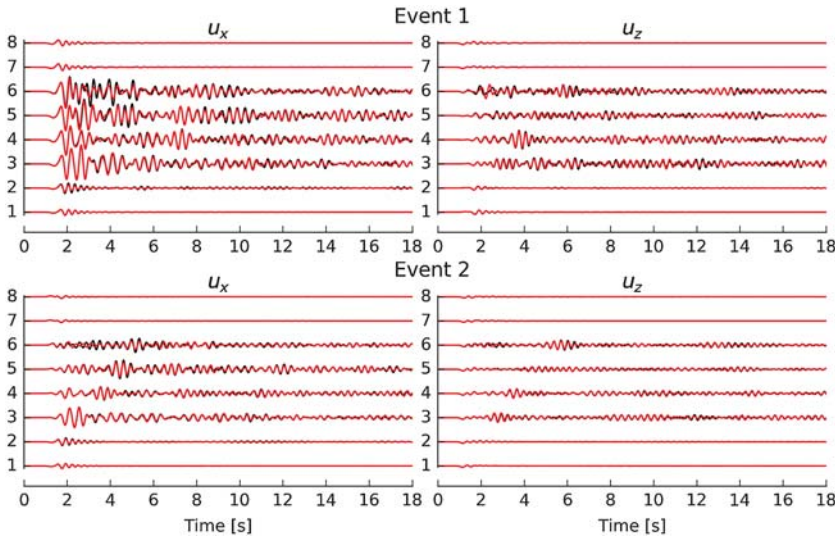

Figure 7. The same as in Figure 6 but for the seismograms simulated for the final inverted model (in red).

2009). The envelope GOF (EG) and phase GOF (PG) are single-valued EM and PM mapped to interval 0–10

$$EG = 10\exp(-|EM|), \qquad (9)$$

$$PG = 10(1 - |PM|), \qquad (10)$$

and thus, unlike EM, they have upper limits.

The TF GOFs are shown in Figure 8 for event 1 (the GOFs for event 2 are similar). Following Anderson (2004) and Kristeková et al. (2009), a score below 4 is a poor fit, a score of 4–6 is a fair fit, a score of 6–8 is a good fit, and a score over 8 is an excellent fit. We can see that there is no

lower fit than fair. It appears only in the GOFs for receiver 6 and is indicated by a solid red line. As expected, the GOF is higher in the phase than in the envelope. Let us note that we show the GOFs based on the misfits normalized with respect to the maximum envelope value at a receiver. Consequently, GOFs well visualize those parts of the seismogram which mostly contribute to the waveform misfit at the receiver. We also show in Figure 8 the single-valued envelope and phase GOFs, EG, and PG, for each TF GOF. The background color indicates the verbal value of fit: green excellent and yellow good. Values of EG and PG may be different from the visually dominant value of TF GOF. This is because they are primarily determined as averages of the point TF EM/PM values weighted locally by the TF envelope of the reference seismogram. Thus, they are less influenced by the parts of the TF plane with a small TF envelope.

Verification in Terms of Misfits

Given the fact that the inverted models were obtained using two sources and eight receivers, it is reasonable to verify the models using alternative sources and receivers. They are all shown in Figure 9. For each of the nine sources (with the same source time function as used for the two sources in the blind test; see Fig. 4) the third-party numerically simulated records at all 16 receivers for the true model. We calculated synthetics for the inverted model (we did not invert the model for the alternative sources and receivers). Then, we evaluated the event misfit for each of the nine sources (events). Figure 10 shows the event misfits between the true and inverted models. The event EMs and PMs χ_{EM}^s and χ_{PM}^s are computed as

$$\chi_M^s = \sqrt{\frac{\sum\limits_{i,r} (M_i^{sr})^2 N_G^{sr}}{\sum\limits_{i,r} N_G^{sr}}}, \qquad (11)$$

in which M_i^{sr} is either of the single-valued EM and PM, i indicates component, s source, r receiver, and $N_G^{sr} = \max_i\{E_i^{sr}\}$ is a global norm (recall that E_i^{sr} is an energy of a record; see relation 7). It is obvious that there is a significant difference between the blue and green event misfits, indicating the considerable effect of the first inversion. The misfits for the additional sources and receivers are
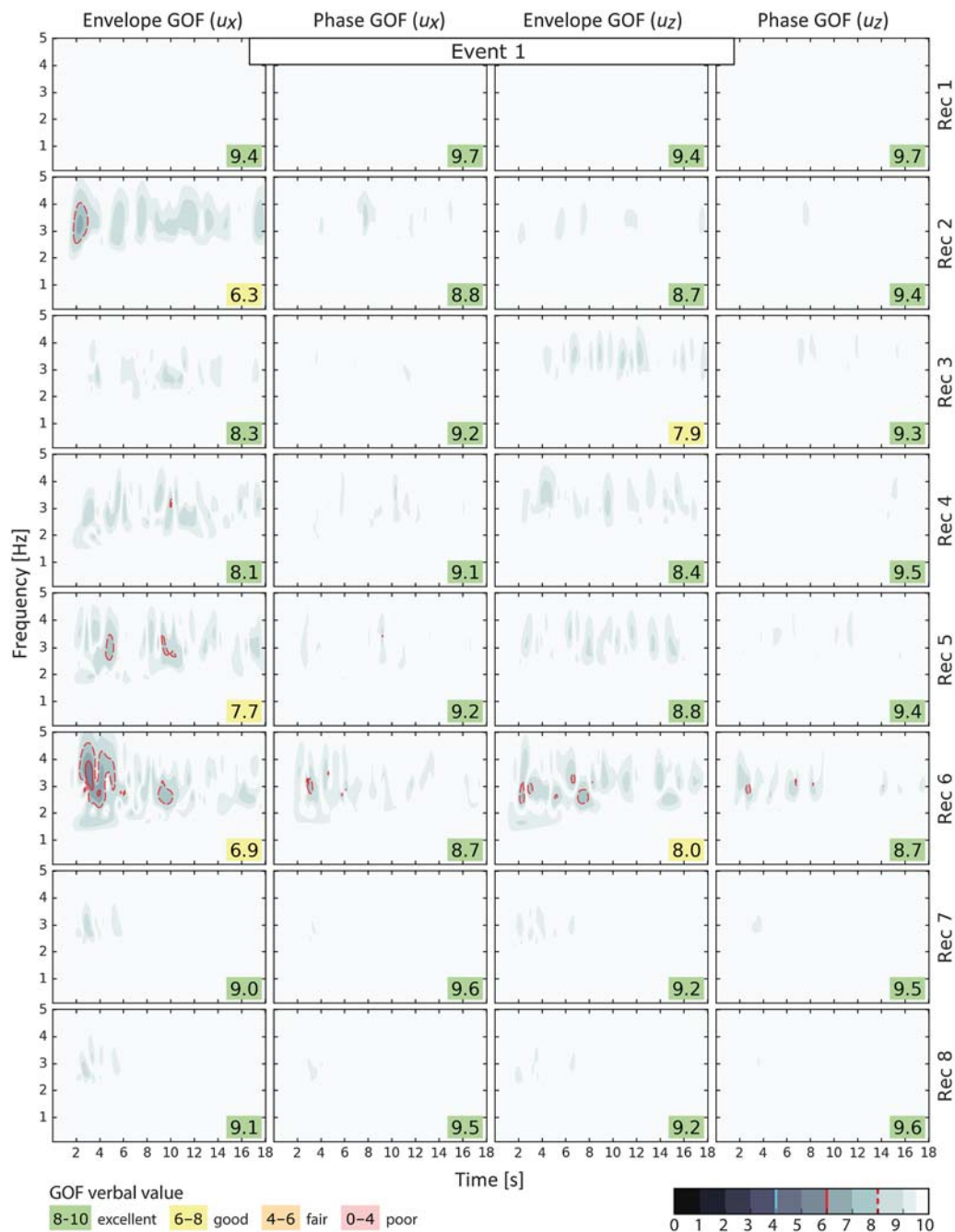
Figure 8. Time–frequency envelope and phase goodness of fit (GOF) between the records and seismograms for the final inverted model for event 1. The corresponding single-valued GOFs are also shown (in the lower right corners). The background color indicates the verbal value of fit: green excellent and yellow good.

very close to those for the sources and receivers used in the inversion. This is significant. The final inverted model is equally accurate for simulating seismic motion for any source in the considered domain.

Verification in Terms of Earthquake Ground-Motion Characteristics

Here, we compare the synthetic seismograms for the initial, intermediate inverted, and final inverted models with

records in terms of the following characteristics of EGM: peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD), spectrum intensity (SI), cumulative absolute velocity (CAV), Arias intensity (I_A), and root mean square acceleration (a_{rms}). Let c_0 be a value of a ground-motion characteristic for a given component of the record at a given receiver and for a given source. Let c be an analogous value for a seismogram to be compared with the record. Then, the corresponding GOF_c may be defined as
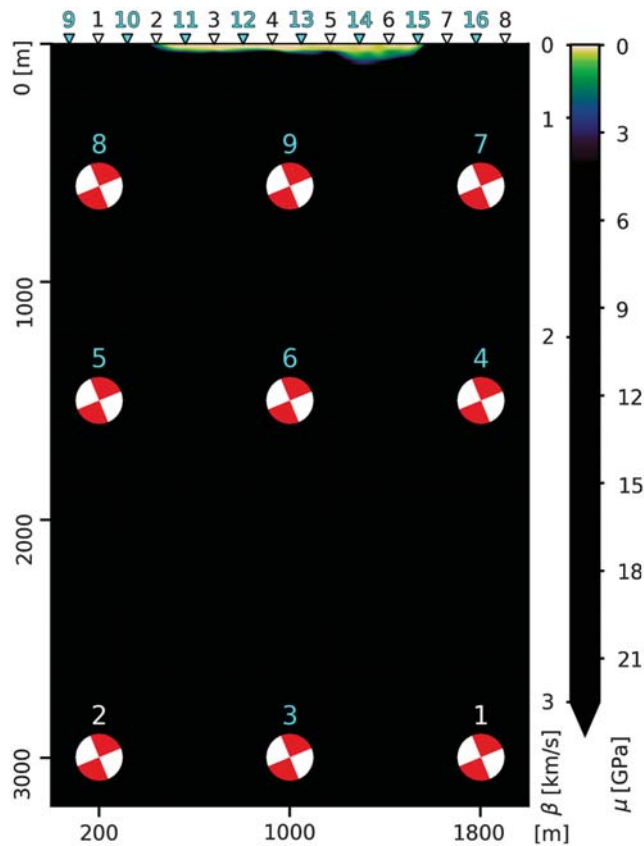
Figure 9. Configuration of sources (focal mechanism plots) and receivers (triangles). White-numbered sources and black-numbered receivers are those used in inversion. Cyan-numbered sources and receivers denote additional sources and receivers used only for verification.
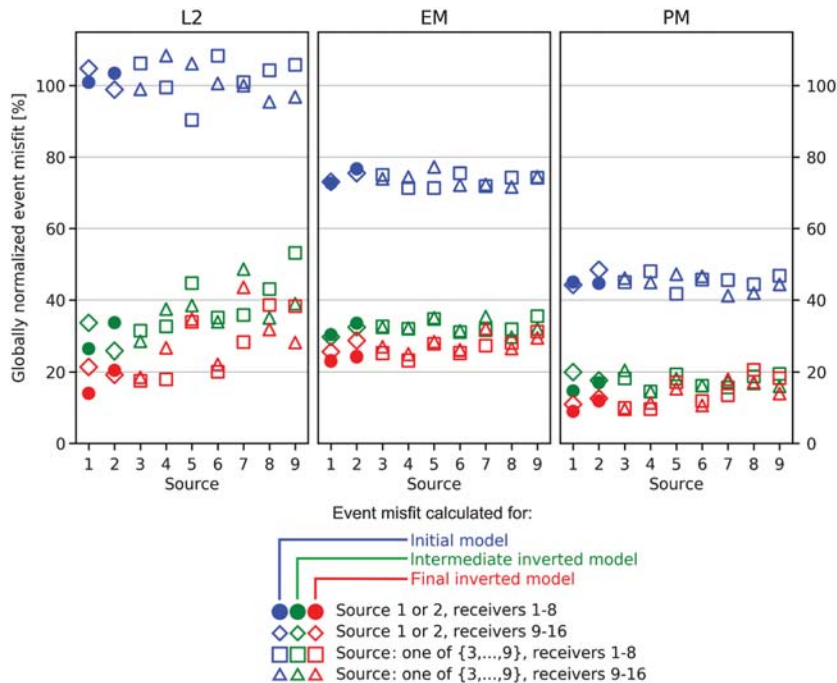

Figure 10. Globally normalized event misfits for the nine considered sources. Recall that only sources 1 and 2 and receivers 1–8 were used in the inversion.

$$\mathrm{GOF}_c = 10\exp\left[-\left(\frac{c-c_0}{c_0}\right)^2\right]. \qquad (12)$$

GOF_c is displayed in Figure 11. We can see that there is a significant difference between GOFs corresponding to the initial model and GOFs corresponding to the intermediate and final inverted models for any of the seven ground-motion characteristics. The arithmetic averages of GOFs corresponding to the intermediate and final inverted models represent the excellent level of agreement.

GOFs for PGA and CAV as a function of receiver position are shown in Figures 12 and 13, respectively. In both cases, an increase of agreement in the intermediate and final inverted models compared to that in the initial model is obvious and significant. Except for the GOF for the horizontal component at receiver 15, the level of agreement is excellent in the intermediate and final inverted models. The agreement is lower at receivers atop the deepest sediments, and the lowest level of agreement (i.e., good) at receiver 15 is related to the position close to the edge of sediments (see Fig. 14). This is a consequence of having the data from only one receiver (i.e., receiver 6) atop the deepest sediments used for the inversion.

Model Comparison

For direct comparison of the intermediate and final inverted models with the true model, the third party provided the true model in the final stage of the study. The comparison of the initial, intermediate inverted, final inverted, and true models is shown in Figure 14. Both inverted models well captured the horizontal extent of the sediments. The shape of the shallow sediments is also relatively well reproduced in the inverted models. The deepest part of sediments is missing in the intermediate inverted model. It is partially captured in the final inverted model: it is shallower (the deepest point being at 90 m) compared to the true model (151 m) by ~40%, and the velocity contrast is smaller. What is the reason for this inaccuracy? Only one (receiver 6) from the eight receivers used for the inversion is located atop the local deep basin. Intuitively, this is insufficient coverage of that part of the sedimentary structure. There is no reason to think that the inversion methodology might be responsible. In the course of developing the methodology, we obtained much better results for deep basins if we used a sufficient number of receivers atop the basin.

Computational Costs

A nonparallelized computational code was used for a multiscale inversion. The computational time for one scenario was
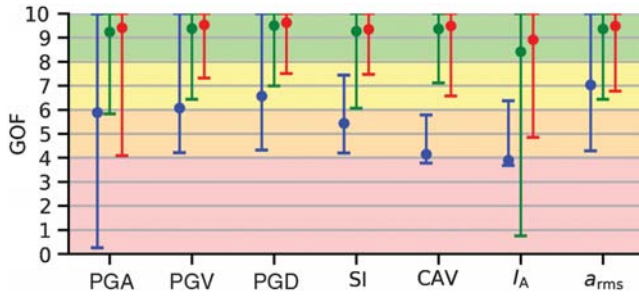
Figure 11. The scatter of the 162 GOF_{PGA} (PGA, peak ground acceleration) values evaluated individually for two components at receivers 3–6 and 11–15 (receivers atop sediments) for 9 sources is shown using the blue, green, and red vertical bars for seismograms simulated for the initial, intermediate inverted, and final inverted models, respectively. Full circles represent the arithmetic averages. GOFs for the other characteristics are shown in the same manner.
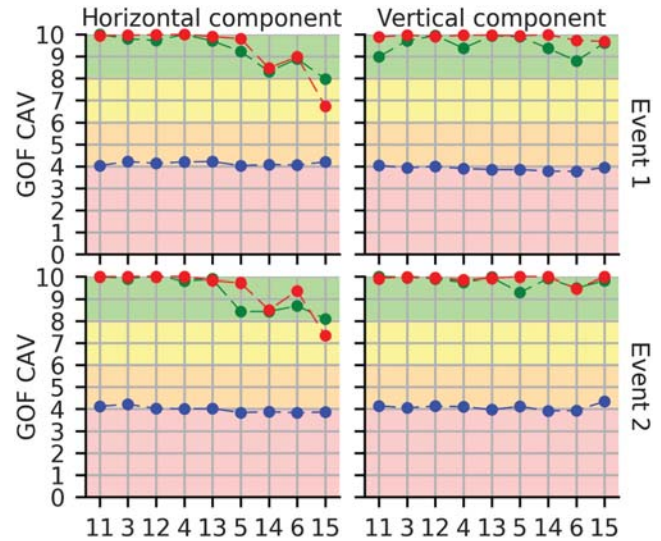

Figure 13. The same as in Figure 12 but for cumulative absolute velocity (CAV).
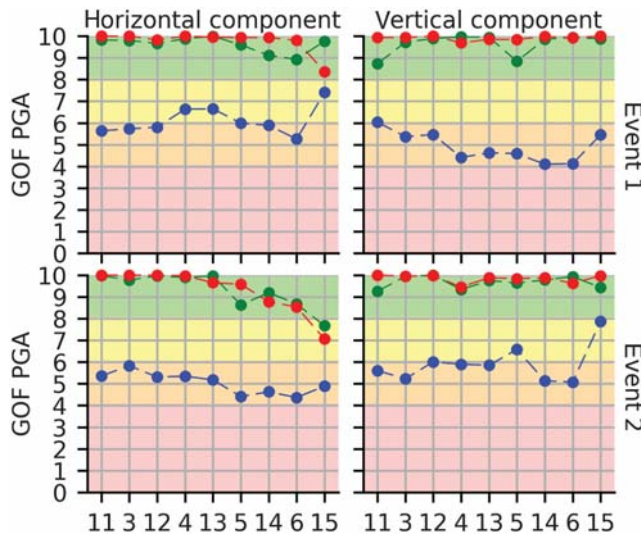

Figure 12. GOFs for PGA evaluated individually for two components at receivers 3–6 and 11–15 for the sources 1 and 2 are shown in the blue, green, and red for seismograms simulated for the initial model, the intermediate inverted model, and the final inverted model, respectively.

~4 days for the energy increase ratio $e_{RI} = 1.3$. Multiple scenarios ran simultaneously on a computer with 64 cores (AMD Opteron 2.2 GHz). The intermediate inverted model was obtained after the first multiscale inversion calculated for 64 scenarios. The final inverted model was obtained after 12 multiscale inversions (~42 days) with, on average, 60 scenarios per one multiscale inversion.

Conclusions

Based on extensive numerical modeling and testing, we developed a procedure for adjoint tomography for 2D LSSS. The procedure comprises:

- L2-norm misfit providing better results than a phase or envelope or travel-time misfit;
- a spatially dependent normalization by maximal absolute values and anisotropic smoothing;
- a spatial mask for restricting a region of inversion;
- a shear modulus (μ) as an inversion parameter. Modulus λ is calculated from the chosen empirical relation between μ and λ;
- minimization of the sum of all waveform misfits at all receivers;
- a robust algorithm for finding the optimal step length instead of the parabolic approximation;
- an adaptive multiscale approach in which a subsequent frequency range is an extension of the current frequency range toward higher frequencies. Energies (in recorded seismograms) in the frequency ranges are logarithmically equidistantly distributed;
- a set of scenarios. We call a complete multiscale inversion using one set of values of inversion parameters a scenario. Because the best set of values cannot be determined at the beginning of the inversion process, it is necessary to try a set of different scenarios;
- repetitive multiscale inversion. The best inverted model from all scenarios is used as a starting model for another set of scenarios.

We numerically demonstrate the developed procedure by presenting a verification blind test. A third party provided (a) seismograms numerically simulated for an undisclosed true structure, (b) source parameters, and (c) material parameters of the bedrock. As the initial model, we assumed the simplest possible model, a homogeneous half-space with the parameters of the bedrock. We evaluated the quality of the obtained inverted models using a direct comparison of seismograms, waveform misfits (L2, EMs, and PMs),
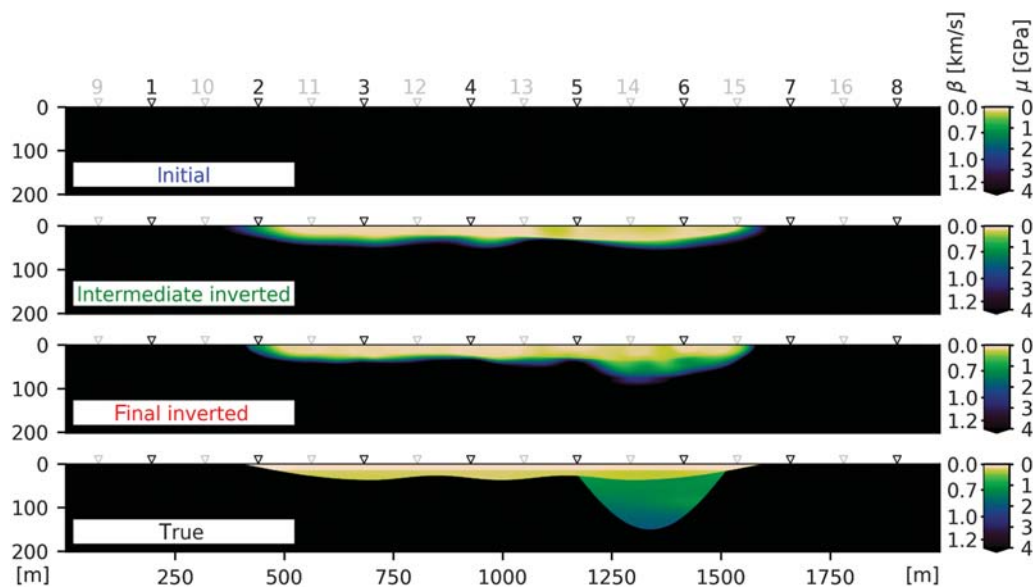
Figure 14. Spatial distribution of S wavespeed β and shear modulus μ in the upper 200 m of the initial, intermediate inverted, final inverted, and true models. Note that the S wavespeed in the homogeneous half-space, and thus in the bedrock, in the true and inverted models is 3000 m/s, and the minimum S wavespeed in sediments in the inverted models and true models is 250 and 160 m/s, respectively. Receiver positions that were not used for inversion are indicated by the gray color.

waveform GOF (TF envelope, TF phase, and single-valued envelope and single-valued phase), and mainly the GOF for selected EGM characteristics (PGA, PGV, PGD, SI, CAV, I_A, and a_{rms}). We also verified the inverted models for other source–receiver configurations not used in the inversion.

We developed and verified the procedure for 2D structures because development including extensive numerical modeling and testing for 3D would be computationally too heavy. We assume that the procedure can be, in principle, applied to 3D structures after some refinements, due to the 3D spatial distribution of sources and receivers.

A 3D blind test and subsequent application to real data will be addressed in the follow-up study.

Data and Resources

The synthetic data for inversions were provided by Andrej Cipciar. For inversion, we developed our own code. Neither the data nor the code is publicly available.

Acknowledgments

References

Anderson, J. G. (2004). Quantitative measure of the goodness-of-fit of synthetic seismograms, *13th World Conf. on Earthquake Engineering*, Vancouver, Canada, Paper 243, available at http://www.iitk.ac.in/nicee/wcee/article/13_243.pdf (last accessed April 2018).

Blom, N., Ch. Boehm, and A. Fichtner (2017). Synthetic inversions for density using seismic and gravity data, *Geophys. J. Int.* **209,** no. 2, 1204–1220.

Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire (2016). Global adjoint tomography: First-generation model, *Geophys. J. Int.* **207,** no. 3, 1739–1766.

Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. Seismol. Soc. Am.* **95,** 2081–2092.

Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent (1995). Multiscale seismic waveform inversion, *Geophysics* **60,** no. 5, 1457–1473.

Cercato, M. (2009). Addressing non-uniqueness in linearized multichannel surface wave inversion, *Geophys. Prospect.* **57,** 27–47, doi: 10.1111/j.1365-2478.2007.00719.x.

Chaljub, E., E. Maufroy, P. Moczo, J. Kristek, F. Hollender, P.-Y. Bard, E. Priolo, P. Klin, F. de Martin, Z. Zhang, *et al.* (2015). 3-D numerical simulations of earthquake ground motion in sedimentary basins: Testing accuracy through stringent models, *Geophys. J. Int.* **201,** 90–111, doi: 10.1093/gji/ggu472.

Chaljub, E., P. Moczo, S. Tsuno, P.-Y. Bard, J. Kristek, M. Kaser, M. Stupazzini, and M. Kristekova (2010). Quantitative comparison of four numerical predictions of 3D ground motion in the Grenoble valley, France, *Bull. Seismol. Soc. Am.* **100,** no. 4, 1427–1455.

Chen, M., F. Niu, Q. Liu, J. Tromp, and X. Zheng (2015). Multi-parameter adjoint tomography of the crust and upper mantle beneath East Asia: Part I: Model construction and comparisons, *J. Geophys. Res.* **120,** no. 3, 1762–1786.

Chen, P., and E.-J. Lee (2015). *Full-3D Seismic Waveform Inversion*, Springer, Cham, Germany.

Choi, Y., and C. Shin (2008). Frequency-domain elastic full waveform inversion using the new pseudo-Hessian matrix: Experience of elastic Marmousi 2 synthetic data, *Bull. Seismol. Soc. Am.* **98,** no. 5, 2402–2415.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pennsylvania.

Fichtner, A. (2011). *Full Seismic Waveform Modelling and Inversion*, Springer, Berlin/Heidelberg, Germany.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. Villasenor (2013). Multiscale full waveform inversion, *Geophys. J. Int.* **194,** 534–556.

Gauthier, O., J. Virieux, and A. Tarantola (1986). Two-dimensional nonlinear inversion of seismic waveforms: Numerical results, *Geophysics* **51,** no. 7, 1387–1403.

Hosseini, M., and S. Pezeshk (2015). A synthetic study into the nature and solution of nonuniqueness in surface-wave inverse problems, *Bull. Seismol. Soc. Am.* **105,** no. 6, 3167–3179, doi: 10.1785/0120150081.

Igel, H., H. Djikpéssé, and A. Tarantola (1996). Waveform inversion of marine reflection seismograms for *P* impedance and Poisson's ratio, *Geophys. J. Int.* **124,** no. 2, 363–371.

Kristeková, M., J. Kristek, and P. Moczo (2009). Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophys. J. Int.* **178,** 813–825, doi: 10.1111/j.1365-246X.2009.04177.x.

Kristeková, M., J. Kristek, P. Moczo, and S. Day (2006). Misfit criteria for quantitative comparison of seismograms, *Bull. Seismol. Soc. Am.* **96,** no. 5, 1836–1850.

Kustowski, B., G. Ekström, and A. Dziewonski (2008). Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *J. Geophys. Res.* **113,** no. B06306, doi: 10.1029/2007JB005169.

Lee, E.-J., and P. Chen (2016). Improved basin structures in Southern California obtained through full-3D seismic waveform tomography (F3DT), *Seismol. Res. Lett.* **87,** no. 4, 874–881.

Lee, E.-J., P. Chen, T. H. Jordan, P. B. Maechling, M. A. M. Denolle, and G. C. Beroza (2014). Full-3-D tomography for crustal structure in southern California based on the scattering-integral and the adjoint-wavefield methods, *J. Geophys. Res.* **119,** 6421–6451.

Liu, Q., and Y. Gu (2012). Seismic imaging: From classical to adjoint tomography, *Tectonophysics* **566/567,** 31–66, doi: 10.1016/j.tecto.2012.07.006.

Luo, Y., R. Modrak, and J. Tromp (2013). Strategies in adjoint tomography, in *Handbook of Geomathematics*, W. Freeden, M. Z. Nashed, and T. Sonar (Editors), Springer, Heidelberg, Germany, 1943–2001.

Marone, F., Y. Gung, and B. Romanowicz (2007). Three-dimensional radial anisotropic structure of the North American upper mantle from inversion of surface waveform data, *Geophys. J. Int.* **171,** 206–222.

Maufroy, E., E. Chaljub, F. Hollender, P.-Y. Bard, J. Kristek, P. Moczo, F. De Martin, M. Theodoulidis, M. Manakou, C. Guyonnet-Benaize, *et al.* (2016). 3D numerical simulation and ground motion prediction: Verification, validation and beyond—Lessons from the E2VP project, *Soil Dynam. Earthq. Eng.* **91,** 53–71.

Maufroy, E., E. Chaljub, F. Hollender, J. Kristek, P. Moczo, P. Klin, E. Priolo, A. Iwaki, T. Iwata, V. Etienne, *et al.* (2015). Earthquake ground motion in the Mygdonian basin, Greece: The E2VP verification and validation of 3D numerical simulation up to 4 Hz, *Bull. Seismol. Soc. Am.* **105,** no. 3, 1398–1418, doi: 10.1785/0120140228.

Métivier, L., R. Brossier, S. Operto, and J. Virieux (2015). Acoustic multiparameter FWI for the reconstruction of P-wave velocity, density and attenuation: Preconditioned truncated Newton approach, *Expanded Abstracts, 85th Annual SEG Meeting*, New Orleans, Louisiana.

Moczo, P., J. Kristek, and M. Gális (2014). *The Finite-Difference Modelling of Earthquake Motions: Waves and Ruptures*, Cambridge University Press, Cambridge, United Kingdom.

Modrak, R., and J. Tromp (2016). Seismic waveform inversion best practices: Regional, global and exploration test cases, *Geophys. J. Int.* **206,** no. 3, 1864–1889.

Nettles, M., and A. M. Dziewonski (2008). Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America, *J. Geophys. Res.* **113,** no. B02303, doi: 10.1029/2006JB004819.

Operto, S., Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux (2013). A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice, *The Leading Edge* **32,** no. 9, 1040–1054.

Pratt, R. G., C. Shin, and G. J. Hicks (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.* **133,** 341–362.

Prieux, V., R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. I. Barkved, and J. H. Kommedal (2011). On the footprint of anisotropy on isotropic full waveform inversion: The Valhall case study, *Geophys. J. Int.* **187,** no. 3, 1495–1515, doi: 10.1111/j.1365-246X.2011.05209.x.

Prieux, V., R. Brossier, S. Operto, and J. Virieux (2013a). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: Imaging compressional wave speed, density and attenuation, *Geophys. J. Int.* **194,** 1640–1664, doi: 10.1093/gji/ggt177.

Prieux, V., R. Brossier, S. Operto, and J. Virieux (2013b). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 2: Imaging compressive-wave and shear-wave velocities, *Geophys. J. Int.* **194,** 1665–1681, doi: 10.1093/gji/ggt178.

Rickers, F., A. Fichtner, and J. Trampert (2013). The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion, *Earth Planet. Sci. Lett.* **367,** 39–51.

Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin (2001). Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion, *Geophysics* **66,** no. 6, 1856–1863.

Simute, S., H. Steptoe, L. Cobden, and A. Fichtner (2016). Full-waveform inversion of the Japanese Islands region, *J. Geophys. Res.* **121,** 3722–3741.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal (2010). Thematic set: Full waveform inversion: The next leap forward in imaging at Valhall, *First Break* **28,** no. 4, 65–70, doi: 10.3997/1365-2397.2010012.

Sirgue, L., O. I. Barkved, J. P. V. Gestel, O. J. Askim, and J. H. Kommedal (2009). 3D waveform inversion on Valhall wide azimuth OBC, *71st EAGE Conference and Exhibition*, available at http://www.earthdoc.org/detail.php?pubid=24070 (last accessed April 2018).

Stopin, A., R.-E. Plessix, and S. Al Abri (2014). Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman, *Geophysics* **79,** no. 3, WA69–WA77.

Tromp, J., H. Zhu, E. Bozdağ, and D. Peter (2011). Sensible choices for adjoint tomography: Misfits, model parameterization, preconditioning, smoothing and iterating, *2nd QUEST Workshop*, available at http://www.quest-itn.org/events/2nd-quest-workshop/files/quest2011_tromp.pdf (last accessed April 2018).

Vigh, D., X. Cheng, K. Jiao, D. Sun, and J. Kapoor (2014). Multiparameter TTI full wave-form inversion on long-offset broadband acquisition: A case study, *SEG Denver 2014 Annual Meeting, Society of Exploration Geophysics, Expanded Abstracts*, 1061–1065.

Vigh, D., K. Jiao, D. Watts, and D. Sun (2014). Elastic full-waveform inversion application using multicomponent measurements of seismic data collection, *Geophysics* **79,** no. 2, R63–R77.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, *et al.* (2013). Anisotropic 3D full-waveform inversion, *Geophysics* **78,** no. 2, R59–R80.

Yang, P., R. Brossier, L. Métivier, and J. Virieux (2016). Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method, *Geophysics* **81,** no. 6, R349–R362.

Yuan, Y. O., and F. J. Simons (2014). Multiscale adjoint waveform-difference tomography using wavelets, *Geophysics* **79,** no. 3, 79–95.

Yuan, Y. O., F. J. Simons, and E. Bozdağ (2015). Multiscale adjoint waveform tomography for surface and body waves, *Geophysics* **80,** no. 5, R281–R302.

Zhu, H., and J. Tromp (2013). Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean, *Science* **341,** 871–875.

Zhu, H., E. Bozdağ, T. S. Duffy, and J. Tromp (2013). Seismic attenuation beneath Europe and the North Atlantic: Implications for water in the mantle, *Earth Planet. Sci. Lett.* **381,** 1–11.

Zhu, H., E. Bozdağ, D. Peter, and J. Tromp (2012). Structure of the European upper mantle revealed by adjoint tomography, *Nature Geosci.* **5,** 493–498.

Zhu, H., E. Bozdağ, and J. Tromp (2015). Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.* **201,** no. 1, 18–52, doi: 10.1093/gji/ggu492.

Appendix

Kernel Preconditioning

There are several problems associated with kernel computations in local surface sedimentary structure (LSSS).

- Radiation patterns of regular and adjoint sources visibly affect kernel values. Obviously, this is an unwanted feature for model update. In LSSS, because of relatively low frequencies, these features take nonnegligible areas in the most important surface layers. The use of higher frequencies is not a solution, because this would increase computational costs, and tighter, higher-frequency kernels would decrease the already small coverage of the model.
- A complex interference wavefield leads to many artefacts in kernels: unrelated regular and adjoint waves coming from different directions overlapping at the same time and position, thus producing spatially scattered values in the kernel. They are often oscillatory in their nature, unrelated to the real structure. A very rough estimate of the number of these features: if an incoming regular wave is split into N phases at interfaces and scattering or diffraction points, and each of them generates corresponding adjoint waves at the receivers that are further split into N^2 adjoint phases, there are potentially $NN^2 = N^3$ overlapping areas. With a very complex wavefield, these kernel features may become dominant. Sufficient smoothing can reduce their effect.
- Wave amplitudes are considerably amplified in slow-surface sediments. The kernel, obtained from a product of the regular and adjoint field quantities that are linearly related to amplitudes, thus reaches very high values just close to the surface. Because of these high values, the inversion would be effectively constrained to an upper part of the model. Some additional preconditioning that accounts for wavefield values and/or velocity structure is therefore necessary.

The kernel, as it is defined, determines the direction of the fastest increase of the misfit. It would be, however, more efficient if the kernel could provide an estimate of a model error.

Kernel (gradient) preconditioning consists of a variety of kernel modifications before it is used for a model update. One of the first applications of preconditioning was a correction by Gauthier *et al.* (1986), who divided the kernel by approximate geometrical spreading to obtain a kernel with sensitivity independent of the distance from sources and receivers. Igel *et al.* (1996) applied preconditioning to correct for the 3D geometrical spreading in a 2D simulation. However, even with full 3D simulations, the preconditioning is necessary to obtain a physically meaningful inverted model. This is because computed kernels are very compli-

cated and not directly suitable for inversion. There are noticeable kernel singularities at point sources and receivers (acting as point sources in adjoint simulations), spatial high-frequency oscillations. Preconditioning may be also used for correction of an uneven source–receiver coverage. Appropriate preconditioning may help avoid local minima and achieve faster convergence of misfit minimization.

Standardly applied preconditioning consists of clipping and smoothing. It is efficient to combine these with additional corrections, for example, ray density (Tromp *et al.*, 2011; Luo *et al.*, 2013) or squared ray density (Luo *et al.*, 2013) or Hessian approximation (e.g., Pratt *et al.*, 1998; Shin *et al.*, 2001; Choi and Shin, 2008; Luo *et al.*, 2013). In the Spatially Dependent Normalization by Maximal Absolute Values section, we introduce another low-cost preconditioning, normalization by the maximal absolute values.

Clipping

Because kernel singularities represent only a small part of the whole model, we can eliminate them by choosing maximal acceptable kernel value and replacing greater values by the chosen maximum. The removal of the kernel values larger than the chosen maximum is substantial for convergence, although the chosen maximum can be just guessed.

Anisotropic Smoothing

To reduce spatial variability of the kernel, we can smooth the kernel. The smoothing removes small-scale heterogeneities and oscillatory values and improves model parts with sparse source–receiver configuration. We use exponential smoothing

$$K_{\text{smooth}}(\mathbf{r}) = K_{\text{smooth}}(x, z) = \frac{1}{A} \iiint_V K(x', z') e^{-q} dx' dz', \tag{A1}$$

in which A is an appropriate norm and q is defined by

$$q = q(x, y, x', z') = \frac{|x' - x|}{s_x} + \frac{|z' - z|}{s_z}, \tag{A2}$$

in which s_x and s_z are the horizontal and vertical smoothing intensities (characteristic lengths of the smoothing window function), respectively. We use an anisotropic smoothing (e.g., Prieux *et al.*, 2011; Zhu *et al.*, 2015) with larger intensity in the horizontal direction. This is in an agreement with a typical horizontally prolonged LSSS. We determine the smoothing intensities as

$$s_x = s\sqrt{r}, \qquad s_z = \frac{s}{\sqrt{r}}, \tag{A3}$$

in which s is the smoothing intensity and r is an estimated ratio of the horizontal-to-vertical dimensions of LSSS. We used $r = 10$ reflecting the estimated dimensions of the sedimentary basin.

The smoothing intensity is proportional to the characteristic wavelength, progressively decreasing as frequencies in the multiscale procedure increase.

Smoothing can be also position-dependent, smoothing less when data coverage is good (e.g., Bozdağ *et al.*, 2016). Required smoothing intensity can be quite large, for example, isotropic Gaussian smoothing with characteristic length of more than 16 characteristic wavelengths (Modrak and Tromp, 2016). This is too much for a typical, relatively small LSSS. The resulting kernel would be too smooth for inverting meaningful heterogeneities.

Spatially Dependent Normalization by Maximal Absolute Values

For removing singular values without excessive smoothing or artificial clipping, we can divide kernel values by spatially dependent norm $N(\mathbf{r})$ which attains the shape of kernel close to singularities. Based on theoretical analysis and numerical tests, we found a spatially dependent normalization by a product of temporal maxima of absolute values of the regular- and adjoint-field quantities used for kernel computation to be a suitable preconditioner for inversions in LSSS. This normalization removes singularities and reduces excessive values in slow sediments.

We apply the following norms for kernel components (equations 1–3):

$$N_\rho(\mathbf{r}) = \max\{\partial_t \mathbf{u}^\dagger(\mathbf{r}, t)\} \max\{\partial_t \mathbf{u}(\mathbf{r}, t)\}, \tag{A4}$$

$$N_\lambda(\mathbf{r}) = \max\{\mathrm{tr}\boldsymbol{\varepsilon}^\dagger(\mathbf{r}, t)\} \max\{\mathrm{tr}\boldsymbol{\varepsilon}(\mathbf{r}, t)\}, \tag{A5}$$

$$N_\mu(\mathbf{r}) = \max\{\|\boldsymbol{\varepsilon}^\dagger(\mathbf{r}, t)\|\} \max\{\|\boldsymbol{\varepsilon}(\mathbf{r}, t)\|\}. \tag{A6}$$

The norm of the strain tensor is defined by

$$\|\boldsymbol{\varepsilon}(\mathbf{r}, t)\| = \sqrt{\sum_{i,j} \varepsilon_{ij}^2(\mathbf{r}, t)}. \tag{A7}$$

Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
Mlynská dolina F1
842 48 Bratislava, Slovakia
moczo@fmph.uniba.sk