Automatic Global Multiscale Seismic Inversion: Insights into Model, Data, and Workflow Management

Michael Afanasiev
michael.afanasiev@erdw.ethz.ch

Christian Boehm
christian.boehm@erdw.ethz.ch

Alexey Gokhberg
alexey.gokhberg@erdw.ethz.ch

Andreas Fichtner
andreas.fichtner@erdw.ethz.ch

Department of Earth Sciences
ETH Zurich
Sonneggstr. 5, 8092 Zurich, Switzerland

ABSTRACT

Modern global seismic waveform tomography is formulated as a PDE-constrained nonlinear optimization problem, where the optimization variables are Earth's visco-elastic parameters. This particular problem has several defining characteristics. First, the solution to the forward problem, which involves the numerical solution of the elastic wave equation over continental to global scales, is computationally expensive. Second, the determinedness of the inverse problem varies dramatically as a function of data coverage. This is chiefly due to the uneven distribution of earthquake sources and seismometers, which in turn results in an uneven sampling of the parameter space. Third, the seismic wavefield depends nonlinearly on the Earth's structure. Sections of a seismogram which are close in time may be sensitive to structure greatly separated in space.

In addition to these theoretical difficulties, the seismic imaging community faces additional issues which are common across HPC applications. These include the storage of massive checkpoint files, the recovery from generic system failures, and the management of complex workflows, among others. While the community has access to solvers which can harness modern heterogeneous computing architectures, the computational bottleneck has fallen to these memory- and manpower-bounded issues.

We present a two-tiered solution to the above problems. To deal with the problems relating to computational expense, data coverage, and the increasing nonlinearity of waveform tomography with scale, we present the Collaborative Seismic Earth Model (CSEM). This model, and its associated framework, takes an open-source approach to global-scale seismic inversion. Instead of attempting to monolithically invert all available seismic data, the CSEM approach focuses on the inversion of specific geographic subregions, and then consistently integrates these subregions via a common computational framework. To deal with the workflow and storage issues, we present a suite of workflow management software, along with a custom designed optimization and data compression library. It is the goal of this paper to synthesize these above concepts, originally developed in isolation, into components of an automatic global-scale seismic inversion.

PASC '16, June 08–10, 2016, Lausanne, Switzerland

ISBN 978-1-4503-4126-4/16/06…$15.00

DOI: http://dx.doi.org/10.1145/2929908.2929910

CCS Concepts

•**Applied computing** → **Earth and atmospheric sciences;** •**Computing methodologies** → *Massively parallel and high-performance simulations;* •**Software and its engineering** → *Software libraries and repositories;*

Keywords

full-waveform inversion, seismic tomography, workflow management, compression

1. INTRODUCTION

Since its conception [20, 3], seismic tomography has developed into one of the most powerful tools to study the Earth's interior. Its ability to address important problems of societal, economic and scientific relevance has driven the impressive densification of seismic station coverage. Thousands of permanent seismic stations together with dense temporary arrays [41, 17] are being used today to record seismic wavefields. In addition to increasing data quality and quantity, technical developments have sharpened our images. Linearized ray tomography used early on [20, 3] has been complemented by methods that account, e.g., for nonlinearity [42, 5, 43] or the finite-frequency nature of seismic waves [56, 13, 28, 57]. Recently, 3D full-waveform inversion has become possible [11, 22, 50, 27], allowing us to exploit complete seismograms for the benefit of improved tomographic resolution.

While tomography has painted a fascinating picture of the Earth, it has reached a stage where individual researchers are unable to take full advantage of the rapidly expanding seismic data volume. Today's tomographic studies therefore mostly operate on either regional [11, 50, 37, 58] *or* global scales [27, 48, 15], using simplified theory and/or

small datasets. Applications that rely on highly-resolved Earth models thus lag behind the potential offered by the available data. Key science questions remain unanswered.

To address this issue, we have introduced the *Collaborative Seismic Earth Model (CSEM)* [1], that goes beyond traditional monoscale seismic tomography by constructing a multiscale Earth model through successive, community-driven, regional refinements. In isolation, the CSEM represents a snapshot of a continuously evolving global seismic inversion, with high resolution regional contributions to global-scale gradient update. It is our goal in this paper to dynamically extend the CSEM framework to include the regional contributions themselves.

This extension poses various computational challenges. These include (i) the spatial and physical parameterization of the Earth model and its efficient user interfacing, (ii) the consistent, collaborative, multiscale updating of the Earth model, (iii) the management of seismic data and the inversion workflow, (iv) the efficient nonlinear optimization of the misfit between observed and synthetic seismograms, and (v) the reduction of the massive storage requirements related to the computation of sensitivity kernels using adjoint methods. In the following sections, we provide a condensed description of approaches that allow us to overcome these challenges and to enable the future construction of a community-driven multiscale whole-Earth model.

2. THE COLLABORATIVE SEISMIC EARTH MODEL

In this section we describe the basic concepts of the CSEM, to provide a context for the subsequent descriptions of the workflow and optimization tools. In a sense, the role of the CSEM in this paper is to tie the other sections to the physical Earth. That is, it acts as factory responsible for the generation of the input Earth models referred to in section 3. As well, it is responsible for amalgamating the results of the corresponding seismic inversions into a consistent global model. To fully understand this process, it is necessary to review several important details of the CSEM's construction.

2.1 Parameterization

Contrary to most Earth models, the CSEM is parameterized in terms of a fully 3D, anisotropic, and attenuative medium. This means that 3D variations in all 21 elastic parameters, shear and compressional attenuation, and density, are stored. Where available, local estimates of resolution are stored as well. Due to the under-determinedness of the seismic inverse problem not all of these parameters can be resolved by a specific dataset. However, this overparameterization offers maximum flexibility, without restricting the model to a particular system of elastic symmetry that may not be compatible with seismic data included in future updates. The initial construction of the model, explained in section 2.2, considered contributions from transversely isotropic models, and as a result most of stored parameters are currently zero. To work with specific forward and inverse problem solvers that require a less general rheology, the fully anisotropic CSEM may be projected onto a lower-dimensional subspace, or transformed into a more classical parameterization.

Geometrically, the model is parameterized as a refinable tetrahedral finite-element mesh representation of the Earth. The default element size is approximately 100 km in the lat/lon direction, and the radial discretization varies between 1 km at the surface, and 50 km near the core-mantle boundary. These scale lengths allow us to capture all structural details present in the initial model, on both crustal and whole-mantle scales. High resolution subregions are incorporated into the model via local mesh refinements, with the mesh size set to one half of the estimated resolution length of the corresponding model.

The user-facing interface to the CSEM comes in the form of a Python/C++ API. The practical purpose of this interface is to abstract the extraction and interpolation of subregions, for use in generic forward problem solvers, and for visualization purposes. Interfaces to the CSEM currently exist for the spectral-element wave propagation solvers SES3D [30], SPECFEM3D_GLOBE [35, 34], the fast-marching traveltime solver FMTOMO [44, 14], and TERRA [8], a mantle convection solver used in geodynamic simulations.

2.2 Initial Model

As a 1-D background model, we used the Preliminary Reference Earth Model (PREM) [19] with the following modifications. The original discontinuity at 220 km depth is replaced by a linear gradient, and we ignore the poorly constrained P wave anisotropy.

Added to this background are the 3D S velocity perturbations from S20RTS [46], which are mapped to P velocity perturbations using the P-to-S scaling given in [47]. The initial crustal model is the one derived by [40, 39], which includes estimations of both S wave velocity and crustal thickness. These parameters are interpolated onto the CSEM mesh via a bilinear interpolation. Below these estimated crustal depths, the mantle values from PREM and S20RTS are stretched upwards if necessary. Within the crust, mappings from S velocity to P velocity and density are those given by [40].

Higher-resolution submodels were incorporated into the initial CSEM for the following regions: Australia [22, 24], Japan [18], the South Atlantic [12], and Europe with an embedded regional model of Anatolia [26, 25, 45]. All of these regional models were obtained by independent full-waveform inversions.

2.3 Workflow

The CSEM workflow comprises of four main stages, illustrated in figure 1. First, a geographic subregion is extracted from the multiscale 'master' CSEM. The region is at this point independent from the master Earth model, and exists on the numerical grid of a chosen solver.

Second, the seismic inverse problem is solved on this subregion. This stage includes the collection of regional data, repeated solutions of the forward problem, and the iterative solution of the optimization problem. The result of this stage is a model which is optimal with regard to the associated regional data, and the forward/inverse methodology selected. Of course, this choice of a specific solver and inverse method is highly problem-dependent. For example, if the target is a detailed crustal model of a particular region, full-waveform inversion methods may be chosen, while if the target is deep mantle structure, classical traveltime-based methods may be preferred. It is here that the other components of this paper, namely the workflow management and

optimization toolbox, are meant to fit in.

Third, the *model update*, or difference between the optimal and initial subregion model, is added back into the master CSEM. The update, rather than the absolute optimal model, is chosen as this (i) minimizes any solver and method-specific biases, and (ii) ensures that any under-resolved structure in the master is not overwritten. Finally, to ensure that any updates to subregions remain consistent with a global seismic dataset, a global-scale iteration of full-waveform inversion is performed on the updated master model. The technical details of this update may also be found in [1].

In broad strokes, these previous sections explain the technical details of the CSEM construction and workflow. With these details explained, it is worth reiterating the central concept: instead of constructing a high-resolution global Earth model by monolithically inverting all available data, we discretize the problem into independent geographic subproblems, and tie the solutions to these subproblems together within an ongoing large-scale global FWI. In the following sections, we describe the technical details of the associated workflow and optimization framework.

3. A HIGH PERFORMANCE SEISMIC INVERSION FACTORY

3.1 Motivation and Objectives

In recent years, massively parallel high-performance computers became the standard instruments for solving forward and inverse problems in seismology [36, 33, 50, 24, 30]. The respective software packages dedicated to forward and inverse waveform modeling specially designed for such computers are now mature and widely available. These packages achieve significant computational performance and provide researchers with an opportunity to solve problems of bigger size at higher resolution within a shorter time. However, a typical seismic inversion process contains various activities that are beyond the common solver functionality. They include management of information on seismic events and stations, 3D models, observed and synthetic seismograms, pre-processing of the observed signals, computation of misfits and adjoint sources, minimization of misfits, and process workflow management. These activities are time consuming, seldom sufficiently automated, and therefore represent a bottleneck that can substantially offset performance benefits provided by even the most powerful modern supercomputers. Furthermore, a typical system architecture of modern supercomputing platforms is oriented towards the maximum computational performance and provides limited standard facilities for automation of the supporting activities.

To address this challenge we have developed a prototype solution that automates all aspects of the seismic inversion process and is tuned for modern massively parallel high performance computing systems. We address several major aspects of the solution architecture, which include (i) design of an inversion state database for tracing all relevant aspects of the entire solution process, (ii) design of an extensible workflow management framework, (iii) integration with wave propagation solvers, (iv) integration with optimization packages, (v) computation of misfits and adjoint sources, and (vi) process monitoring. The software design fits well into the common massively parallel system architecture featuring a large number of computational nodes running distributed applications under control of batch-oriented resource managers. The solution prototype has been implemented on the "Piz Daint" supercomputer provided by the Swiss National Supercomputing Centre (CSCS).

3.2 Solution Architecture

The inversion process, illustrated in figure 2, is initiated with the definition of the modeling and inversion setup that remains invariant during the entire process. The setup comprises the type and parameters of the wave propagation solver, an initial Earth model, attributes of seismic events (e.g., earthquakes) and seismic stations, and observed seismograms for all involved events, stations, and recording channels. The inversion process is implemented as a sequence of iterations; each iteration starts with an initial Earth model and aims at producing a better final model. Each iteration involves multiple runs of the solver, which can be started in either normal or adjoint mode. Each run takes a distinct Earth model version as input and computes synthetic seismograms and, in the adjoint mode, misfit gradients. Based on these results, the misfits between observed and synthetic data are computed; they are passed, together with misfit gradients, to the optimization module, which performs model updates aiming at the minimization of misfits.

For the purposes of misfit computation the observed data are pre-processed. Pre-processing is performed separately for each event and consists of selection of a data slice corresponding to the duration of the event, filtering according to the specified bandwidth, instrument response removal, and interpolation of data samples according to the specified simulation time step. Pre-processing takes place before the first iteration; it is repeated for the following iterations only when the user decides to change the filtering bandwidth; otherwise the existing pre-processed data are reused.

The adjoint computations are performed in two phases, depicted as "adjoint phase 1" and "adjoint phase 2". At the first phase, the forward simulation is performed with the periodical storage of the intermediate wave fields. At the second phase the adjoint simulation proper is performed and the previously stored forward wave fields are combined with the adjoint fields to produce sensitivity kernels.

We use a special database for tracking the state of the inversion process; we call it an inversion state database. This database, shown in figure 3, represents a hierarchical structure with branches for the process setup, inversion iterations, and solver runs. A single setup branch contains all the setup data. For each iteration a separate iteration branch is created. It contains information related to the state of the entire iteration, in particular, the initial and final models, pre-processed observed data, and configuration parameters of various processing components. Iteration entities serve as roots for solver run branches, which contain information describing the state of individual solver runs, in particular, synthetic seismograms and respective misfit values. The setup, iteration, and solver run branches have uniform hierarchical structure specifying information at the event, station and channel levels. Earth models are represented as separate entities in the state database. To store and manage the inversion state we use SQLite, an open source, self-contained, serverless SQL database engine, which fits very well into the high-performance system architecture.

The setup tree specifies: (i) at root level, the initial model,

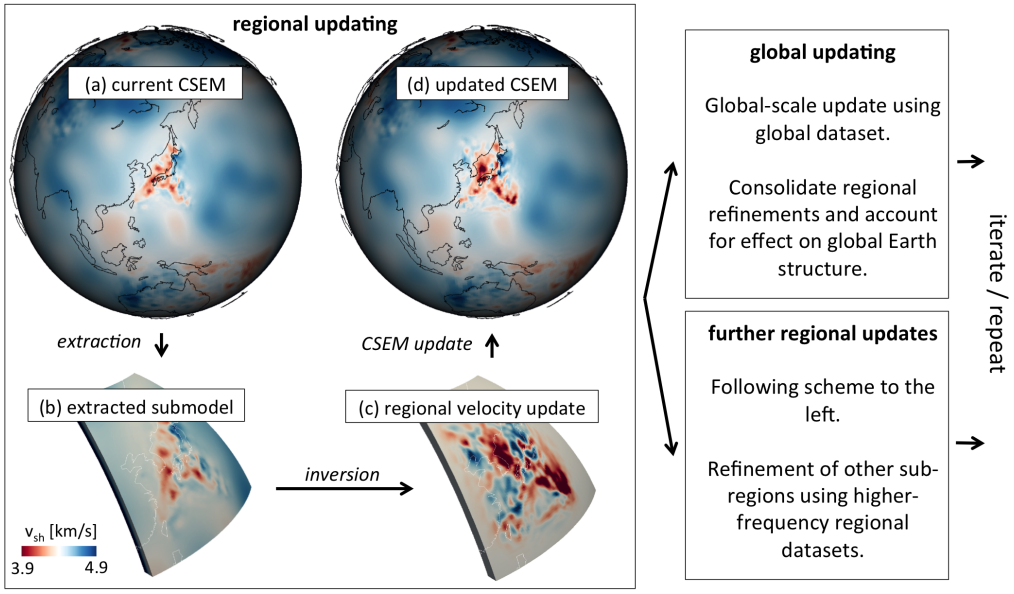

Figure 1: Illustration of the CSEM workflow. From the current CSEM (i) a subregion is extracted. The independent subregion (ii) may be updated using, e.g., full-waveform inversion. Updates to the initial subregion (iii) are then added back to the global CSEM (iv). The workflow may continue either with additional regional updates, or with a global FWI that ensures consistency of the regional refinements with the global longer-period data. Regional refinements can be contributed by external researchers.

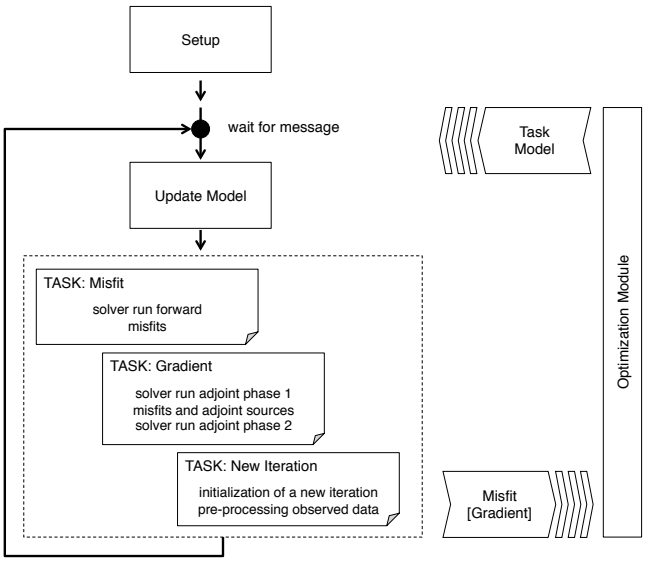

Figure 2: Schematic illustration of the inversion process.

simulation time step and number of steps and the solver configuration parameters; (ii) at event level, the event start time, location, and momentum; (iii) at station level, the station coordinates and definition file format; (iv) at channel level, the references to the station/channel definition file and raw observed data.

The iteration tree specifies: (i) at root level, the parameters for pre-processing, source time function, tapering, window selection, and misfit computations as well as the start and end models and gradient data associated with the iteration; (ii) at event level, the weight and gradient data associated with the event; (iii) at station level, the weight associated with the station; (iv) at channel level, the references to the pre-processed data and adjoint source time functions.

The solver run tree specifies: (i) at root level, the simulation type and phase, model, and misfit associated with the run; (ii) at event level, the misfit associated with the event; (iii) at station level, the misfit associated with the station; (iv) at channel level, the misfit associated with the channel and the reference to the synthetic data.

Each solver run record corresponds either to a single forward computation or to a single phase of an adjoint computation; therefore, adjoint computations produce two consecutive solver run records corresponding to the adjoint phases 1 and 2.

The workflow management framework implements control over the entire inversion process. For this task we could either choose one of the existing scientific workflow management frameworks or design a specialized solution tuned to our needs. In this, we had to address several requirements. The software architecture must be well tuned for job scheduling environments and resource control policies common for the state-of-the-art massively parallel HPC systems;

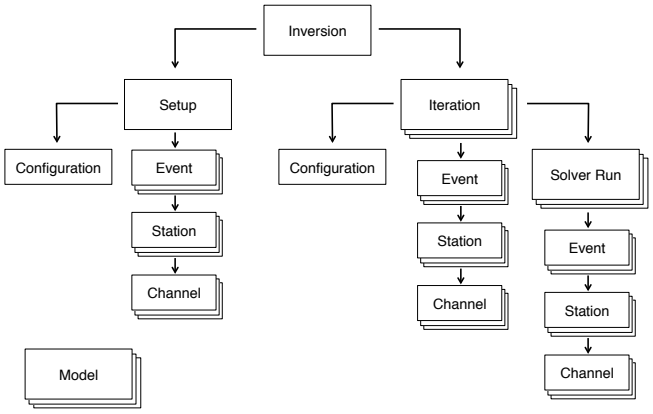

Figure 3: The process setup, inversion iterations, and solver runs are organized in a hierarchical inversion state database.

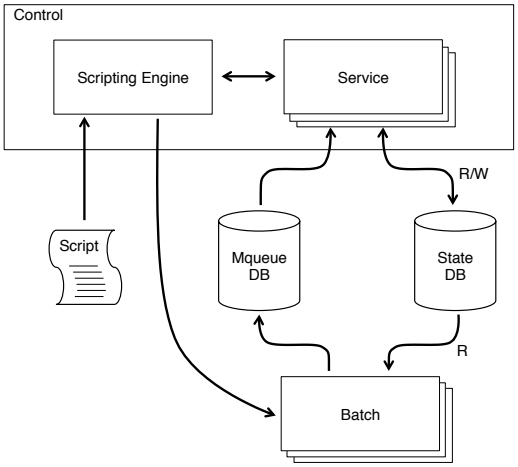

Figure 4: Schematic illustration of the workflow management framework that controls the inversion process.

in particular, it must be compatible with resource managers and job schedulers like SLURM or PBS. Many modern workflow management solutions are based on common Web-based technologies like Web services, however, using of such technologies on the massively parallel HPC platforms seems both impractical and unnecessary. Furthermore, it is important to choose an approach for the workflow specification. For the sake of simplicity, we strongly prefer defining workflows in terms of common structured programming constructs like choices or loops rather than directed acyclic graphs; respectively, for the workflow specification we prefer a simple built-in scripting language to a specialized GUI or a proprietary (e.g., XML-based) specification format. Finally, provided that the intended user community consists mainly of seismologists who are not necessarily experts in computer science, the workflow management software must be easy to learn, install, configure and maintain. In particular, the chosen software architecture must not implement functionality that is not needed on massively-parallel HPC systems or use unnecessarily complex technologies that do not add value for solving the seismic inversion problem.

We have reviewed a few scientific workflow management frameworks, in particular, Kepler [38], Pegasus [16], and Apache Taverna [54], and rejected them for the following reasons: (i) they are too large and complex and require a significant effort for learning, installation, configuration, and maintenance; (ii) they implement extensive functionality for grid and cloud platforms, which we do not need; (iii) they use relatively complex protocols and technologies like Web services, which do not add any value in our case; (iv) they define workflows in terms of directed acyclic graphs, which are specified either in graphical form using a specialized GUI or (in case of Pegasus) as XML-based DAX files generated using the proprietary API.

Therefore, we decided to implement a specialized lean workflow management framework well tuned for solving the seismic wave propagation problems on the massively parallel HPC platforms. This framework is illustrated schematically in figure 4. From the workflow management perspective, the inversion is represented as a sequence of activities, each activity implementing a certain atomic processing function, like seismogram pre-processing, running the solver, computation of misfits and adjoint sources or the model update. The activities can be implemented either as built-in services or external programs started via the batch resource manager of the host system. The external programs may be started in parallel on multiple nodes of the host computer. The workflow manager communicates with the external programs using asynchronous messages. The respective messaging engine is implemented using the SQLite message queue database. The workflow manager and external programs can access the inversion state database, however, to simplify the design, only the workflow manager has write access. The workflow logic, controlling the invocation order of activities, is specified using the process definition. We express process definitions as scripts written in the simple high-level programming language Oberon 2007 [53]. The workflow manager includes an embedded Oberon interpreter, which serves as a scripting engine. All available activities are represented as library functions and thus can be accessed from the scripts. This approach substantially improves the design flexibility and allows definition of various workflow scenarios without rebuilding the underlying software platform.

We prefer Oberon to the popular scripting languages like Python or Lua because it is easy to learn, combines power and simplicity, enforces a systematic approach towards programming and allows an extremely lean and simple implementation. SQLite was chosen as a database engine because its self-contained, serverless and zero-configuration architecture meets well the resource control requirements on the massively parallel HPC platforms.

At present the SES3D wave propagation solver [23, 30] is integrated in the solution. SES3D is a spectral-element implementation of full seismic waveform inversion in spherical 3D coordinates. It is designed for the massively parallel computing systems and can be used in both homogeneous (CPU only) and heterogeneous (equipped with GPU accelerators) hardware configurations. SES3D is integrated with the workflow manager as an external program running in parallel on multiple computing nodes. The workflow manager provides an abstract solver interface that can be adapted to different solvers. In particular, interfacing with

SPECFEM3D_GLOBE [35, 34] is currently work in progress.

A separate framework is designed for interoperability with an optimization module. The workflow manager and optimization process run in parallel and cooperate by exchanging messages according to a specially designed protocol. This protocol implements an abstract interface hiding implementation details of both the workflow manager and the optimization module. This approach facilitates independent design of both modules and improves solution flexibility. The architecture of the optimization module is discussed in detail in the following chapter.

A library of high-performance modules implementing signal pre-processing, misfit and adjoint source computations represents an integral part of the solution. The algorithms for misfit and adjoint source computations are based on the time-frequency transform of observed and synthetic seismograms with the use of phase and envelope misfits [21]. Monitoring is based on information stored in the inversion state database and at present implements a command line interface; design of a graphical user interface is in progress.

4. OPTIMIZATION LIBRARY

In this section, we describe the optimization module of the inversion framework in more detail. The motivation behind developing this library instead of using an existing general-purpose optimization package is twofold. On the one hand, we tailor mathematical optimization methods to the special structure of the seismic inverse problem in order to reduce the computational cost and to improve the quality of the reconstruction. Within the context of time-domain full-waveform inversion, on the other hand, we offer a generic implementation that allows us to easily interchange the parameterization of the Earth model and/or the wave propagation solver.

4.1 Features and Methods

From a mathematical perspective, the most relevant characteristics of the seismic inverse problem can be summarized as follows:

(i) high computational cost to compute the misfit functional and its derivatives with respect to structural parameters, which requires forward and adjoint simulations for each of the seismic events,

(ii) ill-posedness of the inverse problem, which requires some form of regularization,

(iii) non-convexity and the existence of multiple local minimizers, which can be partially circumvented with a suitably chosen misfit functional in combination with a good initial model and starting the inversion using only low frequencies,

(iv) a partially separable structure due to multi-experiment data, where the cost functional is the weighted sum of misfits from several seismic events, which can be computed independently of each other.

The optimization module exploits this problem-specific structure and implements customized trust-region methods using either exact second-order information [7] if available, or a quasi-Newton approximation of the Hessian [29]. Both of the aforementioned methods work matrix-free and are applicable to large-scale problems with millions of unknowns.

We extended those methods to work with inexact derivative information. Here, built-in error estimators of the optimization module adaptively control the level of inexactness to ensure convergence. This enables us, for instance, to exploit the separable structure of the objective function and to use only a subset of the seismic events to approximate the gradient. We combine ideas from the nonlinear Kaczmarz method [9] and related problems in machine learning [10] to reduce the required number of wave simulations for this task. Another source of inexactness in the gradient computation is introduced by lossy compression of the wavefield during the forward run, which in turn reduces the memory requirements significantly (see next section for details).

We distinguish three types of model parameterizations: the physical parameters that enter the elastic wave equation as coefficients, the actual parameters the user wants to infer (which might be a subset of the first parameters), and the optimization variables that the inversion toolbox uses internally. Although this is often done implicitly for gradient-based inversion schemes, we explicitly incorporate this separation in our inversion framework, because it offers a lot of flexibility for incorporating smoothing, scaling and regularization and for using different meshes for the model and the wavefield. The internally used model incorporates depth-scaling and multi-parameter coupling and parameterizes for relative model perturbations instead of absolute values. This improves the scaling of the problem and acts as a preconditioner for gradient-based descent schemes. Furthermore, this yields a better condition number of the linear system that needs to be solved iteratively in a Newton-CG method. Moreover, we directly incorporate a Gaussian filter into the internal parameterization to ensure smooth updates. All of the aforementioned strategies can be treated simultaneously by applying an affine transformation that maps the internally used parameters to the physical model. Likewise, we have to apply the adjoint of this transformation to the gradient with respect to the physical model parameters in order to obtain a consistent discrete gradient.

Smoothing the updates has a regularizing effect, but to further mitigate the ill-posedness of the inverse problem the optimization module offers built-in tools for adding a Tikhonov-type regularization [52] to the objective function.

Due to the tremendous computational cost of repeated calls to the wave propagation solver, we re-use as much information as possible during the iterations. This includes, for instance, bookkeeping of all previous simulations of the wave equation. Furthermore, we interpolate the misfit functional using previous evaluations whenever a step is rejected to obtain a better approximation of the misfit surface compared to classical backtracking approaches.

Prior knowledge to better constrain the nullspace of the problem can be included by restricting the set of feasible Earth models. The optimization module supports box constraints, such as lower and upper bounds on the parameters, as well as linear equality and inequality constraints. These restrictions on the feasible set of models do not involve a PDE, thus evaluating the constraints or their Jacobians is extremely cheap compared to solving the wave equation. Hence, we use a feasible-point trust-region SQP algorithm [55] that ensures feasibility in every iteration by projecting the model onto the set of models that satisfy all constraints. Non-linear constraints can be included as well, but require a custom implementation of the projection subproblem.

4.2 Interface to the Earth Model

The Earth model appears only as a template object within the optimization routines. This enables the use of solver-specific file formats without the need to modify the optimization module itself nor to convert Earth models to different data structures.

Besides basic input/output functionality there are only a few model-specific functions that have to be implemented for each template specialization. These include the functionalities of adding an Earth model to another model, scaling a model by a constant factor as well as point-wise scaling of a model by another model.

In addition, there are two callback functions for computing the inner product of two models and for computing the dual pairing. It is necessary to take spatial dependencies into account when solving the inverse problem, since the unknown Earth model is in fact a vector field defined on a three-dimensional domain. Hence, all computations involving the discrete Earth model should resemble the actions of the continuous vector field. Note that this becomes particularly important when the cells of the discretized Earth model vary in size. Furthermore, the choice of the inner product is known to have a great impact on performance of the minimization algorithm [32]. Hence, we change from the standard Euclidean metric to the one that is induced by the finite element discretization of the Earth model.

4.3 Interface to the Wave Propagation Solver

As has been mentioned in the previous section, the optimization module exchanges information with the workflow manager and triggers activities using a message protocol; cf. figure 2. Interaction with the wave propagation solver is required whenever the optimization module requests the value of the misfit functional or its derivatives for a certain Earth model. For each of these activities the optimization module sends a message specifying the task and a pointer to the corresponding model. Depending on the activity, the workflow management system will either return a double value (in case of computing the misfit value) or a pointer to the gradient, which is stored in the same format as the Earth model itself. The module sends an additional message after successful completion of each iteration. The workflow manager can use this callback, for instance, to select a new set of seismic events or to change time windows or frequencies for the next iteration.

There is also a fourth callback function for computing Hessian-vector products in case the wave propagation solver can handle general space- and time-dependent source terms in addition to point sources.

5. WAVEFIELD COMPRESSION

While the previous parts mainly dealt with the automation of processes within our seismic inversion framework, this section focuses on reducing the massive memory requirements by using compression methods tailored to seismic wavefields.

All of the optimization methods presented in the last section require derivatives of the misfit functional with respect to the Earth model. These sensitivities can be efficiently computed with the help of adjoint techniques. However, one of the main challenges in time-domain full-waveform inversion is the opposite time direction of forward and adjoint simulations and the necessity to access forward and adjoint

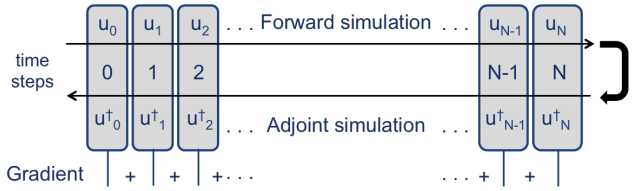

Figure 5: Illustration of the opposite time directions of the forward and adjoint wavefields, needed to compute sensitivity kernels.

wavefields simultaneously at the same time step during the computation of sensitivity kernels, which is sketched in figure 5. Hence, in addition to requiring a huge amount of computational resources for the wave propagation solver already, full-waveform inversion on 3D data sets also demands for massive storage capabilities. Storing the four-dimensional space-time cylinder of the forward wavefield is prohibitively expensive for large data sets and potential remedies like checkpointing [31, 4, 49, 2] or solving a so-called backward wave equation [51] introduce a significant computational overhead of at least one additional simulation. Hence, input/output (I/O) operations and the amount of auxiliary data that has to be transferred to and from disk increase the memory-boundedness of full-waveform inversion.

To overcome this bottleneck, we utilize techniques for the lossy compression of the spatio-temporal wavefield that

(i) provide a significant reduction of the I/O overhead without the need for additional simulations, and

(ii) maintain a similar rate of convergence when solving the inverse problem with inexact gradient information resulting from the compressed wavefield.

Following the modular approach that has been outlined in the previous sections, the compression techniques work independently of the PDE solver. Note, however, that spectral element methods using a high-order continuous Galerkin discretization and an explicit Newmark time stepping scheme achieve the highest performance in terms of compression rate and computational overhead. This numerical scheme is widely used in the seismological community, for instance, in the open source packages SPECFEM3D_GLOBE [35, 34] and SES3D [23, 30].

The compression module that we integrated into the seismic inversion framework extends the methods developed in [6] and embeds them into the minimization algorithm by adaptively steering the accuracy of the inexactly computed gradients. This approach combines a moving-window reinterpolation on a coarse temporal grid with spatial compression on hierarchical meshes using an adaptively chosen floating-point precision. Furthermore, we identify "shadow zones" in which forward and adjoint waves do not overlap, i.e., where neither storing nor even computing the wavefield is required.

It is important to note that the compression only affects the computation of the gradient of the misfit functional, whereas seismograms, misfits and adjoint sources are always computed with the full accuracy. Here, the compression algorithm exploits the problem-specific structure of the Fréchet derivative of the wavefield w.r.t. structural parameters and ensures that spatial errors (absolute and/or rela-

tive) remain below predefined thresholds. Hence, instead of enforcing a certain compression factor a-priori, we control the pointwise errors in the decompressed wavefield. These thresholds are chosen adaptively during the inversion and ensure that more accurate gradients are computed when approaching a local minimum. Thus, this strategy prevents losing high-resolution information during the compression. On the other hand, the achievable compression factor might decrease with an increasing model resolution.

This approach achieves an effective compression factor of three orders of magnitude in numerical experiments without slowing down the rate of convergence. Moreover, it is computationally cheap by adding only a few percents to the costs of a forward simulation.

With an ever-increasing resolution of the numerical mesh, even the reduced memory requirements using compression might become less efficient than checkpointing at some point, which of course depends on both, the problem size and the HPC architecture. In this case our compression methods can be used in combination with checkpointing, i.e., to compress the snapshots at the checkpoints.

6. CONCLUSIONS

We developed an ensemble of computational methods intended to facilitate the construction of a *Collaborative Seismic Earth Model*, i.e., a multiscale model of the Earth's interior that evolves successively by community-driven regional refinements. This ensemble has the following components: (i) A fully anisotropic, visco-elastic whole-Earth model, parameterized in terms of tetrahedra that can be refined in regions where higher-resolution refinements have been contributed. (ii) An inversion paradigm that ensures the consistent refinement of subregions, using different forward and inverse modeling techniques. (iii) A high-performance system for the management of seismic data and nonlinear, iterative inversions. (iv) An optimization library that is specifically tailored to full-waveform tomographic inverse problems. (v) A suite of compression algorithms that reduce the storage requirements of adjoint-based gradient calculations by three orders of magnitude without reducing convergence speed.

It is worth a quick conceptual tracing of the passage of an Earth model through these frameworks, now that their details have been explained. To begin, a subregion is extracted from the CSEM, and passed off to the inversion factory described in section 3. Here, during each nonlinear iteration, the model is passed down to the optimization library described in section 4, along with updates calculated with the assistance of the compression techniques described in section 5. After the final iteration, the updated model then works its way back up to the CSEM, where it becomes a component of a global-scale update. This process is then repeated as data and compute time allow.

The optimization and streamlining of these components, needed to bring the *Collaborative Seismic Earth Model* to production-stage maturity, are currently work in progress.

7. ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the paper.

This work was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under the CHRONOS project ch1 and PASC project GeoScale.

8. REFERENCES

[1] M. Afanasiev, D. Peter, K. Sager, S. Simutė, L. Ermert, L. Krischer, and A. Fichtner. Foundations for a multiscale collaborative global Earth model. *Geophys. J. Int.*, 204:39–58, 2016.

[2] V. Akçelik, G. Biros, and O. Ghattas. Parallel Multiscale Gauss-Newton-Krylov Methods for Inverse Wave Propagation. In *Proceedings of the 2002 ACM/IEEE Conference on Supercomputing*, SC '02, Baltimore, MD, USA, November 2002. IEEE/ACM.

[3] K. Aki and W. H. K. Lee. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes - 1. A homogeneous initial model. *J. Geophys. Res.*, 81:4381–4399, 1976.

[4] J. E. Anderson, L. Tan, and D. Wang. Time-reversal checkpointing methods for RTM and FWI. *Geophysics*, 77(4):S93–S103, 2012.

[5] H. Bijwaard, W. Spakman, and E. R. Engdahl. Closing the gap between regional and global traveltime tomography. *J. Geophys. Res.*, 103:30055–30078, 1998.

[6] C. Boehm, M. Hanzich, J. de la Puente, and A. Fichtner. Wavefield compression for adjoint methods in full-waveform inversion. *submitted to Geophysics*, 2015.

[7] C. Boehm and M. Ulbrich. A semismooth Newton-CG method for constrained parameter identification in seismic tomography. *SIAM J. Sci. Comput.*, 37(5):S334–S364, 2015.

[8] H.-P. Bunge, M. A. Richards, and J. R. Baumgardner. A sensitivity study of three-dimensional spherical mantle convection at 108 rayleigh number: effects of depth-dependent viscosity, heating mode, and an endothermic phase change. *J. Geophys. Res.*, 102:11991–12007, 1997.

[9] M. Burger and B. Kaltenbacher. Regularizing Newton–Kaczmarz methods for nonlinear ill-posed problems. *SIAM J. Numer. Anal.*, 44(1):153–182, 2006.

[10] R. H. Byrd, G. M. Chin, J. Nocedal, and Y. Wu. Sample size selection in optimization methods for machine learning. *Math. Program.*, 134(1):127–155, 2012.

[11] P. Chen, L. Zhao, and T. H. Jordan. Full 3D tomography for the crustal structure of the Los Angeles region. *Bull. Seis. Soc. Am.*, 97:1094–1120, 2007.

[12] L. Colli, A. Fichtner, and H.-P. Bunge. Full waveform tomography of the upper mantle in the South Atlantic region: Imaging westward fluxing shallow asthenosphere? *Tectonophysics*, 604:26–40, 2013.

[13] F. Dahlen, S.-H. Hung, and G. Nolet. Fréchet kernels for finite-frequency traveltimes – I. Theory. *Geophys. J. Int.*, 141:157–174, 2000.

[14] M. de Kool, N. Rawlinson, and M. Sambridge. A practical grid-based method for tracking multiple refraction and relection phases in 3d heterogeneous media. *Geophys. J. Int.*, 167:253–270, 2006.

[15] E. Debayle and Y. Ricard. Seismic observations of large-scale deformation at the bottom of fast-moving

plates. *Earth and Planetary Science Letters*, 376:165–177, 2013.

[16] E. Deelman, K. Vahi, G. Juve, M. Rynge, S. Callaghan, P. J. Maechling, R. Mayani, W. Chen, R. Ferreira da Silva, M. Livny, and K. Wenger. Pegasus: a workflow management system for science automation. *Future Gener. Comp. Sy.*, 46:17–35, 2015.

[17] J. Díaz, A. Villaseñor, J. Gallart, J. Morales, A. Pazos, D. Córdoba, J. Pulgar, J. L. García-Lobón, M. Harnafi, and TopoIberia Seismic Working Group. The IBERARRAY broadband seismic network: A new tool to investigate the deep structure beneath Iberia. *ORFEUS Newsletter*, 8:1–6, 2009.

[18] H. Diaz-Steptoe. *Full seismic waveform tomography of the Japan region using adjoint methods*. Master thesis, Utrecht University, 2013.

[19] A. M. Dziewoński and D. L. Anderson. Preliminary reference Earth model. *Phys. Earth Planet. In.*, 25:297–356, 1981.

[20] A. M. Dziewoński, B. H. Hager, and R. J. O'Connell. Large-scale heterogeneities in the lower mantle. *J. Geophys. Res.*, 82:239–255, 1977.

[21] A. Fichtner, B. L. N. Kennett, H. Igel, and H.-P. Bunge. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophys. J. Int.*, 175:665–685, 2008.

[22] A. Fichtner, B. L. N. Kennett, H. Igel, and H.-P. Bunge. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophys. J. Int.*, 179:1703–1725, 2009.

[23] A. Fichtner, B. L. N. Kennett, H. Igel, and H.-P. Bunge. Spectral-element simulation and inversion of seismic waves in a spherical section of the Earth. *J. Num. An. Ind. Appl. Math.*, 4:11–22, 2009.

[24] A. Fichtner, B. L. N. Kennett, H. Igel, and H.-P. Bunge. Full waveform tomography for radially anisotropic structure: New insight into present and past states of the Australasian upper mantle. *Earth Planet. Sci. Lett.*, 290:270–280, 2010.

[25] A. Fichtner, E. Saygin, T. Taymaz, P. Cupillard, Y. Capdeville, and J. Trampert. The deep structure of the North Anatolian Fault Zone. *Earth Planet. Sci. Lett.*, 373:109–117, 2013.

[26] A. Fichtner, J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. Villasenor. Multi-scale full waveform inversion. *Geophys. J. Int.*, 194:534–556, 2013.

[27] S. French, V. Lekic, and B. Romanowicz. Waveform tomography reveals channeled flow at the base of the oceanic lithosphere. *Science*, 342:227–230, 2013.

[28] W. Friederich. The S-velocity structure of the East Asian mantle from inversion of shear and surface waveforms. *Geophys. J. Int.*, 153:88–102, 2003.

[29] M. E. Gertz. A quasi-newton trust-region method. *Math. Program.*, 100(3):447–470, 2004.

[30] A. Gokhberg and A. Fichtner. Full-waveform inversion on heterogeneous HPC systems. *Comput. Geosci.*, 89:260–268, 2016.

[31] A. Griewank. Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation. *Optim. Methods Softw.*, 1(1):35–54, 1992.

[32] M. Heinkenschloss and L. N. Vicente. An interface optimization and application for the numerical solution of optimal control problems. *ACM Trans. Math. Softw.*, 25(2):157–190, June 1999.

[33] D. Komatitsch, G. Erlebacher, D. Göddeke, and D. Michea. High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster. *J. Comp. Phys.*, 229:7692–7714, 2010.

[34] D. Komatitsch and J. Tromp. Spectral-element simulations of global seismic wave propagation, part I: validation. *Geophys. J. Int.*, 149:390–412, 2002.

[35] D. Komatitsch and J. Tromp. Spectral-element simulations of global seismic wave propagation, part II: 3-D models, oceans, rotation, and gravity. *Geophys. J. Int.*, 150:303–318, 2002.

[36] D. Komatitsch, S. Tsuboi, C. Ji, and J. Tromp. A 14.6 billion degrees of freedom, 5 teraflops, 2.5 terabyte earthquake simulation on the Earth Simulator. In *Proceedings of the 2003 ACM/IEEE Conference on Supercomputing*, SC '03, pages 4–11, Phoenix, AZ, USA, 2003. ACM.

[37] E.-J. Lee, P. Chen, T. H. Jordan, P. B. Maechling, M. Denolle, and G. C. Beroza. Full-3D tomography (F3DT) for crustal structure in Southern California based on the scattering-integral (SI) and the adjoint-wavefield (AW) methods. *J. Geophys. Res.*, 119(8):6421–6451, 2014.

[38] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, and Y. Zhao. Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, 18(10):1039–1065, 2006.

[39] U. Meier, A. Curtis, and J. Trampert. Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophysical Research Letters*, 34(16), 2007. L16304.

[40] U. Meier, A. Curtis, and J. Trampert. Global crustal thickness from neural network inversion of surface wave data. *Geophys. J. Int.*, 169:706–722, 2007.

[41] A. Meltzer, R. Rudnick, P. Zeitler, A. Levander, G. Humphreys, K. Karlstrom, G. Ekström, R. Carlson, T. Dixon, M. Gurnis, P. Shearer, and R. D. van der Hilst. USArray initiative. *GSA Today*, 8-10, 1999.

[42] G. Nolet. Partitioned waveform inversion and two-dimensional structure under the Network of Autonomously Recording Seismographs. *J. Geophys. Res.*, 95:8499–8512, 1990.

[43] N. Rawlinson, A. M. Reading, and B. L. N. Kennett. Lithospheric structure of Tasmania from a novel form of teleseismic tomography. *J. Geophys. Res.*, 111(B2), 2006.

[44] N. Rawlinson and M. Sambridge. Wavefront evolution in strongly heterogeneous layered media using the fast marching method. *Geophys. J. Int.*, 156:631–647, 2004.

[45] F. Rickers, A. Fichtner, and J. Trampert. 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, 2013.

[46] J. Ritsema, H. van Heijst, and J. H. Woodhouse. Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science*, 286:1925–1928, 1999.

[47] J. Ritsema and H. J. van Heijst. Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle from P, PP, PPP and PKPab traveltimes. *Geophys. J. Int.*, 149:482–489, 2002.

[48] A. J. Schaeffer and S. Lebedev. Global shear speed structure of the upper mantle and transition zone. *Geophys. J. Int.*, 194:417–449, 2013.

[49] W. W. Symes. Reverse time migration with optimal checkpointing. *Geophysics*, 72(5):SM213–SM221, 2007.

[50] C. Tape, Q. Liu, A. Maggi, and J. Tromp. Seismic tomography of the southern California crust based upon spectral-element and adjoint methods. *Geophys. J. Int.*, 180:433–462, 2010.

[51] J. Tromp, D. Komatitsch, and Q. Liu. Spectral-element and adjoint methods in seismology. *Commun. Comput. Phys.*, 3(1):1–32, 2008.

[52] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, PA, 2002.

[53] N. Wirth. The programming language Oberon. https://www.inf.ethz.ch/personal/wirth/Oberon/Oberon07.Report.pdf, 2015. Revision 1.10.2013 / 4.3.2016, accessed 20.04.2016.

[54] K. Wolstencroft, R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, S. Soiland-Reyes, I. Dunlop, A. Nenadic, P. Fisher, J. Bhagat, K. Belhajjame, F. Bacall, A. Hardisty, A. Nieva de la Hidalga, M. P. Balcazar Vargas, S. Sufi, and C. Goble. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Res.*, 41(W1):W557–W561, 2013.

[55] S. J. Wright and M. J. Tenny. A feasible trust-region sequential quadratic programming algorithm. *SIAM J. Optim.*, 14(4):1074–1105, 2004.

[56] K. Yomogida. Fresnel zone inversion for lateral heterogeneities in the Earth. *Pure Appl. Geophys.*, 138:391–406, 1992.

[57] K. Yoshizawa and B. L. N. Kennett. Sensitivity kernels for finite-frequency surface waves. *Geophys. J. Int.*, 162:910–926, 2005.

[58] H. Zhu, E. Bozdağ, and J. Tromp. Seismic structure of the European upper mantle based on adjoint tomography. *Geophys. J. Int.*, 201:18–52, 2015.