A Data Centric View of Large-Scale Seismic Imaging Workflows

(Invited Paper)

Matthieu Lefebvre*, Ebru Bozdağ*, Henri Calandra†, Judy Hill‡, Wenjie Lei*, Daniel Peter§, Norbert Podhorszki‡, David Pugmire‡, Herurisa Rusmanugroho*, James Smith*, Jeroen Tromp*
*Princeton University, Princeton, NJ, USA
†Total SA E&P, USA
‡Oak Ridge National Laboratory, Oak Ridge, TN, USA
§ETH Zurich, Switzerland

***Abstract*—We discuss I/O challenges encountered in seismic imaging workflows on large HPC systems. Seismic tomography is widely used to image Earth's interior on all scales, from hydrocarbon reservoirs to the entire planet. The data volumes involved are large, and the computational requirements associated with the iterative imaging process are considerable. While software optimization still remains an important concern for superior performance, large-scale experiments and big data sets create bottlenecks in optimization-type workflows, causing significant I/O challenges. We address this problem by integrating parallel I/O libraries and new data formats in the workflow.**

I. Introduction

Knowledge about Earth's interior comes mainly from seismic observations and measurements. Seismic tomography is the most powerful technique for determining 3D images of the Earth —usually in terms of wavespeeds, density, or attenuation— using seismic waves generated by earthquakes or man-made sources recorded by a set of receivers. Advances in the theory of wave propagation and 3D numerical solvers together with dramatic increases in the amount and quality of seismic data and rapid developments in high-performance computing offer new opportunities to improve our understanding of the physics and chemistry of Earth's interior. Adjoint methods provide an efficient way of incorporating 3D numerical wave simulations in seismic imaging, and have been successfully applied for regional- and continental-scale problems [1], [2], [3] and —to some extent— in exploration seismology [4], [5]. However, it has so far remained a challenge on the global scale and in 3D exploration, mainly due to computational limitations.

In the context of adjoint tomography, scientific workflows are well defined. They consist of a few collective steps (e.g., mesh generation, model updates, etc.) and of a large number of independent steps (e.g., forward and adjoint simulations for each seismic event, pre- and post-processing of seismic data, etc.). The goal is to increase the accuracy of seismic models while keeping the entire procedure as efficient and stable as possible. While computational power still remains an important concern [6], large-scale experiments and big data sets create bottlenecks in workflows causing significant I/O problems on HPC systems.

Legacy seismic data formats were initially designed for specific seismic applications involving limited data sets, with little concern for performance. We are developing a new modern seismic data format based on ORNL's ADIOS libraries –called the Adaptable Seismic Data Format (ASDF)– that is suited for a variety of seismic workflows, allowing users to retain provenance related to observed and simulated seismograms. Here, we give examples from global adjoint tomography and exploration seismology, which are two of the extreme cases in seismic imaging. Pre-processing tools (resampling, filtering, window selection, computing adjoint sources, etc.) are modified to take advantage of this new data format. We accommodate the ADIOS libraries in our numerical solvers to reduce the number of files that are read and written during simulations (i.e., meshes, kernels, models, etc.) to drastically decrease disk access. We adjust post-processing tools (i.e., summing, pre-conditioning and smoothing gradients, model updates, etc.) accordingly. Moreover, parallel visualization tools, such as VisIt [7], take advantage of metadata included in our ADIOS outputs to extract features and display massive data.

II. Seismic Imaging Workflow

A. Overview of the workflow

In seismic tomography, the aim is to minimize differences between a set of observed data, generated by N_s sources and recorded by N_r receivers, and corresponding synthetic data through a pre-defined misfit function. The source can be passive (i.e., earthquakes) or active (i.e., explosions, air guns, etc.). The receivers have, in general, three components for earthquake studies and can have one or three components in exploration seismology. Seismic data are recorded as time series of a physical quantity, such as displacement, velocity or acceleration. In adjoint tomography, the gradient of a chosen misfit function is computed through the interaction of a forward seismic wavefield with its adjoint wavefield; the latter is generated by back-projecting measurements made on seismic data [8]. This procedure requires only two numerical simulations: one for the forward wavefield from source to receiver, and another for the adjoint wavefield from receiver to source. Seismic Earth models are iteratively updated based on

pre-conditioned conjugate gradient or L-BFGS optimization techniques.

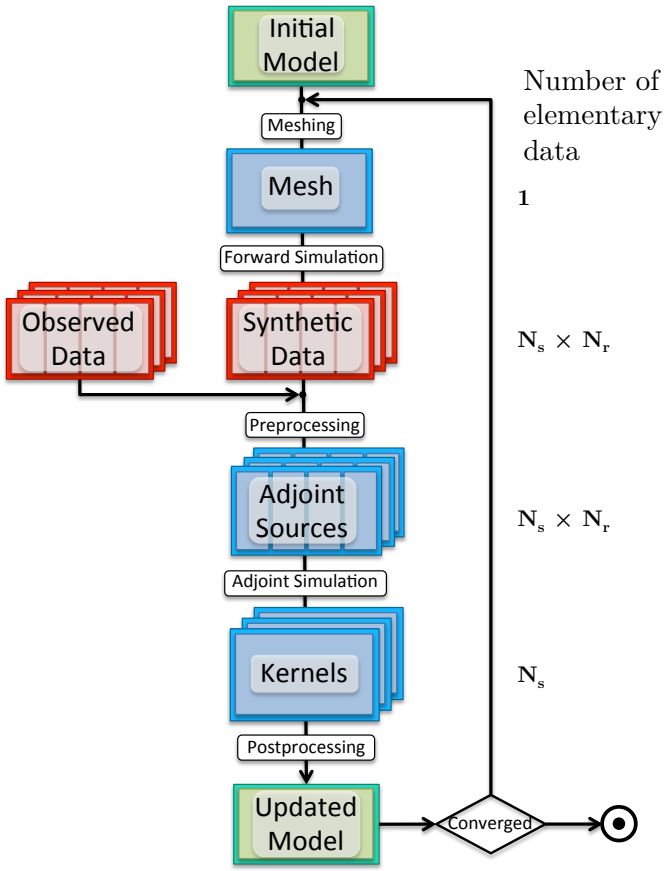

Fig. 1. General workflow for adjoint tomography. The focus is on the data involved at each step. Seismic data are depicted by red boxes and for each of the N_s seismic events they are recorded by N_r receivers. Computational data are represented by green and blue boxes. The amount of elementary data varies depending on the workflow stage and can eventually be grouped into a smaller number of files.

A typical workflow for adjoint tomography is shown in Figure 1, which consists of three major steps: 1) numerical simulations of forward and adjoint wavefields, 2) preprocessing, and 3) post-processing. To make the whole workflow efficient, we need to address each step while discussing the importance of the new data format for both seismograms and the products of our numerical solvers. It must be noted that most of the workflow is based on parallel or embarrassingly parallel processing:

- Forward and adjoint simulations are performed either with the SPECFEM3D package, which is used for simulations in exploration seismology, or its global counterpart SPECFEM3D_GLOBE used for continental- to global-scale simulations. A spectral-element method is used to solve the anelastic or acoustic wave equation numerically in realistic 3D models [9], [10]. The simulations are run for each source, also called an *event*, independent of the number of receivers. Thus we need to perform N_s forward and N_s adjoint simulations to obtain *event kernels*, which are the total sum of data sensitivity from all source-receiver pairs. Then the gradient of the misfit function is simply the sum of all event kernels.
- Observed and synthetic seismic data are not suitable for direct analysis. The pre-processing stage is dedicated to making measurements by comparing observed and simulated seismograms and to calculate adjoint sources to initiate adjoint simulations. This stage consists of three major steps which run in an embarrassingly parallel fashion: 1) data processing (i.e., tapering, resampling, filtering, deconvolution of instrument response, etc.), 2) window selection to determine the usable parts of seismograms to make measurements based on the automated phase picking algorithm FLEXWIN [11], and 3) computing adjoint sources based on the MEASURE_ADJ package [1]. In the case of exploration seismology, the adjoint source is simply propagated backward in time without windowing [4], [5].
- In the post-processing stage, once the gradient is obtained by summing the event kernels resulting from the adjoint simulations, it is pre-conditioned and smoothed. Then the model update is performed based on a conjugate gradient or L-BFGS [12] optimization scheme. This stage also involves a line-search to determine the step-length in the (conjugate) gradient direction.

B. Data

The workflow sketched in the previous sections deals with two main types of data: seismic and computational data. The entire process is based on measurements on seismic data for each source-receiver pair. A pair of observed and synthetic seismograms (depicted in red in Figure 1) are lightweight ($\sim$ 10 KB), but it should be noted that each seismic event can be recorded by thousands to tens of thousands of receivers on three components. For instance, in global inversions, it is common to assimilate data from more than 2,000 globally distributed seismic stations on three components, which easily leads to tens of thousands of $\sim$ 200 min long seismic records per event. In the case of exploration seismology, for instance 3D marine data acquisition, the streamers can contain sixty thousand hydrophones, and the number of shots can reach fifty thousand, depending on region of interest. The time record length depends on the length of the streamer or cable, but is typically about 12 s. Due to limitations in conventional marine seismic surveys, Ocean Bottom Seismometers or Ocean Bottom Nodes data containing three-component geophones and a single hydrophone are sometimes acquired, which increases the volume of seismic data in inversion workflows.

Computational data are, in general, characterized by discretization and representation of the scientific problem. In our case, these are mesh and model files, and data sensitivity kernels, which are the output of SPECFEM3D and SPECFEM3D_GLOBE used in the post-processing stage. They are shown in blue and green in the workflow chart shown in Figure 1. The size of these files depends on the spatial and temporal resolutions. For instance, a transversely

isotropic global adjoint simulation (100 min long seismograms at a resolution going down to 27 s with 1300 receivers) reads 49 GB of computational data and writes out 8 GB of data for adjoint data sensitivity kernels. When increasing the resolution of the simulations by going down to a shortest period of 9 s, all these numbers should be multiplied by 3^3, yielding about 1.3 TB of computational data.

III. Enhancing I/O performance and data utilization

A. Constraints

The previous section described two different types of data. The different purposes of these data imply different constraints.

The main bottlenecks in the adjoint tomography workflow stem from the number of files to be read and written, which reduces performance significantly and creates problems on filesystems due to heavy I/O traffic. Classical seismic data formats, which describe each seismic trace as a single file, exacerbate this problem. Moreover, since we are considering large seismic simulations on the latest supercomputers, we have a particular concern about parallel performance.

It is apparent that the classical data formats neither fulfill the computational requirements on HPC systems nor the provenance of computations and analysis for reproducibility of experiments. Furthermore, the lack of a common and flexible data format, both seismic and computational, has been a major problem in the seismological community, restricting the exchange of data and Earth models, and thus collaborative science.

B. Efficient and structured I/O with the ADIOS library

The large number of input and output files involved in 3D solvers and the pre- and post-processing stages creates a severe limitation on scaling jobs on HPC systems by stressing filesystems with heavy I/O loads. Therefore, it is a requirement to drastically reduce the number of files, both for computational and seismic data.

Parallel performance imposes use a library allowing parallel file access file. Several choices are available, but the most straightforward choice is MPI-IO. MPI-IO demands the definition of correct access patterns for each MPI process and to tune the software very carefully for each computer architecture, network and file systems. Recently, more portable solutions have become available. For instance, netCDF, HDF5 and their parallel counterparts allow users to write out files together with associated metadata. They provide an efficient way to organize data in accordance with requirements of the scientific problem and to keep track of the evolution of computations within the frame of workflows. An alternative is the ADIOS format [13], [14] released by ORNL. Compared to netCDF and HDF5, it works on simpler data structures since its main focus is on parallel performance. Besides metadata availability similar to other formats mentioned before, it also lets users change the transport method to target the most efficient I/O method for a particular system.

C. Optimizing computational data

Computational data, in general, do not require complex metadata since they are well structured within our numerical solvers. Until now, the way the computational data is written on disk was not problematic on local clusters for smaller size scientific problems (e.g., regional- or continental-scale wave propagation, small seismic data sets, etc.). However, to run simulations more efficiently on HPC systems for more challenging problems, such as global adjoint tomography or increased resolution regional- and exploration-scale tomography, we need to revise the way the solver handles computational data. In the old version, for each variable or set of closely related variables, a file was created for each MPI process. The number of files, for a single seismic event, was proportional to the number of MPI processes P. For a full step of the iterative workflow the number of files was $\mathcal{O}(P.N_s)$. Accessing these files during large-scale simulations did not only have an impact on performance, but also on the filesystem due to heavy I/O traffic. The new implementation uses ADIOS to limit the number of files accessed during reading and writing of computational data, independent of the number of processes, that is $\mathcal{O}(N_s)$. As an additional benefit, using ADIOS, HDF5 or netCDF will let us define readers for popular visualization tools such as Paraview and VisIt.

D. Toward a new parallel and adaptable seismic data format

Seismic data are stored in various legacy formats that differ from one application to another. Earthquake seismologists usually prefer the SAC (Seismic Analysis Code) format, in which every seismic trace is written in a separate file, i.e., the number of files equals to the number of seismic traces. It allows users to store some metadata related to earthquake and station information, however, in a rigid and limited way. On the other hand, in exploration seismology, Seismic Unix (SU) or SEGY formats are usually preferred, which allow users to gather a set of seismic traces, but again with limited metadata.

All these formats embed headers designed to record information at fixed locations. Hence, they often fail to satisfy the requirements of modern seismological applications involving complex workflows and big data. To address all these issues, we are working on a new data format named the Adaptable Seismic Data Format (ASDF) based on the ADIOS libraries. It is intended to store many seismic traces in a single file (for instance, a single file per seismic event rather than having thousands of files from each component of each receiver). It also gives users the flexibility to design metadata according to their problem without any restriction on size or number of headers. Computation wise, gathering seismic traces through a parallel I/O library helps lessen the impact on the filesystem due to the availability of data aggregation methods.

IV. Conclusion and perspectives

We have outlined the difficulties in modern seismology workflows mainly related to handling large data sets on HPC systems. Even though the data volume is not comparable to what is commonly referred as “big data”, workflows and

data management tools are likely to create performance and filesystem issues on supercomputers. Using file formats that include metadata and embedded optimized I/O techniques not only helps increase computational performance but also ensures reproducibility, and in the long term brings a standard to seismic and computational data which would ultimately increase collaborations within the seismological community. As our seismic workflow is composed of well defined steps, introducing Scientific Workflow Management Software, such as Swift [15], [16], should help us focus more on scientific results and avoid the burden of manually dealing with a large number of processing steps.

REFERENCES

[1] C. Tape, Q. Liu, A. Maggi, and J. Tromp, "Adjoint tomography of the southern California crust." *Science*, vol. 325, no. 5943, pp. 988–92, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19696349

[2] 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," *Geophysical Journal International*, vol. 179, no. 3, pp. 1703–1725, 2009. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-246X.2009.04368.x

[3] H. Zhu, E. Bozdağ, D. Peter, and J. Tromp, "Structure of the European upper mantle revealed by adjoint tomography," *Nature Geoscience*, vol. 5, no. 7, pp. 493–498, 2012. [Online]. Available: http://www.nature.com/doifinder/10.1038/ngeo1501

[4] H. Zhu, Y. Luo, T. Nissen-Meyer, C. Morency, and J. Tromp, "Elastic imaging and time-lapse migration based on adjoint methdos," *GEOPHYSICS*, vol. 74, no. 6, pp. WCA167–WCA177, 2009. [Online]. Available: http://library.seg.org/doi/abs/10.1190/1.3261747

[5] Y. Luo, J. Tromp, B. Denel, and H. Calandra, "3D coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods," *GEOPHYSICS*, vol. 78, no. 4, pp. S193–S202, 2013. [Online]. Available: http://library.seg.org/doi/abs/10.1190/geo2012-0462.1

[6] M. Rietmann, P. Messmer, T. Nissen-Meyer, D. Peter, P. Basini, D. Komatitsch, O. Schenk, J. Tromp, L. Boschi, and D. Giardini, "Forward and adjoint simulations of seismic wave propagation on emerging large-scale GPU architectures," in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, ser. SC '12. Los Alamitos, CA, USA: IEEE Computer Society Press, 2012, pp. 38:1—-38:11. [Online]. Available: http://dl.acm.org/citation.cfm?id=2388996.2389048

[7] H. Childs, E. Brugger, B. Whitlock, J. Meredith, S. Ahern, K. Bonnell, M. Miller, G. H. Weber, C. Harrison, D. Pugmire, T. Fogal, C. Garth, A. Sanderson, E. W. Bethel, M. Durant, D. Camp, J. M. Favrek, O. Rubel, P. Navratil, M. Wheelera, P. Selbya, and F. Vivodtzev, "VisIt: An End-User Tool for Visualizing and Analyzing Very Large Data," in *SciDAC*, 2011.

[8] A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984. [Online]. Available: http://www.ipgp.fr/ tarantola/Files/Professional/Papers_PDF/InversionOfSeismic.pdf

[9] D. Komatitsch and J. Tromp, "Spectral-element simulations of global seismic wave propagation-I. Validation," *Geophysical Journal International*, vol. 149, no. 2, pp. 390–412, 2002. [Online]. Available: http://doi.wiley.com/10.1046/j.1365-246X.2002.01653.x

[10] ——, "Spectral-element simulations of global seismic wave propagation II . Three-dimensional models , oceans , rotation and self-gravitation," *Geophysical Journal International*, vol. 150, no. 1, pp. 303–318, 2002. [Online]. Available: http://doi.wiley.com/10.1046/j.1365-246X.2002.01716.x

[11] A. Maggi, C. Tape, M. Chen, D. Chao, and J. Tromp, "An automated time-window selection algorithm for seismic tomography," *Geophysical Journal International*, vol. 178, no. 1, pp. 257–281, 2009. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-246X.2009.04099.x

[12] J. Nocedal, "Updating Quasi-Newton Matrices with Limited Storage," *Mathematics of Computation*, vol. 35, no. 151, pp. 773–782, 1980. [Online]. Available: http://www.jstor.org/stable/2006193?origin=crossref

[13] J. Lofstead, S. Klasky, K. Schwan, N. Podhorszki, and C. Jin, "Flexible IO and Integration for Scientific Codes Through The Adaptable IO System (ADIOS)," *Computing*, pp. 15–24, 2008. [Online]. Available: http://portal.acm.org/citation.cfm?id=1383533

[14] Q. Liu, J. Logan, Y. Tian, H. Abbasi, N. Podhorszki, J. Y. Choi, S. Klasky, R. Tchoua, J. Lofstead, R. Oldfield, M. Parashar, N. Samatova, K. Schwan, A. Shoshani, M. Wolf, K. Wu, and W. Yu, "Hello adios: the challenges and lessons of developing leadership class i/o frameworks," *Concurrency and Computation: Practice and Experience*, 2013. [Online]. Available: http://dx.doi.org/10.1002/cpe.3125

[15] M. Wilde, I. Foster, K. Iskra, P. Beckman, Z. Zhang, A. Espinosa, M. Hategan, B. Clifford, and I. Raicu, "Parallel Scripting for Applications at the Petascale and Beyond," *Computer*, vol. 42, no. 11, pp. 50–60, 2009. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/MC.2009.365

[16] Y. Zhao, M. Hategan, B. Clifford, I. Foster, G. Von Laszewski, V. Nefedova, I. Raicu, T. Stef-Praun, and M. Wilde, "Swift: Fast, Reliable, Loosely Coupled Parallel Computation," *2007 IEEE Congress on Services Services 2007*, vol. 0, no. Services, pp. 199–206, 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4278797