Solid Earth Discuss., 7, 957–1005, 2015
www.solid-earth-discuss.net/7/957/2015/
doi:10.5194/sed-7-957-2015

Solid Earth Discussions
Open Access

Instaseis: instant global seismograms based on a broadband waveform database

M. van Driel[1], L. Krischer[2], S. C. Stähler[2,3], K. Hosseini[2], and T. Nissen-Meyer[4]

[1]Dept. of Earth Sciences, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland
[2]Dept. of Earth Sciences, Ludwig-Maximilians-Universität München, Theresienstraße 41, 80333 München, Germany
[3]Leibniz-Institute for Baltic Sea Research, Warnemünde, Seestraße 15, 18119 Rostock, Germany
[4]Dept. of Earth Sciences, University of Oxford, South Parks 49 Road, Oxford OX1 3AN, UK

Received: 16 January 2015 – Accepted: 9 February 2015 – Published: 6 March 2015

Correspondence to: M. van Driel (martin@vandriel.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

We present a new method and implementation (*Instaseis*) to store global Green's functions in a database which allows for near-instantaneous (on the order of milliseconds) extraction of arbitrary seismograms. Using the axisymmetric spectral element method (*AxiSEM*), the generation of these databases, based on reciprocity of the Green's functions, is very efficient and is approximately half as expensive as a single *AxiSEM* forward run. Thus, this enables the computation of full databases at half the cost of the computation of seismograms for a single source in the previous scheme and allows to compute databases at the highest frequencies globally observed. By storing the basis coefficients of the numerical scheme (Lagrange polynomials), the Green's functions are 4th order accurate in space and the spatial discretization respects discontinuities in the velocity model exactly. High order temporal interpolation using Lanczos resampling allows to retrieve seismograms at any sampling rate. *AxiSEM* is easily adaptable to arbitrary 1-D models or other spherical objects such as Mars. In this paper, we present the basic rationale and details of the method as well as benchmarks and illustrate a variety of applications. The code is open source and available with extensive documentation at www.instaseis.net.

1 Introduction

Despite the exponential growth of computational power and substantial progress of 3-D numerical methods for seismic wave propagation in the last 15 years (Igel et al., 2000; Komatitsch and Tromp, 2002a; Tromp, 2007; Tromp et al., 2010), the simulation of the highest frequencies observed in seismic waves on the global scale remains a high-performance computing challenge and is not yet done routinely. This is why many seismologists still rely on approximate methods to compute and analyze high-frequency body waves such as ray-theoretical traveltimes (e.g. the *TauP*-toolkit described in Crotwell et al., 1999), WKBJ synthetics (Chapman, 1978), the reflectivity

SED
7, 957–1005, 2015

method (Fuchs and Müller, 1971) or the frequency-wavenumber integration method (Kikuchi and Kanamori, 1982). More recently, several methods that include the full physics in solving the seismic wave equation while reaching the highest observable frequencies by assuming spherically symmetric models have become available, see Fig. 1 for an example. These methods include the direct solution method (*DSM* Geller and Ohminato, 1994; Kawai et al., 2006), the frequency domain integration method (*GEMINI* Friederich and Dalkolmo, 1995) and a generalization of it including self gravitation (*Yspec* Al-Attar and Woodhouse, 2008).

As detailed by Nissen-Meyer et al. (2007b), the main drawback of these methods when applied to computing wavefields rather than single seismograms, is their scaling proportional to the number of points in space where the wavefield is sampled. This motivated the development of a direct time-domain approach, where the displacement as a function of space and time is a natural field variable and only needs to be written to disk (Nissen-Meyer et al., 2007a, b, 2008). The implementation of this axisymmetric spectral element method *AxiSEM* was recently extended to include anisotropy and attenuation (van Driel and Nissen-Meyer, 2014a, b), published under public license (Nissen-Meyer et al., 2014) and is available at www.axisem.info.

As computing full global waveforms especially at higher frequencies requires substantial computational resources, several initiatives serve to deliver waveforms by means of databases without having to run a full numerical solver. The *ShakeMovie* project (http://global.shakemovie.princeton.edu) provides synthetics for earthquakes from the CMT catalogue (www.globalcmt.org) recorded at permanent GSN and FDSN stations in 1-D and 3-D velocity models (Tromp et al., 2010). The *Pyrocko* toolbox (http://emolch.github.io/pyrocko) provides a Python interface to generate and access Green's function databases, which for the global case are based on *GEMINI*, several databases are offered for download. In this paper we present a method that uses *AxiSEM* to generate global Green's function databases and provides a Python interface for convenient extraction of seismograms. The advantage over *ShakeMovie* synthetics are the possible higher frequencies and arbitrary source and receiver combinations

SED
7, 957–1005, 2015

independent of catalogues and real stations. Compared to *Pyrocko* with *GEMINI* synthetics, *AxiSEM* is more efficient in generating the databases, allowing to routinely compute them for a large number of different background models or specialized applications (e.g. limited depth/distance ranges). Also, by using the Lagrangian polynomials in the SEM mesh as basis functions, it achieves higher spatial accuracy.

This paper is structured as follows. In Sect. 2 we present the technical aspects and argue for the choices made in the spatial and temporal discretization. Section 3 gives a short overview of the Python interface. In Sect. 4 we show the performance with respect to accuracy, speed and disk space requirements for the databases. Finally, we depict a variety of applications in Sect. 5.

2 Methods

2.1 Computing Green's functions with *AxiSEM*

AxiSEM was from the beginning designed with the application of computing global wavefields rather then single seismograms in mind (Nissen-Meyer et al., 2007b). This becomes apparent in the following main advantages in this application: it uses a 2-D discretization (Fig. 2), with an analytical decomposition of the 3-D wavefield into several 2-D wavefields. For moment tensor sources, four 2-D wavefields are needed, for force sources two. As it is a time domain method, the displacement field in space-time is a natural field variable of the numerical scheme and simply needs to be written to disk without any extra computational cost when larger regions of Earth are included in the database. *AxiSEM* uses a spectral element scheme for spatial discretization which lends itself well to parallelization on HPC systems. As it is based on the weak formulation of the wave equation, it naturally includes the free surface boundary condition and allows for highly accurate modelling of surface waves.

Nissen-Meyer et al. (2014) argued against using collective parallel I/O since the availability of the NetCDF libraries (Rew and Davis, 1990) was not granted on all

SED
7, 957–1005, 2015

supercomputers. For that reason, we implemented a round robin I/O scheme, which remains advantageous when running *AxiSEM* on less then about 100 cores in parallel and to avoid installation problems on systems where NetCDF is not available as a pre-compiled package. On supercomputers however, the situation has since improved and NetCDF compiled with parallel support seems now to be wide spread. For this reason, we implemented a collective parallel I/O scheme that performs well, even when running on more than 1000 cores, see Table 1. In this scheme, all processes that have to write data to disk communicate via the message passing interface (MPI) and then write collectively at the same time to the parallel file system. This way we achieved throughputs of up to 4 GB s^{-1} on *SuperMUC*.

2.2 Forward and backward databases

Instaseis has the capability of dealing with *forward* wavefields, i.e. the waves are propagated from a moment-tensor point source at fixed depth (i.e. receivers exist throughout the medium), as well as *backward* or *reciprocal* wavefields, where the wavefields are propagated from a single-force point source at fixed depth and recorded throughout the medium (i.e. sources exist throughout the medium).

Potential applications of *forward databases* are the generation of 3-D wave-propagation movies (Holtzman et al., 2013), the computation of incoming teleseismic waves in 1-D/3-D hybrid methods (e.g. Monteiller et al., 2012; Masson et al., 2013) or the forward field in the computation of sensitivity kernels (Nissen-Meyer et al., 2007a) for seismic tomography. To generate a *forward* database, a total of four runs with *AxiSEM* are needed (Nissen-Meyer et al., 2007b).

In contrast, *reciprocal* databases utilize the reciprocity of the Green's functions, and are useful in all cases where the receivers are at fixed depth, thus for instance mimicking earthquake catalogues recorded at stations along the surface. The source can be located anywhere in the region where the Green's functions are recorded in the simulation, thus allowing for unlimited choices in the source-receiver geometry. To generate a *reciprocal* database, a total of two runs with *AxiSEM* are needed, one for

SED
7, 957–1005, 2015

the vertical component and one for both horizontal components of the seismogram (Nissen-Meyer et al., 2007b). It is also possible to compute a database for the vertical component seismograms only, which is then a factor of three faster and uses only about 40 % of the disk space.

2.3 The spatial scheme

For the spatial discretization we choose to keep the same basis as used in *AxiSEM*. The displacement $\boldsymbol{u}$ is expanded in terms of Lagrangian polynomials l_i (see Fig. 3) of order N defined on the integration points of the spectral element scheme (see Fig. 4) within each element:

$$\boldsymbol{u}(\xi,\eta,t) = \sum_{ij=0}^{N} \boldsymbol{u}_{ij}(t) l_i(\xi) l_j(\eta) \tag{1}$$

ξ and η are the reference coordinates of the element and N typically has a value of 4. This approach has several advantages:

- The wavefield is represented by polynomials, typically of degree 4, interpolation is hence of 4th order accuracy.
- The basis is local and only few coefficients are needed to represent the wavefield inside an element (typically 25), in contrast to e.g. global basis functions such as spherical harmonics.
- Discontinuities in the model that cause discontinuities in the strain Green's functions are respected by the mesh.
- The strain tensor (representing the moment tensor in the reciprocal case) can be computed on the fly from the stored displacements at high accuracy. This reduces the storage by a factor of two as the displacement has three degrees of freedom, compared to six for the strain.

SED
7, 957–1005, 2015

– Since the displacement is continuous also at model discontinuities and element boundaries, we need to store it only once at all Gauss–Lobatto–Legendre (GLL) points that belong to multiple elements, reducing the storage by another factor of 16/25 = 0.64 (see Fig. 4).

– Storing the displacement allows to use force sources as well without any extra computation or storage requirements.

Figure 5 visualizes the spatial representation for a long period mesh (50 s) for the Rayleigh wave train and the $G_{rr,r}$ component of the strain Green's tensor: the strain is smooth also across the doubling layer of the mesh where the background model (*ak135f*, Montagner and Kennett, 1996) is smooth as well. Still, the discontinuities of the model and hence the strain are explicitly represented by this discretization and the resolution of the mesh is adapted to the local wavelength, as for instance in the crust. Figure 6 shows an example for 2 s shortest period and compares our sampling to regular depth sampling. In the regular sampling case with nearest neighbour interpolation, the phase and envelope errors can be quite large, especially close to the model discontinuities (up to 80 % envelope misfit and 4 % phase misfit as defined by Kristekova et al., 2009) and for very shallow sources (up to 40 % envelope misfit and 14 % phase misfit).

2.3.1 Finite element mapping

One performance-critical step in the spatial scheme is to find the reference coordinates (ξ, η) inside the element that includes a point given in global coordinates (s, z). While the opposite mapping is trivial because this is how the elements of the SEM are defined (Nissen-Meyer et al., 2007a, Appendix A1), it cannot be generally inverted easily. Hua (1990) present an analytical inverse solution for quadrilateral elements, which is quite involved and not easy to generalize to the semicircular elements used in *AxiSEM*.

Instead, we follow a two-step strategy: first, we find the six closest element midpoints. This allows to limit the search to a small number of elements in which the point could

SED
7, 957–1005, 2015

be. The number six is specific to the *AxiSEM* mesh, where each corner point can belong to a maximum of six elements in the doubling layers, see Fig. 7. This step can be seen as approximating the *AxiSEM* mesh with Voronoi cells. For most points, the closest midpoint will already indicate the correct element, in the worst case the second step has to be performed for all six candidates.

In a second step, the reference coordinates (ξ, η) of the given point are computed for the six candidate elements sorted by the distance of the midpoints. If both ξ and η are in the interval $[-1, 1]$, the element is found. (ξ, η) are computed using an iterative gradient scheme adopted from *SPECFEM3D* (Komatitsch and Tromp, 2002b). Starting from the midpoint of the candidate element, updated values are found by linear approximation of the mapping:

$$\begin{pmatrix} \xi_{n+1} \\ \eta_{n+1} \end{pmatrix} = \begin{pmatrix} \xi_n \\ \eta_n \end{pmatrix} + \mathcal{J}^{-1}(\xi_n, \eta_n) \cdot \begin{pmatrix} s_\mathrm{p} - s(\xi_n, \eta_n) \\ z_\mathrm{p} - z(\xi_n, \eta_n) \end{pmatrix} \tag{2}$$

with the Jacobian matrix defined as

$$\mathcal{J}(\xi, \eta) = \begin{pmatrix} \partial_\xi s & \partial_\eta s \\ \partial_\xi z & \partial_\eta z \end{pmatrix}, \tag{3}$$

and the mapping $s(\xi, \eta)$ and $z(\xi, \eta)$ depending on the element type as defined in Nissen-Meyer et al. (2007b). In the *AxiSEM* mesh, this iteration converges to numerical accuracy within less then ten iterations and is not performance critical for *Instaseis* as it is only used on the few candidate elements. Also, this two step approach requires only the midpoints of all elements in the mesh to be read from file on initialization and can be implemented efficiently using the kd-tree provided by the SciPy package (Jones et al., 2001).

2.4 The temporal scheme

The design of the temporal scheme is guided by a number of constraints on the spectrum of the source time function: the spectrum should decay steep enough above

SED
7, 957–1005, 2015

the highest frequency resolved by the mesh, such that the least number of samples according to the Nyquist criterion can be used without introducing aliasing. On the other hand, it should not decay too steeply, such that it is still possible to deconvolve and convolve with another source time function. Additionally, the spectrum should be as flat as possible within the usable frequency range and "earthquake-like" without the necessity of deconvolving it when extracting a seismogram from the database. An actual delta function as would be required for true Green's functions cannot be represented in a discrete approximation as it is not bandlimited.

We found a Gaussian source time function with $\sigma = \tau/3.5$ to fulfill these requirements, where τ is the shortest period resolved by the mesh. Figure 8 shows the amplitude spectra of this source time function as well as a corresponding velocity seismogram at a distance of 40°. The two spectra have a very similar general shape and decay to 10^{-3} of the maximum at half the shortest period. This motivates that sampling with four samples per period will not introduce aliasing artifacts. It is desirable to retrieve seismograms from the database with arbitrary time steps, which requires interpolation or resampling. Popular time domain schemes such as interpolation by low order polynomials or splines do not work well close to the Nyquist frequency. On the other hand, frequency domain resampling by zero-padding the discrete Fourier transform of the signal can only resample to rational multiples of the original sampling interval. Finally, the kernel from the theoretically exact reconstruction according to the Nyquist-Shannon sampling theorem (i.e. the sinc function) has infinite support which renders it impractical as well.

Therefore, we adopt the Lanczos resampling scheme, which is popular in image processing and an approximation to the sinc-resampling with finite support. The Lanczos kernel is defined as the sinc function multiplied by the Lanczos window function (Burger and Burge, 2009):

$$L(t) = \begin{cases} \mathrm{sinc}(t)\,\mathrm{sinc}(t/a) & \text{if } t \in [-a, a] \\ 0 & \text{else,} \end{cases} \tag{4}$$

SED
7, 957–1005, 2015

where a is a parameter to control the number of samples to be used in the interpolation and the sinc function is defined as

$$\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}. \tag{5}$$

Interpolation is then performed by convolving the discrete signal s_i with this kernel and evaluating it at the new timesamples t_j (Burger and Burge, 2009):

$$S(t_j) = \sum_{i=\lfloor x \rfloor - a + 1}^{\lfloor x \rfloor + a} s_i L(t_j - i), \tag{6}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Figure 9 shows the Lanczos kernel for different values of a, Fig. 10 shows a practical example of resampling a seismogram. In Fig. 11 we test a number of values for a for the first 1800 s of the same seismogram and we find $a = 12$ to be a reasonable compromise between cost (using 25 samples in the interpolation) and accuracy (RMS error of 0.03 %).

3 Python API

Instaseis is implemented as a library for the Python programming language; some performance critical parts are written in Fortran. Furthermore it directly integrates with the ObsPy package (Megies et al., 2011; Beyreuther et al., 2010) and utilizes the Python bindings to NetCDF 4 (Rew and Davis, 1990). This enables it to take advantage of the strong scientific Python ecosystem built on top of the SciPy Stack (Jones et al., 2001).

Reasons for choosing Python include its growing popularity in the sciences and it being easy to learn and use while still sufficiently powerful for complex scenarios. It is particularly well suited for big data applications and the integration with web services and databases which suits the potential use cases of *Instaseis*.

SED
7, 957–1005, 2015

Figure 12 shows how to use the Python API in the most simple case. *Instaseis* provides an object oriented interface: in addition to the shown `Source` and `Receiver` classes it furthermore provides `ForceSource` and `FiniteSource` objects. These can also be created by providing data in most commonly used file formats like StationXML, QuakeML, and the Standard Rupture Format. Please refer to the *Instaseis* documentation for further details (www.instaseis.net). Combining and integrating these features enables the construction of modern and clean workflows to solve new problems. A big advantage of this approach is that no temporary files need to be created and the synthetic seismograms can be extracted from the databases on demand when and where they are needed.

The Python API furthermore implements a client/server approach for remote *Instaseis* database access over HTTP. This enables organizations to host high-frequency databases and serve them to users over the internet. This eliminates the need and upfront cost to calculate, store, and distribute *Instaseis* databases for most users while still offering enough performance for many use cases. The Python interface is data-source agnostic; from a usage perspective it thus does not matter if the databases are available locally or via the internet.

Instaseis is developed with a test-driven approach utilizing continuous integration. It is well documented, has a high test coverage, and we intend to maintain it for the next couple of years providing a solid foundation for future applications built on top of it. It will be licensed free and open source for non-profit use, the source code and issue tracker are hosted on GitHub.

4 Benchmarks

4.1 Accuracy

As we already provided some rigorous validation comparing *AxiSEM* synthetics to a reference solution (*Yspec* Al-Attar and Woodhouse, 2008) in van Driel and Nissen-

SED
7, 957–1005, 2015

Meyer (2014a, b), the purpose of this section is only to confirm that using the new scheme with reciprocal computations, spatial interpolation and temporal resampling does not decrease accuracy. Figure 13 shows a record section and some details for *Instaseis*, *AxiSEM* and *Yspec* seismograms computed in the anisotropic, visco-elastic *PREM* model for an event at 126 km depth beneath Tonga bandpass filtered to 50 to 2 s period.

While this figure is similar and the *AxiSEM* and *Yspec* reference data actually the same as presented in van Driel and Nissen-Meyer (2014b, Fig. 11), it is important to note that they were generated in very different ways: here we computed a whole Green's function database for all epicentral distances and down to 700 km source depth and changing source or receivers would cost a few milliseconds only. In our previous approach, this would have required a full new *AxiSEM* simulation on the order of 10 K CPU hours computational cost. Also, in contrast to van Driel and Nissen-Meyer (2014b), we used default mesh parameters for 2 s period and time step close to the stability limit of the 4th order symplectic time scheme (Nissen-Meyer et al., 2008). Still, the phase misfit (Kristekova et al., 2009) is well below 1 % in all zoom windows and the maximum of the envelope misfit is 2 % for the PPP phase on station ALE.

The fact that these traces are virtually indistinguishable for such a demanding setup of wave propagation over 800 wavelengths verifies that the entire workflow of computing and interrogating the database are correctly implemented. In particular, numerical reciprocity (i.e. the different force and moment sources), on-the-fly calculation of the strain tensor as well as temporal and spatial sampling have no significant adverse effect on accuracy, i.e. any remaining errors vanish within numerical accuracy of the forward solver *AxiSEM*.

4.2 Database size

One major constraint for computing a database beside the CPU cost is the permanent storage requirement. Here, we summarize the most important parameters and the related scaling of the required disk space. The amount of data scales with the third

SED
7, 957–1005, 2015

power of the highest frequency resolved by the mesh, but zip compression is slightly more efficient for longer traces, resulting in an empirical exponent of 2.7, see Fig. 14. Scaling with the length of the seismograms is slightly stronger than linear, again because the compression is more efficient on the zeros before the first P arrival. Scaling with depth and epicentral distance range is linear, where the prefactor for depth scaling is halved at each doubling layer of the mesh. The databases for vertical (40 %) and horizontal (60 %) components are computed and therefore usable independently.

Several examples are shown in Fig. 14: for Earth, a complete reciprocal database including all three components, all epicentral distances and sources down to 700 km and one hour of seismogram length accurate down to 2 s period is about 1 TB in size. Calculating them once and storing them on a central server will give any user arbitrary and immediate access to short period synthetic seismograms without any further cost. More specialized databases are possible: for example to study inner core phases for shallow events in an epicentral distance from 140 to 160°, 200 GB storage suffices to store a database with a frequency of 2 Hz.

4.3 Performance

To evaluate the overall performance of *Instaseis*, two distinct parts have to be analyzed: first, the databases have to be generated with *AxiSEM*. Though very efficient, the database generation at short periods is a high performance computing task. However, *AxiSEM* scales well on up to 10 000 cores which allows to do the computations even at the highest frequencies within hours on a supercomputer. Detailed performance and scaling tests of *AxiSEM* can be found in Nissen-Meyer et al. (2014), here we just show the total CPU time required to compute full databases (i.e. horizontal and vertical component) for 1 h long seismograms for two different time schemes (2nd order Newmark and 4th order symplectic Nissen-Meyer et al., 2008) and two planets (Earth and Mars) at a variety of resolutions, see Fig. 15. The general scaling of *AxiSEM* is proportional to T^{-3}, where T is the shortest period resolved by the mesh. The slight discrepancy from this power law at longer periods is due to the thin crustal

SED
7, 957–1005, 2015

layers causing a smaller global time step in the simulation. Simulations for Mars are approximately a factor five faster than for Earth, due to the smaller radius.

The performance of the second part, the seismogram extraction, on the other hand is rarely limited by raw computing power. It scales linearly with increasing frequency of the databases' Green's functions and can easily be accomplished on a standard laptop. The limiting factor in most cases is the latency of the storage system, e.g. the time until it starts reading from the database. To alleviate this issue we implemented partial memoization on the functions reading data from the files. The Green's functions from a whole element of the numerical grid are read once and cached in memory. If data from the same element is needed again at a later stage it will already be in memory thus avoiding repeated disc access. Once the cache memory limit is reached, the data with the earliest last access time is deallocated, effectively resulting in a priority queue sorted by last access time. This optimization is very effective for most common use cases as they oftentimes require seismograms in a small range of epicentral distances and depths.

Instaseis comes with a number of integrated benchmarks to judge its performance for a certain database on a given system. The benchmarks emulate the computational requirements and data access patterns of some typical use cases like finite source simulations and source parameter inversions. Finite sources within the benchmarks are simulated by calculating waveforms for moment tensor sources on an imaginary fault plane along the equator ranging from the surface to a depth of 25 km. One source is calculated for each kilometer in depth until the bottom of the fault is reached. This is repeated each kilometer along the fault's surface trajectory until the benchmark terminates. A source parameter inversion is simulated by calculating seismograms from moment tensor sources randomly scattered within 50 km distance to a fixed point. Results for four runs are shown in Fig. 16. As is the case with all benchmarks they have to be interpreted carefully, nonetheless they demonstrate the behaviour and performance characteristics of *Instaseis* on real machines.

SED
7, 957–1005, 2015

5 Applications

In this section we depict several possible use cases of *Instaseis*. This list is not exhaustive and deliberately unconnected to provide a broad overview.

5.1 Graphical user interface

To prominently highlight the features and nearly instantaneous seismogram extraction for arbitrary source and receiver combinations of *Instaseis*, we developed a cross-platform graphical user interface (GUI), shown in Fig. 17. It ships with the standard *Instaseis* package and is written in PyQt, a Python wrapper for the Qt toolkit.

Most evidently, this may be used for visual inspection and verification of any given *AxiSEM* Green's function database. *Instaseis'* performance permits an immediate visual feedback to changing parameters. This also delivers quantitative insight for an intuitive understanding of the features and parameter sensitivities of seismograms. Examples of this are the polarity flips of first arrivals when crossing a moment tensor's nodal planes, the triplication of phases for shallow sources, the Hilbert transformed shape of reflected phases and the relative amplitude of surface waves (especially overtones) depending on the earthquake depth. Furthermore, the GUI allows the calculation of seismograms from finite sources and the exploration of waveform differences in comparison to best-fitting points sources.

5.2 IRIS web-interface

To enable usage of *Instaseis* seismograms to a broader community, we aim to remove all hurdles of computing and storing large databases locally. To this end, and in collaboration with IRIS, we plan to establish a webinterface to the *Instaseis* databases. In contrast to the *ShakeMovie* approach (Tromp et al., 2010), this interface will be able to handle arbitrary sources and receivers independent from catalogue data or other

SED
7, 957–1005, 2015

parameter limitations. The interface and databases will be described and benchmarked in detail in a separate publication.

5.3 Finite frequency tomography

In finite frequency tomography (e.g. Nolet, 2008) information is extracted by comparing observed to synthetic seismograms, for example by using matched filters based on cross-correlation (Sigloch and Nolet, 2006; Colombi et al., 2014). For body waves, short periods down to 1 s are commonly used (e.g. Stähler et al., 2012; Hosseini and Sigloch, 2015). Typical datasets contain thousands of earthquakes (e.g. Auer et al., 2014), each recorded at hundreds of stations, resulting in up to a million waveforms. For wave propagation solvers that solve the forward problem for each event separately, this presents a formidable computational challenge, which is why previous studies resorted to approximate solutions like WKBJ (Chapman, 1978) or the reflectivity method (Fuchs and Müller, 1971). *Yspec* (Al-Attar and Woodhouse, 2008) is about an order of magnitude faster than *AxiSEM* in computing seismograms for a single source, however, the cost scales linearly with the number of events.

With *Instaseis*, we can now build the whole database for all possible sources with only two runs of *AxiSEM*. Figure 18 compares the computational cost of computing the reference synthetics down to 2 s period assuming that each event was recorded at 1000 three component stations. Ignoring the cost of computing the database, *Instaseis* is comparable in performance to WKBJ, but actually returns full seismograms including all phases, see Fig. 19. In contrast to WKBJ, where each crustal reverberation has to be defined separately, it automatically calculates the full crustal response. Also, it appropriately models diffracted phases such as *Pdiff* and triplicated phases from upper mantle discontinuities. If we include the database generation, *Instaseis* breaks even in computational cost with *Yspec* already at about 14 000 waveforms, i.e. five events with 1000 three component stations each. At about 5×10^8 waveforms, the cost of extracting the seismograms from the database becomes dominant over the database generation. Assuming 2000 seismograms per event, this is equivalent to 10 000 earthquakes, i.e.

SED
7, 957–1005, 2015

in the order of available earthquake catalogues. However, generating seismograms with different source locations or moment-tensor radiation patterns, which is often necessary in tomography, does not require a new database generation.

5.4 Probabilistic source inversion

Uncertainties in source parameters have been shown to have a strong influence on waveform tomography (Valentine and Woodhouse, 2010). Probabilistic point source inversion estimates the uncertainties of source parameters and their correlation. From these, the effect on seismic tomography can be estimated (Stähler and Sigloch, 2014). It requires the repeated calculation of synthetic waveforms for varying moment tensors, depths and source time functions to calculate the likelihood and posterior probability density of models in a Bayesian sense. Changing source time function and moment tensor is extremely efficient from an *Instaseis* perspective, and the limitation to a fixed epicenter means that the I/O buffering can be done very efficiently, which is reflected in the *Source Inversion* testcase in the benchmark (Fig. 13).

From a previous study (Stähler and Sigloch, 2014), we assume that for an inversion for depth, the moment tensor and the source time function, a 20-dimensional model space has to be sampled, which requires to perform roughly 60 000 forward simulations. Using 100 seismic stations and three-component seismograms, this means that roughly 1.8×10^7 waveforms have to be calculated for one source inversion, costing on the order of 50–100 CPU hours (Fig. 18).

5.5 Finite sources

Finite sources can be represented in *Instaseis* by a cloud of point sources without limitations on the fault geometry or source time functions. Each point source needs to be attached with a moment tensor, a sliprate function and a time shift relative to the origin time. These can for instance be retrieved from a standard rupture file (.srf). As a show case, we computed the seismograms for the source inversion validation (SIV)

SED
7, 957–1005, 2015

exercise #3 (http://equake-rc.info). The source is a $M = 7.8$ strike-slip earthquake on the San Andreas Fault represented by $\approx 10^5$ point sources, where each source has a different mechanism and sliprate function. The 52 stations are in 30–90° epicentral distance (see Fig. 20), where the *P* wave arrival is well separated. Excluding the cost of generating the database, it costs a total of 12 CPU hours to compute the 52 one hour long three component seismograms accurate down to 5 s.

Figure 21 compares the *Instaseis* seimograms to *P* phases computed with the frequency wavenumber integration method (fk) by Kikuchi and Kanamori (1982), where only direct and surface reflected phases where taken into account. The largest differences for the closer stations (left panel) are due to additional phase arrivals within the time window. However, also for the more distant stations, the differences remain larger than the differences caused by a local 1-D crustal model. Besides the computational performance, this suggests that more accurate waveforms should be beneficial for finite source inversions.

5.6 Insight/Mars

The upcoming NASA-lead Mars Insight mission (Banerdt et al., 2013), to be launched in March 2016 and scheduled to land September 2016, will deploy a single station with both a broad-band and a short period seismometer on Mars. This will be the first extra-terrestrial seismic mission since the Apollo lunar landings (1969–1972) and Mars Viking missions (1975) with the goal of elucidating the interior structure of a planet other than Earth. The instrument will record local, regional, and more distant marsquakes, including meteorite impacts, and send data back to Earth for analysis.

Our knowledge of the seismic structure of Mars is limited, because of lack of resolution of currently available areophysical data (e.g., Khan and Connolly, 2008) and the limited sensitivity of the Viking seismometers due to their installation on board of the lander. For this reason, we will generate databases of "reference" seismic waveforms for a comprehensive collection (order of magnitude 1000) of 1-D Martian models to be used by modelers and analysts in preparation for the Insight mission. The models

SED
7, 957–1005, 2015

are constructed from current areophysical data (mean mass, mean moment of inertia, tidal Love number, and tidal dissipation) and thermodynamic modeling methods and summarize our current understanding of the internal constitution of the planet. *AxiSEM* and hence *Instaseis* can readily be used to propagate waves on Mars, see Fig. 22, allowing us to build these databases very efficiently.

5.7 Ambient seismic noise

As mentioned in Sect. 2.3, seismograms generated by force sources can be extracted from the same reciprocal databases. This is particularly interesting for studying ambient seismic noise. By cross-correlating noise recorded at two stations, using long enough time series and under certain assumptions (uncorrelated, isotropically distributed white noise sources), it is possible to retrieve the Green's function of the medium between the two stations (e.g. Sanchez-Sesma, 2006; Gouédard et al., 2008). However, not all of these assumptions are met in nature, e.g. the noise sources are not evenly distributed (Tsai, 2009; Froment et al., 2010; Basini et al., 2013). Also, the noise sources themselves are not yet well understood, especially with respect to the generation of Love waves in the microseismic band (Nishida et al., 2008).

Instaseis provides a basis to quickly generate noise synthetics to study such effects, which we illustrate in Fig. 23. We computed noise cross correlations, accurate down to a period of 5 s, for a total of 20 days of noise data generated with 100 000 noise sources. The calculation only took 1 CPU hour. In the first case, the noise sources consist of vertical forces with a random source-time function, all have the same amplitude and are distributed evenly on the globe. The resulting cross correlation is in good agreement with the Green's function, which is obtained by introducing an impulse source at each of the stations in Zurich and Munich. In the second case, sources are located in the oceans only, their amplitude proportional to the significant wave height (Gualtieri et al., 2013). For the two stations located in Zurich and Munich the close sources are thus solely located in the west, which leads to strong asymmetry in the retrieved correlations

SED
7, 957–1005, 2015

(Stehly et al., 2006). *Instaseis* thus enables users to study noise on the global scale across the microseismic band, by generating realistic waveforms at negligible cost.

6 Conclusions and outlook

In this paper we presented a readily available methodology and code to extract seismograms for spherical earth models from a Green's function database. High efficiency in the generation of databases and very fast extraction (on the order of milliseconds per seismogram) of highly accurate seismograms (indistinguishable from conventional forward solvers) can then replace previously employed approximations such as WKBJ, reflectivity or frequency-wavenumber integration methods that were used for computational reasons in many applications of global seismology. *Instaseis* code is open source and available with extensive documentation at www.instaseis.net.

Future developments include Cartesian local domains with layered models, which are not yet supported by *AxiSEM*. As a large fraction of earthquakes are located below oceans and receivers on continents, it may be beneficial for body waves studies to take advantage of the axisymmetric capability of *AxiSEM* and place the receiver on a circular 'island' of continental crust within a global oceanic crustal model.

The Supplement related to this article is available online at doi:10.5194/sed-7-957-2015-supplement.

Author contributions. M. van Driel and L. Krischer implemented *Instaseis*, M. van Driel, S. C. Stähler and T. Nissen-Meyer continuously develop *AxiSEM* and added the database output, and K. Hosseini prepared the finite frequency example. M. van Driel prepared the manuscript with contributions from all co-authors.

Acknowledgements. We thank Alex Hutko and Chad Trabant (IRIS) for valuable discussions on the database selection and implementation as well as user demand. Amir Khan provided

SED
7, 957–1005, 2015

the Mars model and Olaf Zielke the reference data used in Fig. 21. Celine Hadziioannou supported us to design the noise example. We gratefully acknowledge support from the European Commission (Marie Curie Actions, ITN QUEST, www.quest-itn.org) and the EU-FP7 VERCE project (number 283543). Data processing and downloading was done using ObsPy (Megies et al., 2011; Beyreuther et al., 2010) and ObsPyLoad (Scheingraber et al., 2013). Computations where performed at the ETH central HPC cluster (Brutus), the Swiss National Supercomputing Center (CSCS), the UK National Supercomputing Service (HECToR) and the Leibniz Supercomputing Center (LRZ), whose support is gratefully acknowledged.

References

Al-Attar, D. and Woodhouse, J. H.: Calculation of seismic displacement fields in self-gravitating earth models – applications of minors vectors and symplectic structure, Geophys. J. Int., 175, 1176–1208, doi:10.1111/j.1365-246X.2008.03961.x, 2008. 959, 967, 972

Astiz, L., Earle, P., and Shearer, P.: Global stacking of broadband seismograms, Seismol. Res. Lett., 67, 8–18, doi:10.1785/gssrl.67.4.8, 1996. 983

Auer, L., Boschi, L., Becker, T. W., Nissen-Meyer, T., and Giardini, D.: Savani: a variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, J. Geophys. Res.-Sol. Ea., 119, 3006–3034, doi:10.1002/2013JB010773, 2014. 972

Banerdt, W. B., Smrekar, S., Lognonné, P., Spohn, T., Asmar, S. W., Banfield, D., Boschi, L., Christensen, U., Dehant, V., Folkner, W., Giardini, D., Goetze, W., Golombek, M., Grott, M., Hudson, T., Johnson, C., Kargl, G., Kobayashi, N., Maki, J., Mimoun, D., Mocquet, A., Morgan, P., Panning, M., Pike, W. T., Tromp, J., van Zoest, T., Weber, R., Wieczorek, M. A., Garcia, R., and Hurst, K.: InSight: A Discovery Mission to Explore the Interior of Mars, in: 44th Lunar Planet. Sci. Conf., The Woodlands, Texas, 18–22 March, Abstract Number 1915, 2013. 974

Basini, P., Nissen-Meyer, T., Boschi, L., Casarotti, E., Verbeke, J., Schenk, O., and Giardini, D.: The influence of nonuniform ambient noise on crustal tomography in Europe, Geochem. Geophy. Geosy., 14, 1471–1492, doi:10.1002/ggge.20081, 2013. 975

SED
7, 957–1005, 2015

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J.: ObsPy: a Python toolbox for seismology, Seismol. Res. Lett., 81, 530–533, doi:10.1785/gssrl.81.3.530, 2010. 966, 977

Burger, W. and Burge, M.: Principles of Digital Image Processing: Core Algorithms, Springer, London, 2009. 965, 966

Chapman, C. H.: A new method for computing synthetic seismograms, Geophys. J. Int., 54, 481–518, doi:10.1111/j.1365-246X.1978.tb05491.x, 1978. 958, 972

Colombi, A., Nissen-Meyer, T., Boschi, L., and Giardini, D.: Seismic waveform inversion for core-mantle boundary topography, Geophys. J. Int., 198, 55–71, doi:10.1093/gji/ggu112, 2014. 972

Crotwell, H. P., Owens, T. J., and Ritsema, J.: The TauP Toolkit: flexible seismic travel-time and ray-path utilities, Seismol. Res. Lett., 70, 154–160, 1999. 958

Friederich, W. and Dalkolmo, J.: Complete synthetic seismograms for a spherically symmetric earth by a numerical computation of the Green's function in the frequency domain, Geophys. J. Int., 122, 537–550, doi:10.1111/j.1365-246X.1995.tb07012.x, 1995. 959

Froment, B., Campillo, M., Roux, P., Gouédard, P., Verdel, A., and Weaver, R. L.: Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations, Geophysics, 75, SA85–SA93, doi:10.1190/1.3483102, 2010. 975

Fuchs, K. and Müller, G.: Computation of synthetic seismograms with the reflectivity method and comparison with observations, Geophys. J. Roy. Astr. S., 23, 417–433, 1971. 959, 972

Geller, R. J. and Ohminato, T.: Computation of synthetic seismograms and their partial derivatives for heterogeneous media with arbitrary natural boundary conditions using the direct solution method, Geophys. J. Int., 421–446, doi:10.1111/j.1365-246X.1994.tb01807.x, 1994. 959

Gouédard, P., Stehly, L., Brenguier, F., Campillo, M., Colin de Verdière, Y., Larose, E., Margerin, L., Roux, P., Sánchez-Sesma, F. J., Shapiro, N. M., and Weaver, R. L.: Cross-correlation of random fields: mathematical approach and applications, Geophys. Prospect., 56, 375–393, doi:10.1111/j.1365-2478.2007.00684.x, 2008. 975

Gualtieri, L., Stutzmann, E., Capdeville, Y., Ardhuin, F., Schimmel, M., Mangeney, A., and Morelli, A.: Modelling secondary microseismic noise by normal mode summation, Geophys. J. Int., 193, 1732–1745, doi:10.1093/gji/ggt090, 2013. 975

Holtzman, B., Candler, J., Turk, M., and Peter, D.: Seismic Sound Lab: sights, sounds and perception of the Earth as an acoustic space, in: Sound, Music, Motion, edited by:

SED
7, 957–1005, 2015

Aramaki, M., Derrien, O., Kronland-Martinet, R., and Ystad, S. L., Springer, Marseille, 161–174, 2013. 961

Hosseini, K. and Sigloch, K.: Finite frequency measurements of core-diffracted P-waves (Pdiff), Geophys. J. Int., submitted, 2015. 972

Hua, C.: An inverse transformation for quadrilateral isoparametric elements: analysis and application, Finite Elem. Anal. Des., 7, 159–166, 1990. 963

Igel, H., Takeuchi, N., Geller, R. J., Megnin, C., Bunge, H.-P., Clévédé, E., Dalkolmo, J., and Romanowicz, B.: The COSY Project: verification of global seismic modeling algorithms, Phys. Earth Planet. In., 119, 3–23, doi:10.1016/S0031-9201(99)00150-8, 2000. 958

Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: open source scientific tools for Python, available at: http://www.scipy.org/, last access: 23 February 2015, 2001. 964, 966

Kawai, K., Takeuchi, N., and Geller, R. J.: Complete synthetic seismograms up to 2 Hz for transversely isotropic spherically symmetric media, Geophys. J. Int., 164, 411–424, doi:10.1111/j.1365-246X.2005.02829.x, 2006. 959

Khan, A. and Connolly, J. A. D.: Constraining the composition and thermal state of Mars from inversion of geophysical data, J. Geophys. Res., 113, E07003, doi:10.1029/2007JE002996, 2008. 974

Kikuchi, M. and Kanamori, H.: Inversion of complex body waves, B. Seismol. Soc. Am., 72, 491–506, 1982. 959, 974, 1003

Komatitsch, D. and Tromp, J.: Spectral-element simulations of global seismic wave propagation – II. Three-dimensional models, oceans, rotation and self-gravitation, Geophys. J. Int., 150, 303–318, doi:10.1046/j.1365-246X.2002.01716.x, 2002a. 958

Komatitsch, D. and Tromp, J.: Spectral-element simulations of global seismic wave propagation – I. Validation, Geophys. J. Int., 149, 390–412, doi:10.1046/j.1365-246X.2002.01653.x, 2002b. 964

Kristekova, M., Kristek, J., and Moczo, P.: 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, 2009. 963, 968, 988, 995

Masson, Y., Cupillard, P., Capdeville, Y., and Romanowicz, B.: On the numerical implementation of time-reversal mirrors for tomographic imaging, Geophys. J. Int., 196, 1580–1599, doi:10.1093/gji/ggt459, 2013. 961

SED
7, 957–1005, 2015

Megies, T., Beyreuther, M., Barsch, R., Krischer, L., and Wassermann, J.: ObsPy – what can it do for data centers and observatories?, Ann. Geophys.-Italy, 54, 12 pp., doi:10.4401/ag-4838, 2011. 966, 977

Montagner, J. and Kennett, B. L. N.: How to reconcile body-wave and normal-mode reference earth models, Geophys. J. Int., 125, 229–248, doi:10.1111/j.1365-246X.1996.tb06548.x, 1996. 963, 988

Monteiller, V., Chevrot, S., Komatitsch, D., and Fuji, N.: A hybrid method to compute short-period synthetic seismograms of teleseismic body waves in a 3-D regional model, Geophys. J. Int., 192, 230–247, doi:10.1093/gji/ggs006, 2012. 961

Nishida, K., Kawakatsu, H., Fukao, Y., and Obara, K.: Background Love and Rayleigh waves simultaneously generated at the Pacific Ocean floors, Geophys. Res. Lett., 35, L16307, doi:10.1029/2008GL034753, 2008. 975

Nissen-Meyer, T., Dahlen, F. A., and Fournier, A.: Spherical-earth Fréchet sensitivity kernels, Geophys. J. Int., 168, 1051–1066, doi:10.1111/j.1365-246X.2006.03123.x, 2007a. 959, 961, 963

Nissen-Meyer, T., Fournier, A., and Dahlen, F. A.: A two-dimensional spectral-element method for computing spherical-earth seismograms – I. Moment-tensor source, Geophys. J. Int., 168, 1067–1092, doi:10.1111/j.1365-246X.2006.03121.x, 2007b. 959, 960, 961, 962, 964

Nissen-Meyer, T., Fournier, A., and Dahlen, F. A.: A 2-D spectral-element method for computing spherical-earth seismograms – II. Waves in solid-fluid media, Geophys. J. Int., 174, 873–888, doi:10.1111/j.1365-246X.2008.03813.x, 2008. 959, 968, 969

Nissen-Meyer, T., van Driel, M., Stähler, S. C., Hosseini, K., Hempel, S., Auer, L., Colombi, A., and Fournier, A.: AxiSEM: broadband 3-D seismic wavefields in axisymmetric media, Solid Earth, 5, 425–445, doi:10.5194/se-5-425-2014, 2014. 959, 960, 969, 984

Nolet, G.: A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun, Cambridge University Press, Cambridge, UK, 2008. 972

Nyquist, H.: Certain topics in telegraph transmission theory, Transactions of the A.I.E.E., 617–644, doi:10.1109/5.989875, 1928. 991

Rew, R. and Davis, G.: NetCDF: an interface for scientific data access, IEEE Computer Graphics & Applications, 10, 76–82, 1990. 960, 966

Sanchez-Sesma, F. J.: Retrieval of the Green's function from cross correlation: the canonical elastic problem, B. Seismol. Soc. Am., 96, 1182–1191, doi:10.1785/0120050181, 2006. 975

SED
7, 957–1005, 2015

Scheingraber, C., Hosseini, K., Barsch, R., and Sigloch, K.: ObsPyLoad: a tool for fully automated retrieval of seismological waveform data, Seismol. Res. Lett., 84, 525–531, doi:10.1785/0220120103, 2013. 977

Sigloch, K. and Nolet, G.: Measuring finite-frequency body-wave amplitudes and traveltimes, Geophys. J. Int., 167, 271–287, doi:10.1111/j.1365-246X.2006.03116.x, 2006. 972, 1001

Stähler, S. C. and Sigloch, K.: Fully probabilistic seismic source inversion – Part 1: Efficient parameterisation, Solid Earth, 5, 1055–1069, doi:10.5194/se-5-1055-2014, 2014. 973

Stähler, S. C., Sigloch, K., and Nissen-Meyer, T.: Triplicated P-wave measurements for waveform tomography of the mantle transition zone, Solid Earth, 3, 339–354, doi:10.5194/se-3-339-2012, 2012. 972

Stehly, L., Campillo, M., and Shapiro, N. M.: A study of the seismic noise from its long-range correlation properties, J. Geophys. Res., 111, B10306, doi:10.1029/2005JB004237, 2006. 976

Tolman, H.: User manual and system documentation of WAVEWATCH-III version 3.14, Tech. Rep. 276, NOAA/ National Weather Service, MD, 2009. 1005

Tromp, J.: Theory and observations – forward modeling and synthetic seismograms: 3-D numerical methods, in: Treatise on Geophysics, Amsterdam, Elsevier B.V., 192–217, 2007. 958

Tromp, J., Komatitsch, D., Hjörleifsdóttir, V., Liu, Q., Zhu, H., Peter, D., Bozdag, E., McRitchie, D., Friberg, P., Trabant, C., and Hutko, A.: Near real-time simulations of global CMT earthquakes, Geophys. J. Int., 183, 381–389, doi:10.1111/j.1365-246X.2010.04734.x, 2010. 958, 959, 971

Tsai, V. C.: On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, Geophys. J. Int., 178, 1555–1564, doi:10.1111/j.1365-246X.2009.04239.x, 2009. 975

Valentine, A. P. and Woodhouse, J. H.: Reducing errors in seismic tomography: combined inversion for sources and structure, Geophys. J. Int., 180, 847–857, doi:10.1111/j.1365-246X.2009.04452.x, 2010. 973

van Driel, M. and Nissen-Meyer, T.: Seismic wave propagation in fully anisotropic axisymmetric media, Geophys. J. Int., 199, 880–893, doi:10.1093/gji/ggu269, 2014a. 959, 967

van Driel, M. and Nissen-Meyer, T.: Optimized viscoelastic wave propagation for weakly dissipative media, Geophys. J. Int., 199, 1078–1093, doi:10.1093/gji/ggu314, 2014b. 959, 968

SED
7, 957–1005, 2015

Table 1. I/O performance for a typical setup of *AxiSEM* on SuperMUC. The simulation parameters were: 2 s shortest period, 3600 s simulation length, model: *ak135f*, vertical component, maximum source depth 700 km. The resulting uncompressed wavefield file has a size of 675 GB. The I/O throughput is not affected much by the number of CPUs involved. The throughput between different runs varies, which is probably caused by the changing I/O load on the system.

#CPUs	Runtime	I/O time	Throughput	Rel. I/O time
4624	1091 s	196 s	3.44 GB s^{-1}	18.0 %
2304	1802 s	281 s	2.40 GB s^{-1}	15.6 %
1152	2359 s	167 s	4.04 GB s^{-1}	7.0 %
576	4482 s	193 s	3.50 GB s^{-1}	4.3 %

SED
7, 957–1005, 2015

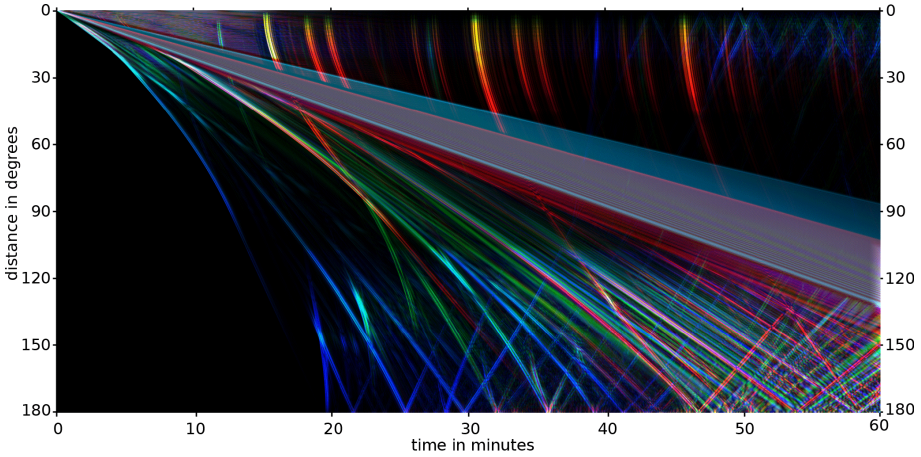

Figure 1. Global stack of 1 h of seismograms accurate to a shortest period of 2 s for an earthquake in 27 km depth computed with *Instaseis*. The displacement is color-coded analogous to the IRIS global stack (Astiz et al., 1996), i.e. red = transversal component, green = radial component, blue = vertical component. An automatic gain control (*AGC*) with a window of 100 s length is used to balance large amplitude variations between the various phases. Note that creating this plot does not require to define the source depth at the time of database calculation. A high resolution version of this plot and one for 5 h long seismograms is added as Supplement.

SED
7, 957–1005, 2015

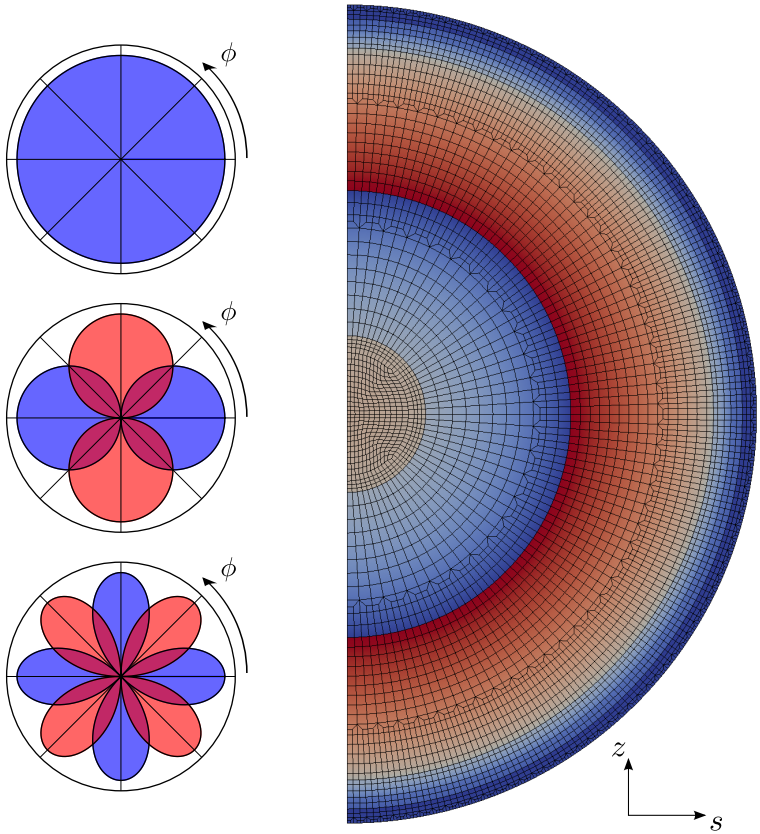

Figure 2. The 3-D wavefield is decomposed analytically into monopole, dipole and quadrupole radiation patterns (left) and the remaining 2-D problem is solved on a D-shaped domain (right) using the spectral element method. While the forward databases require a total of four 2-D computations, it is only two for the backward databases using reciprocity of the Green's function: one for the vertical and one for the horizontal components (modified from Nissen-Meyer et al., 2014).

SED
7, 957–1005, 2015

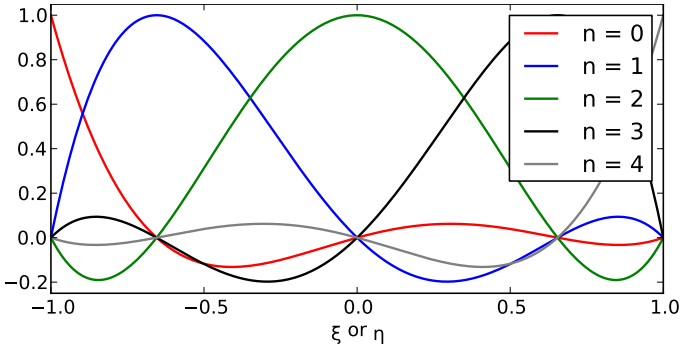

Figure 3. Lagrangian basis polynomials $l_n(\xi)$ of fourth order in one dimension. At the collocation points, all but one are zero, such that the value of the interpolated function at this point coincides with the coefficient in this basis expansion.

SED
7, 957–1005, 2015

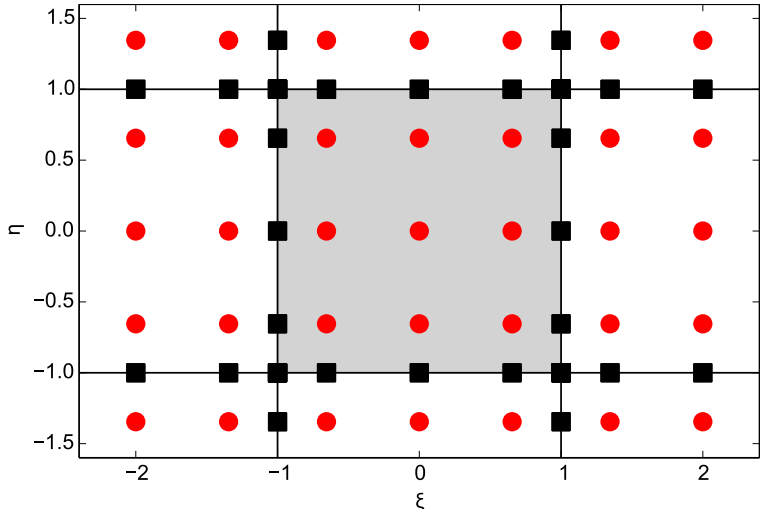

Figure 4. Lagrange interpolation points inside an element (gray) and its neighbours. Coordinates ξ and η are the reference coordinates of the gray element. Points on the edges (black squares) are shared between neighbours and function values at these points need to be stored only once if the function is continuous (e.g. displacement). The number of global degrees of freedom per element of such functions is thus approximately 16 compared to 25 for discontinuous functions (e.g. strain).

SED
7, 957–1005, 2015

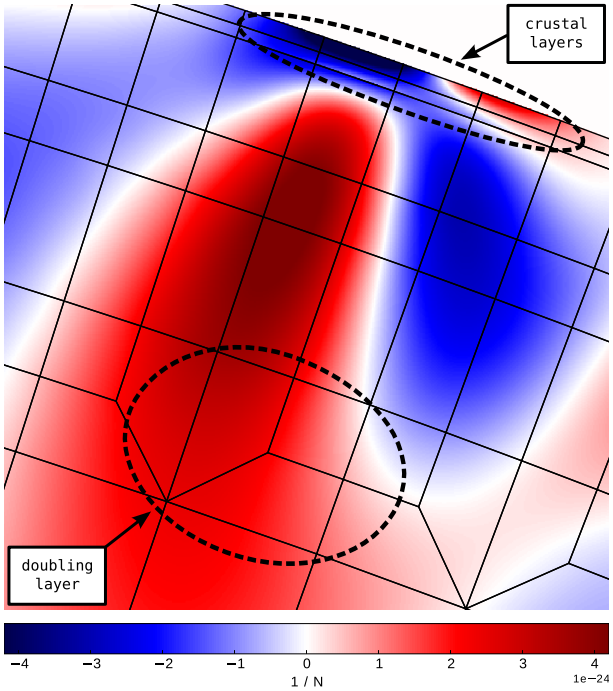

Figure 5. Snapshot of one component of the Green's tensor ($G_{rr,r}$) as represented in the SEM basis for a shortest period of 50 s. Discontinuities such as caused by the crustal layers are exactly represented and the wavefield is smooth across doubling layers of the mesh.

SED
7, 957–1005, 2015

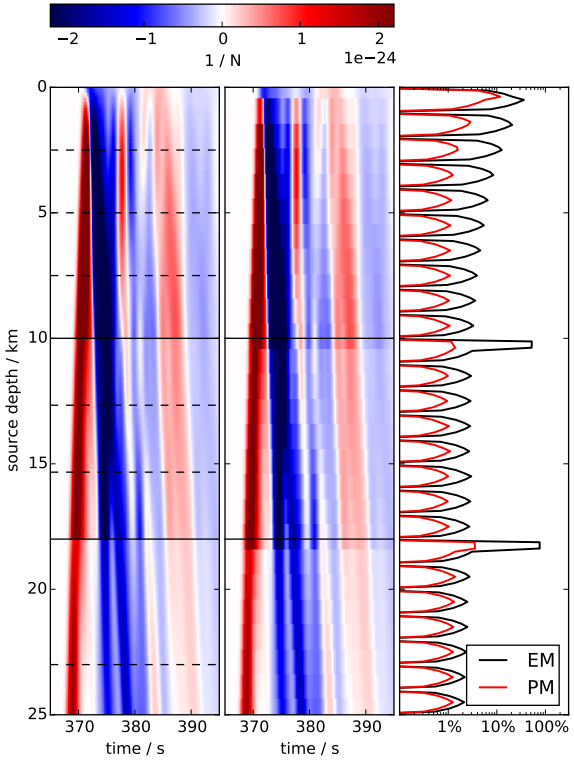

Figure 6. One component of the strain Green's tensor ($G_{rr,r}$) for a distance of 30° as a function of time and depth with a shortest period of 2 s. SEM basis (left) vs. regular sampling with 1 km distance (middle), phase and envelope misfits (EM and PM) as defined by Kristekova et al. (2009) computed in the period range 1–20 s. Dashed lines in the left panel sketch the spectral elements. The crustal discontinuities of *ak135f* (Montagner and Kennett, 1996) are located at 10 and 18 km depth (solid lines).

SED
7, 957–1005, 2015

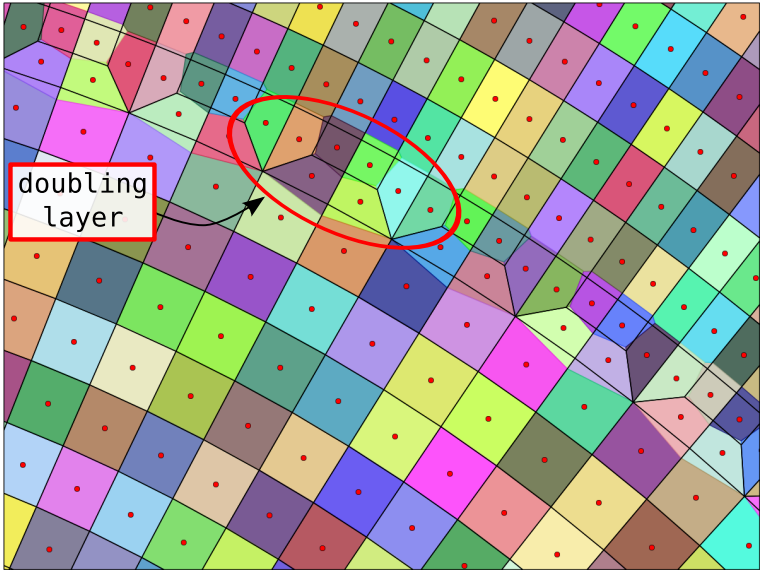

Figure 7. Voronoi approximation (colored) of the *AxiSEM* mesh (black lines) using the midpoints of the elements (red circles) only, zoomed onto a doubling layer for a 50 s mesh. For most elements, the Voronoi cell coincides almost exactly with the *AxiSEM* element, note that most of the *AxiSEM* elements have edges of concentric circles while the edges of the Voronoi cells are all straight lines. In the worst case six *AxiSEM* elements have to be tested whether a point is inside or not.

SED
7, 957–1005, 2015

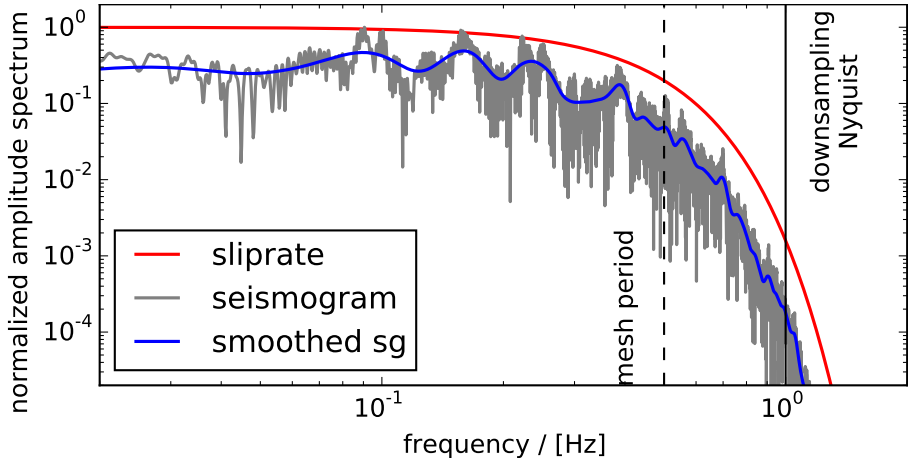

Figure 8. Normalized amplitude spectra of the Gaussian source time function (sliprate) used at 2 s mesh period and a vertical component synthetic seismogram recorded at 40° epicentral distance. The vertical lines denote the resolution of the mesh and the Nyquist frequency of the downsampling using 4 samples per mesh period.

SED
7, 957–1005, 2015

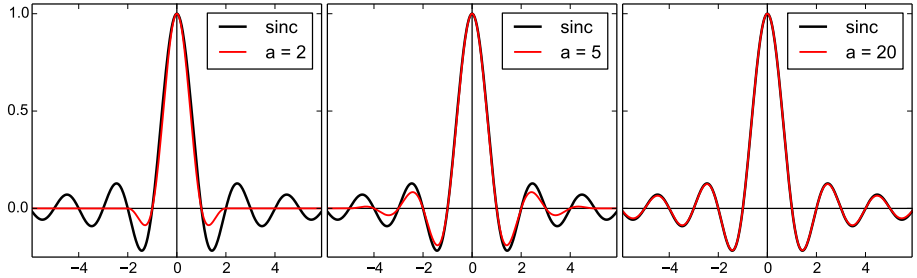

Figure 9. Lanczos kernels used for resampling. For large values of the parameter *a*, it converges towards the sinc function, which is the kernel that allows exact reconstruction for bandlimited signals as stated in the Nyquist sampling theorem (Nyquist, 1928).

SED
7, 957–1005, 2015

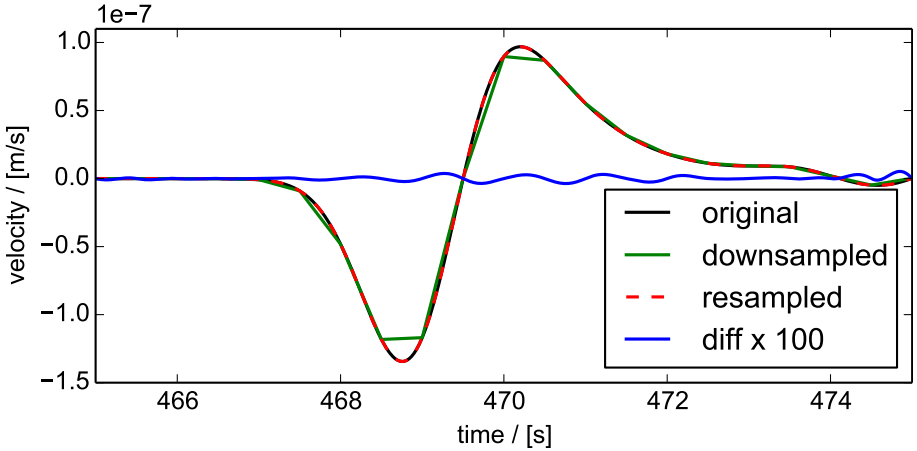

Figure 10. Resampling using a Lanczos kernel with $a = 12$ of a P arrival velocity seismogram convolved with the source time function from Fig. 8 recorded at 40° distance, and the resampling error multiplied by a factor of 100. The relative error is on the order of 0.05 % of RMS, see Fig. 11.

SED
7, 957–1005, 2015

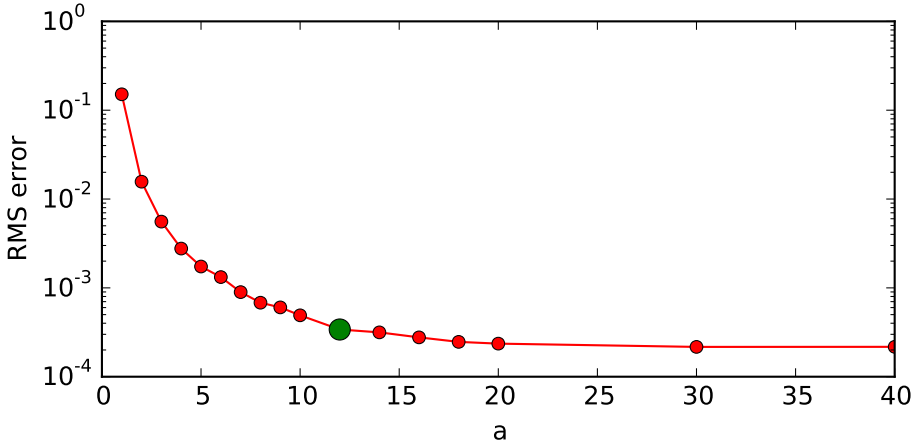

Figure 11. RMS error of the resampling using the first 1800 s of the seismogram from Fig. 10. Convergence is reached around $a = 20$. It does not converge to zero, because some high frequency energy was negelected in the downsampling, see Fig. 8.

SED
7, 957–1005, 2015

```
>>> import instaseis
>>> db = instaseis.open_db("./AK135")
>>> receiver = instaseis.Receiver(network="BW", station="ZUGS",
                                  latitude=47.416, longitude=10.979)
>>> source = instaseis.Source(
        latitude=89.91, longitude=0.0, depth_in_m=12000,
        m_rr = 4.71e+17, m_tt = 3.81e+15, m_pp =-4.74e+17,
        m_rt = 3.99e+16, m_rp =-8.05e+16, m_tp =-1.23e+17)
>>> st = db.get_seismograms(source=source, receiver=receiver)
>>> print(st)
3 Trace(s) in Stream:
BW.ZUGS..LXZ | 1970-01-01T00:00:00.00Z - ... | 2.1 Hz, 7746 samples
BW.ZUGS..LXN | 1970-01-01T00:00:00.00Z - ... | 2.1 Hz, 7746 samples
BW.ZUGS..LXE | 1970-01-01T00:00:00.00Z - ... | 2.1 Hz, 7746 samples
```

Figure 12. The *Instaseis* Python API demonstrated in a short interactive Python session. A `Source` and a `Receiver` object are created and then passed to the `get_seismograms()` method of an `InstaseisDB` object. This will extract the Green's functions from the databases and perform all necessary subsequent steps resulting in directly usable three component seismograms in form of an ObsPy `Stream` object. Please refer to the *Instaseis* documentation for details.

SED
7, 957–1005, 2015

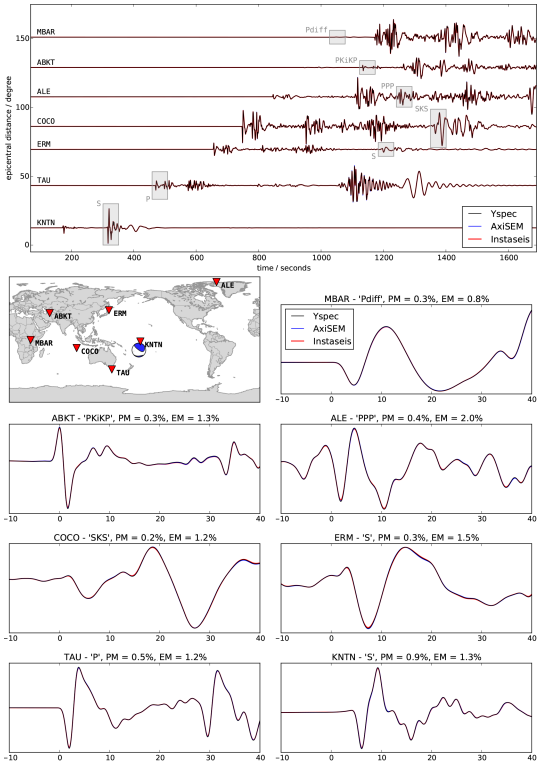

Figure 13. Comparison of vertical displacement seismograms (bandpass filtered from 50 to 2 s period) for a moment magnitude $M_w = 5.0$ event in 126 km depth under the Tonga islands, computed with *Instaseis*, *AxiSEM* and *Yspec* in the anisotropic *PREM* model without ocean but including attenuation. The traces are recorded at the GSN stations indicated in the map. The zoom windows are depicted with gray rectangles in the record section and the time scale is relative to the ray-theoretical arrival. *EM* and *PM* (Kristekova et al., 2009) denote the envelope and phase misfit between *Instaseis* and *Yspec* traces in the corresponding time window.

SED
7, 957–1005, 2015

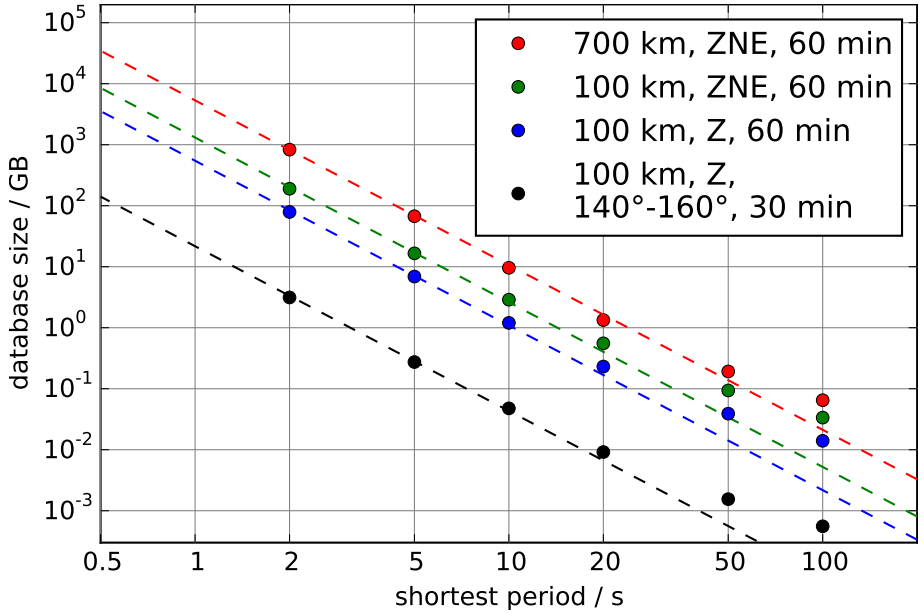

Figure 14. Storage requirements of the reciprocal databases for *PREM* after zip compression for all three components and several parameter sets (maximum source depth, components, seismogram length and epicentral distance range). Dashed lines are fitted functions $g(f) = a\,f^{2.7}$ with frequency f. The exponent is slightly smaller than the expected 3, because the zip compression is more efficient for longer time traces. At long periods, element sizes are governed by the layer thickness rather than the wavelength, resulting in the discrepancy from the power law at long periods.

SED
7, 957–1005, 2015

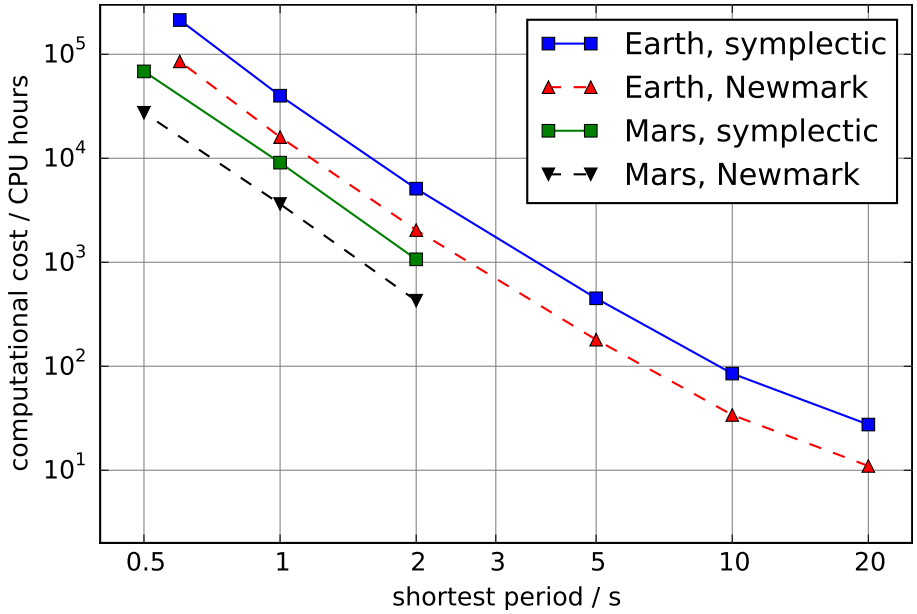

Figure 15. Computational cost in CPU hours (measured on *Monte Rosa*, a Cray XE6, for Earth and *Piz Daint*, a Cray XC30, for Mars) to generate full *Instaseis* databases with 1 h long seismograms for two time schemes: 2nd order Newmark and 4th order symplectic.

SED
7, 957–1005, 2015

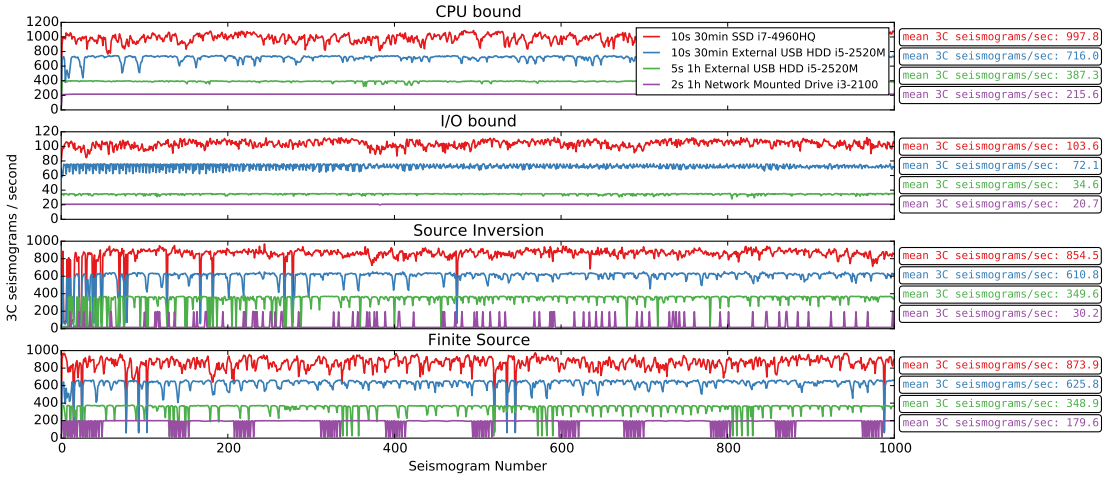

Figure 16. Results of benchmarks for four typical use cases run on different hardware with a variety of shortest periods. The graphs show the inverse time for the calculation of the ith three-component seismogram. Each run calculated 1000 three-component seismograms, is repeated ten times with the same random seed, the top and bottom values are discarded, and the mean of the remaining eight values is plotted. The CPU and I/O bound scenarios illustrate the speed with a fully efficient and a deactivated cache, respectively. The two bottom scenarios emulate real use cases, see the the main text for details. Amongst other things they show the consequence of a too small cache in the source inversion scenario for the 2 s run and the efficiency of the buffer in the finite source scenario for the same database.

SED
7, 957–1005, 2015

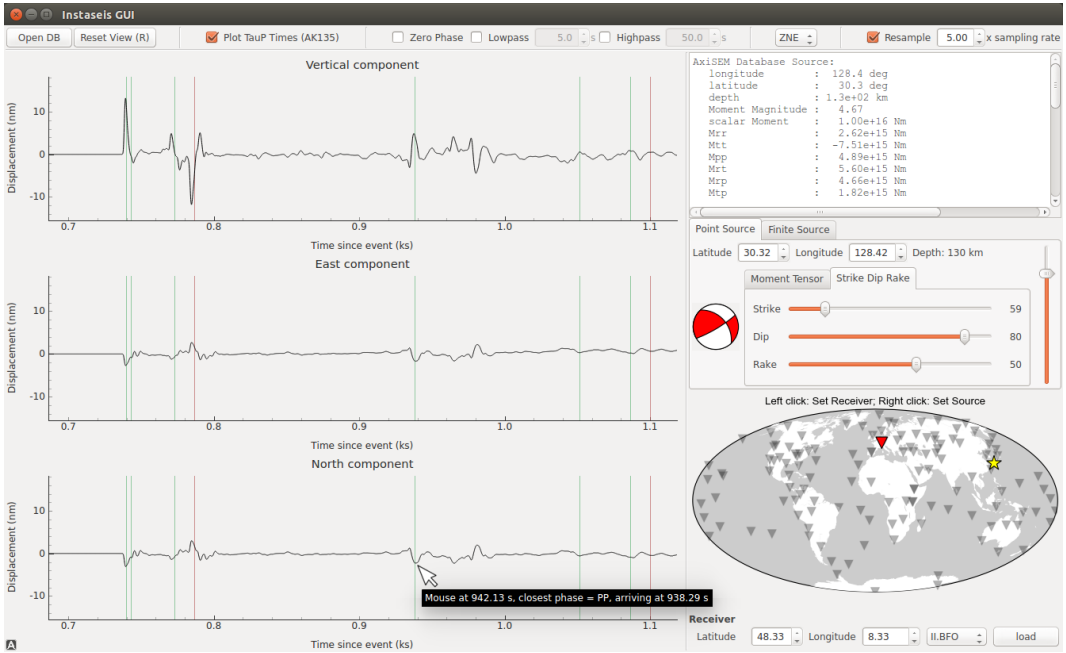

Figure 17. Screenshot of the *Instaseis* graphical user interface (GUI). Aside from quickly exploring the characteristics of a given Green's function database it is a great tool for understanding and teaching many aspect of seismograms. The speed of *Instaseis* enables an immediate visual response to changing source and receiver parameters. The left hand side shows three component seismograms where theoretical arrival times of various seismic phases are overlaid as vertical lines. The bar at the top is used to change filter and resampling settings and the section on the right side is used to modify source and receiver parameters.

SED
7, 957–1005, 2015

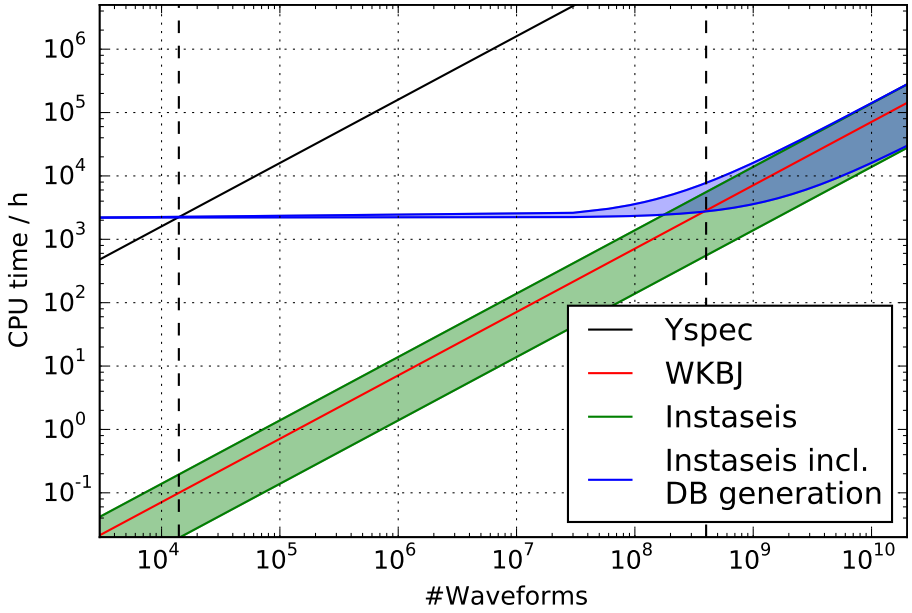

Figure 18. Computational cost to compute many synthetic seismograms for finite frequency tomography with a shortest period of 2 s using different methods. For *Yspec* we assume that for every source there are 1000 receivers with 3 components each. The shaded regions for *Instaseis* indicate the dependence of the performance on the actual source receiver distribution, compare Fig. 16. Including the cost to generate the database with *AxiSEM*, *Instaseis* breaks even with *Yspec* for 14 000 waveforms, that is equivalent to about 5 sources in this configuration.

SED
7, 957–1005, 2015

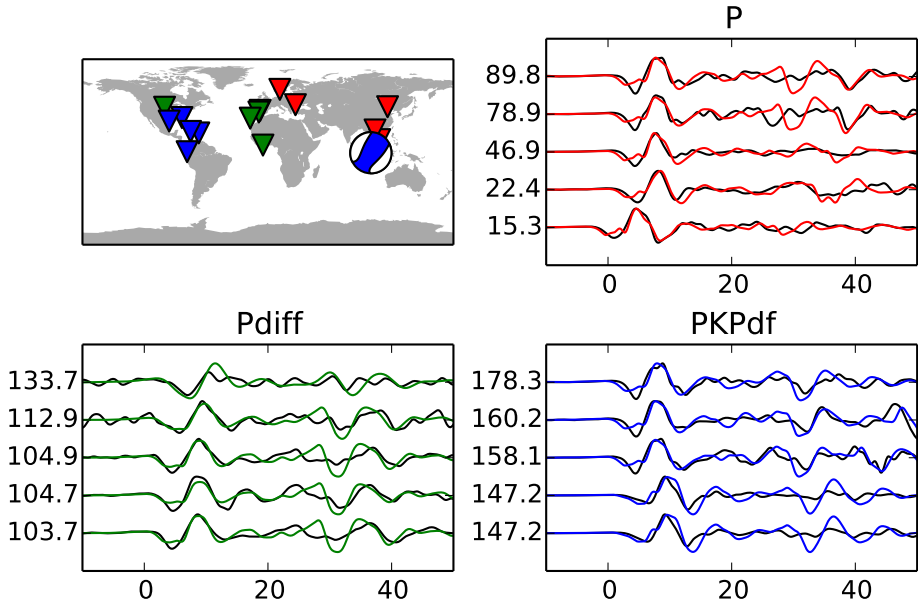

Figure 19. Comparison between observed seismograms (black) and *Instaseis* synthetics for the Sumatra earthquake on 30 September 2009 with magnitude M_w = 7.5 at 82 km depth. Vertical axis labels are epicentral distances, horizontal is time relative to the ray-theoretical arrivals. A Gabor filter with 3.7 s central period is applied to all traces and the synthetics are convolved with an inverted source time function (Sigloch and Nolet, 2006). The waveforms are aligned by computing relative time-shifts between data and synthetic seismograms by cross-correlation (similar to actual finite-frequency tomography).

SED
7, 957–1005, 2015

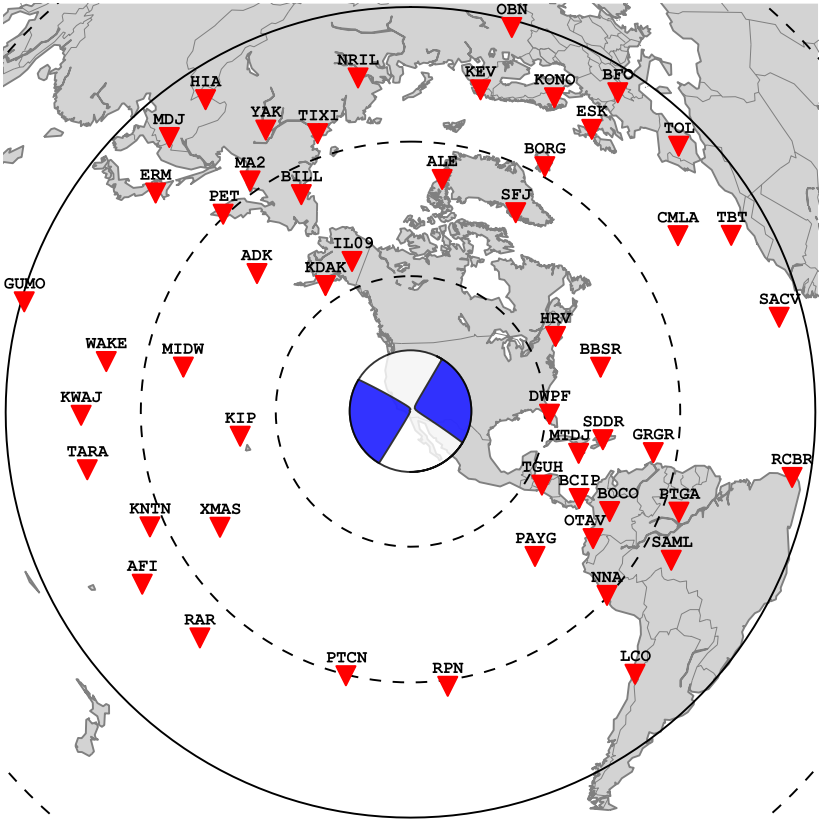

Figure 20. Stations used in the Source Inversion Validation (SIV) exercise. Circles mark 30, 60 and 90° epicentral distance. The source is a strike-slip earthquake in Southern California represented by $\approx 10^5$ point sources.

SED
7, 957–1005, 2015

Instaseis: instant global seismograms

M. van Driel et al.

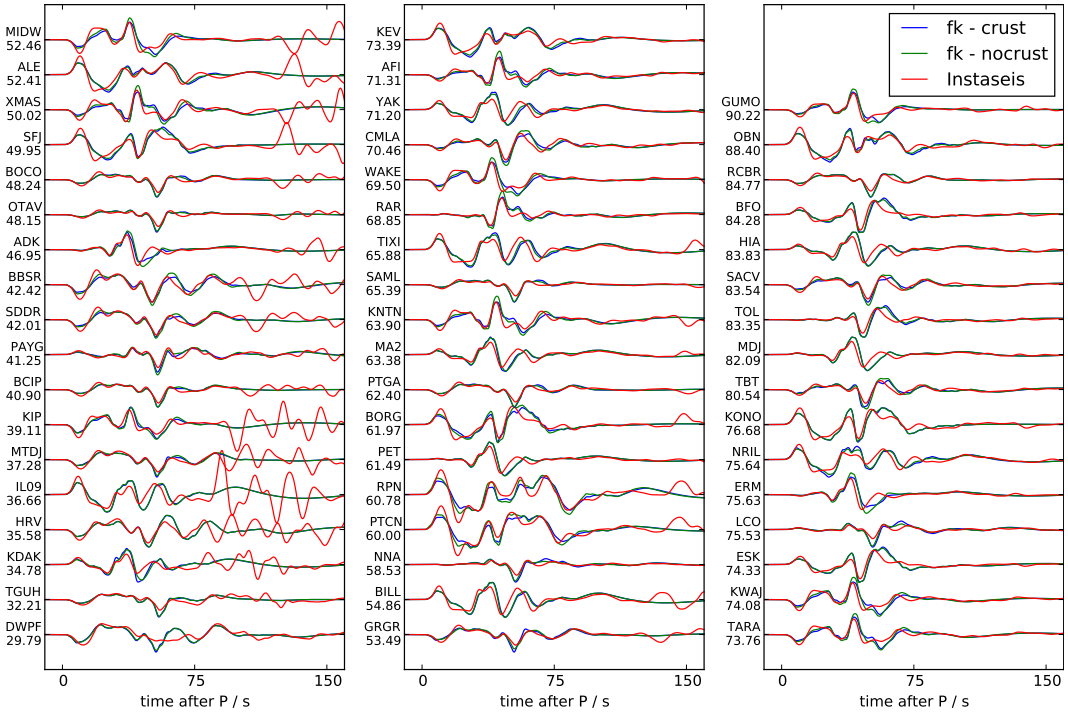

Figure 21. Seismograms for the SIV benchmark, *Z* component aligned on the *P* arrival. The labels denote the station code and epicentral distance. In the frequency-wavenumber integration (*fk*, Kikuchi and Kanamori, 1982), only direct *P* and the depth phases were included, while *Instaseis* provides full seismograms. The differences between *fk* and *Instaseis* are generally larger than the effect of the crust, suggesting that using full waveforms might be beneficial for finite source inversion.

SED
7, 957–1005, 2015

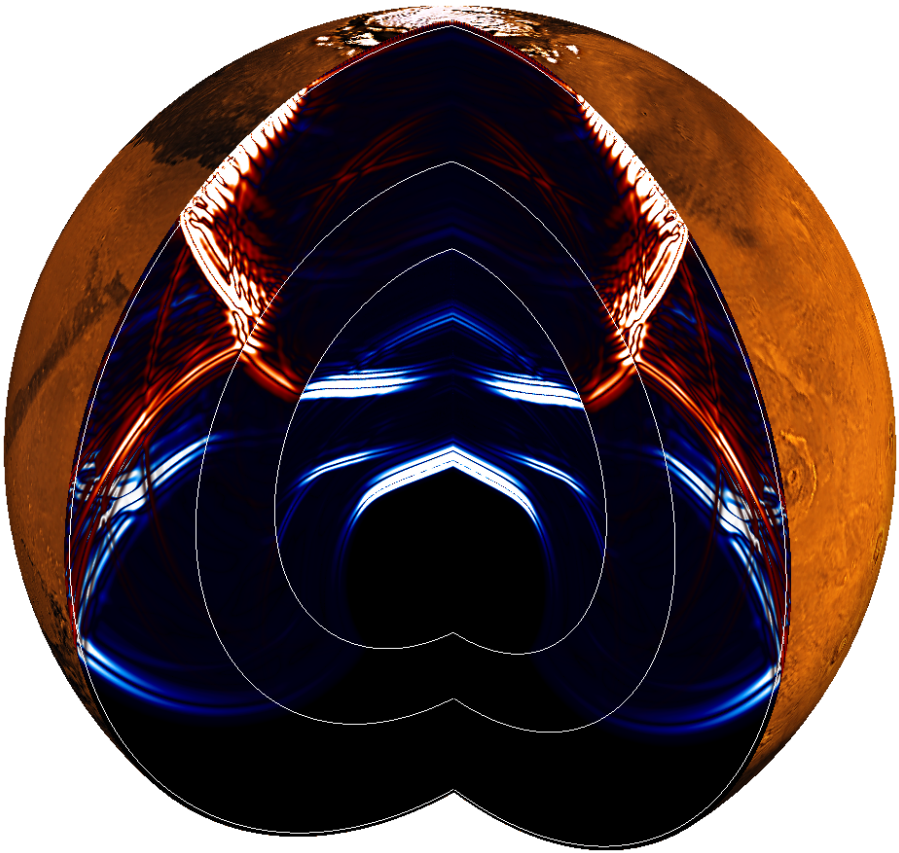

Figure 22. Seismic waves travelling in Mars after a meteorite impact at its north pole computed with *AxiSEM*. P waves are shown in blue and S waves and surface waves in red.

SED
7, 957–1005, 2015

Instaseis: instant global seismograms

M. van Driel et al.

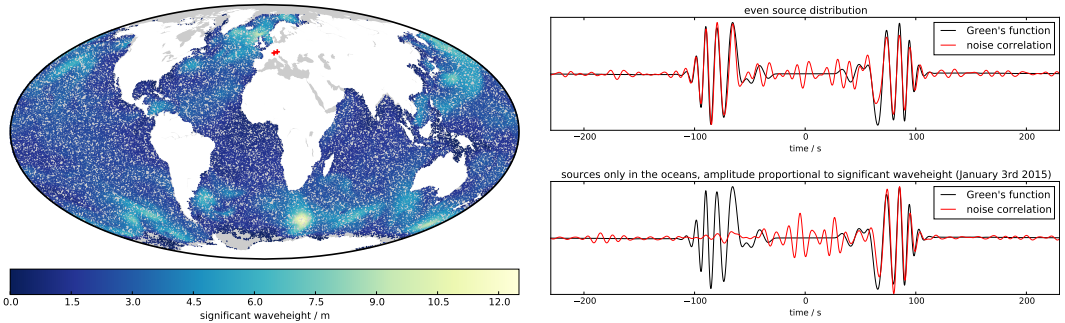

Figure 23. Ambient seismic noise cross correlations computed with *Instaseis*. Left: 100 000 vertical force sources located in the oceans and amplitude proportional to the significant wave height from the NOAA WAVEWATCH III model on 3 January 2015 (Tolman, 2009). Red crosses indicate the receivers located in Munich and Zurich. Right: cross correlations of 20 days of noise for (top) evenly distributed noise sources and (bottom) the sources in the map, the traces are normalized to their maximum amplitude.

SED
7, 957–1005, 2015