

3-D scattering of elastic waves by small-scale heterogeneities in the Earth's mantle

K. Leng,^{1,2} J. Korenaga² and T. Nissen-Meyer¹

¹Department of Earth Science, University of Oxford, South Parks Road, Oxford OX1 3AN, UK. E-mail: kuangdai.leng@earth.ox.ac.uk

²Department of Geology & Geophysics, Yale University, 210 Whitney Avenue, New Haven, CT 06511, USA

Accepted 2020 June 29. Received 2020 June 21; in original form 2020 January 2

SUMMARY

Small-scale heterogeneities in the Earth's mantle, the origin of which is likely compositional anomalies, can provide critical clues on the evolution of mantle convection. Seismological investigation of such small-scale heterogeneities can be facilitated by forward modelling of elastic wave scattering at high frequencies, but doing so with conventional 3-D numerical methods has been computationally prohibitive. We develop an efficient approach for computing high-frequency synthetic wavefields originating from small-scale mantle heterogeneities. Our approach delivers the exact elastodynamic wavefield and does not restrict the geometry or physical properties of the local heterogeneity and the background medium. It combines the technique of wavefield injection and a numerical method called AxiSEM3D. Wavefield injection can decompose the total wavefield into an incident and a scattered part. Both these two parts naturally have low azimuthal complexity and can thus be solved efficiently using AxiSEM3D under two different coordinate systems. With modern high-performance computing (on an order of magnitude of 10^5 CPU-hr), we have achieved a 1 Hz dominant frequency for global-scale problems with strong deep Earth scattering. Compared with previous global injection approaches, ours allows for a 3-D background medium and yields the exact solution without ignoring any higher-order scattering by the background medium. Technically, we develop a traction-free scheme for realizing wavefield injection in a spectral element method, which brings in several flexibilities and simplifies the implementation by avoiding stress or traction computation on the injection boundary. For a spherical heterogeneity in the mid-lower mantle, we compare the 3-D full-wave solution with two approximate ones obtained, respectively, by the perturbation theory and in-plane (axisymmetric) modelling. As a comprehensive application, we study S-wave scattering by a 3-D ultra-low velocity zone, incorporating 3-D crustal structures on the receiver side as part of the background model.

Key words: Composition and structure of the mantle; Computational seismology; Theoretical seismology; Wave propagation; Wave scattering and diffraction.

1 INTRODUCTION

Lateral heterogeneities in the Earth's mantle and their roles in mantle dynamics have long been discussed in solid Earth geophysics. Seismic tomography, as one of the primary probings of the Earth's interior, has exhibited good agreement on the predominance of long-wavelength heterogeneities in the mantle (Becker & Boschi 2002; Auer *et al.* 2014), such as the two large low-shear-velocity provinces (LLSVP; Zhao 2004; Garnero *et al.* 2016) atop the core–mantle boundary (CMB), stagnant or deep-thrusting subducted slabs (Sigloch *et al.* 2008; Fukao & Obayashi 2013) and deep-rooted mantle plumes (Montelli *et al.* 2006; French & Romanowicz 2015), manifesting the first-order pattern of large-scale

mantle convection (Silver *et al.* 1988; Zhao 2004; Garnero & McNamara 2008). Apart from these long-wavelength features, a number of seismological studies have indicated the abundance of small-scale heterogeneities across the entire mantle (Hedlin *et al.* 1997; Kaneshima 2016). Such heterogeneities have been associated with subducted oceanic plates that have undergone various processes such as convective mixing (Kellogg & Turcotte 1987; van Keken *et al.* 2002), crustal segregation (Christensen & Hofmann 1994; Karato 1997), phase transition (Ringwood & Irifune 1988; Bina *et al.* 2001) and partial melting (Defant & Drummond 1990; Peacock *et al.* 1994). Constraining the distribution, morphology and strength of such small-scale heterogeneities, in turn, can provide insights into the nature of these complex processes. In addition to

subducted slabs, isolated small patches known as the ultra-low velocity zones (ULVZ) have been discovered on the CMB (Garnero & Helmberger 1995; Helmberger *et al.* 2000), commonly within or near the edges of a LLSVP (Yu & Garnero 2018), whose origin has been ascribed primarily to partial melting or some mantle heterogeneity reacting with the core (Williams & Garnero 1996; Garnero & McNamara 2008). Assessing the geometry and seismic properties of the ULVZs can help constraining their chemical identity and understanding the evolution of the LLSVPs or thermochemical piles in the context of deep mantle flow (Thorne *et al.* 2013; Li *et al.* 2017).

Forward modelling can play a fundamental role in seismological studies of small-scale mantle heterogeneities. First, it can help revealing the physics of wave scattering as a basis for data interpretation. In this paper, the term ‘wave scattering’ generally means that an elastic wavefield is altered by a heterogeneity such as due to reflection, refraction or diffraction. As opposed to the weak, long-wavelength mantle inhomogeneity that smoothly alters the travel-time and amplitude of an incident body wave, a strong, localized heterogeneity can involve more complicated wave scattering phenomena, usually with new seismic phases generated upon multiple wave encounters on its boundaries. Such scattering effects can depend non-linearly on the location, geometry and physical properties of the heterogeneity, leading to elusive observations on the surface. With the help of forward modelling, one can zoom into the scattering processes and explain their surface expressions, such as for narrow-conduit plumes (Hwang *et al.* 2011; Maguire *et al.* 2016), subducted slabs (Haugland *et al.* 2017) and ULVZs (Rondenay *et al.* 2010; Cottaar & Romanowicz 2012; Thorne *et al.* 2013; Vanacore *et al.* 2016). Secondly, physically realistic synthetic waveforms are required to justify existing inversion methodologies. Broad-band array data of scattered body waves have been utilized in different ways for structural inversion or imaging. For instance, from a random media viewpoint, stochastic analyses have suggested the presence of small-scale scatterers (on an order of magnitude from 10 to 100 km) throughout the mantle, using various scattered phases such as *P* coda (Shearer & Earle 2004), Pdiff coda (Earle & Shearer 2001) and PKP precursors (Hedlin *et al.* 1997; Margerin & Noclet 2003). Regional distributions of such scatterers may be further constrained by a deterministic location method (Kaneshima & Helffrich 1998; Korenaga 2015; Frost *et al.* 2018), which projects the scattered energy back to the Earth’s interior based on the ray theory. These data-intensive array approaches are persistently challenged by a low signal-to-noise ratio and sparse observations, for which waveform stacking plays an essential role (Rost & Thomas 2002; Korenaga 2013). From a theoretical perspective, nearly every array approach (either stochastic or deterministic) implies the linear perturbation theory or the Born approximation (Hudson & Heritage 1981; Dalkolmo & Friederich 2000) so that the influence of each point scatterer in the interior can be superposed at an observation point on the surface. Because of these challenges, the efficacy of the array approaches must be verified using broadband synthetic data (along with artificial noise) obtained from 3-D waveform modelling. Finally, forward modelling is the basic element of full-waveform inversion techniques such as adjoint tomography (Bozdağ *et al.* 2016) and Bayesian inference with Monte Carlo sampling (Hong & Sen 2009; Gebraad *et al.* 2019). Such inversion techniques seem particularly suitable for an isolated heterogeneity (such as a ULVZ) because they can potentially handle the strong non-linearity involved in wave scattering while the scale of the inverse problem can be restricted by the small target area. Nevertheless, their applications are still hindered due to inadequate computing power for forward modelling.

A fine structural scale necessitates forward modelling at a high frequency. This paper is targeted at a dominant frequency at 1 Hz, capable of resolving a length scale of 2–10 km in the deep mantle. At a high frequency near 1 Hz, comprehensive 3-D numerical methods such as finite difference method (FDM; Igel *et al.* 2002; Moczo *et al.* 2002) and spectral element method (SEM; Komatitsch & Tromp 2002; Afanasiev *et al.* 2019) are not accessible for the foreseeable future (with a cost estimation of 10^8 CPU-hr for one global-scale simulation with a 1-hr seismogram length). Recent forward studies of this kind have been conducted at a period around 10 s (e.g. Cottaar & Romanowicz 2012; Maguire *et al.* 2016). A new hybrid 3-D method, AxiSEM3D (Leng *et al.* 2016, 2019), has been recently developed. By exploiting the azimuthal sparsity of a global wavefield, it delivers a speedup between 2 and 3 orders of magnitude for a 3-D tomographic mantle model near 1 Hz. In the presence of a localized mantle heterogeneity, such a speedup can be maintained with an in-depth tuning of the azimuthal parametrization of the wavefield using a technique called ‘wavefield scanning’ (Leng *et al.* 2019); however, wavefield scanning must start from a trial simulation with an oversampling azimuthal parametrization, which can be too expensive to achieve at a high frequency.

To overcome the insufficiency of computing power, two approximations can be adopted, the perturbation theory and 2-D in-plane modelling: the former simplifies the wave scattering physics and the latter reduces the dimensionality of the heterogeneity. The linear perturbation theory, also known as the Born approximation (Hudson & Heritage 1981; Dalkolmo & Friederich 2000), regards a heterogeneity as an assembly of independent point scatterers and thus ignores any multiscattering effects within the heterogeneity. Consequently, the computation of the scattered waveforms can be simplified as an integral (known as the Love integral) over space and time of the product of three constituents (Dalkolmo & Friederich 2000): the model perturbation, the incident wavefield and the Green’s function. Expensive 3-D wave propagation is avoided because both the incident wavefield and the Green’s function are computed with the background model (usually spherically symmetric). Nevertheless, the Born approximation has good accuracy only for a rather limited parameter space with a weak perturbation strength ($|\delta v| < 5$ per cent) and a subwavelength structural scale (Hudson & Heritage 1981; Wu & Aki 1985; Korneev & Johnson 1993). Though the limit of perturbation strength can be relaxed (to cover $|\delta v| < 50$ per cent) via the ‘long-wave’ extension (Gubernatis 1979; Gritto *et al.* 1995), the requirement of subwavelength structural scales still implies the incapability of resolving structural details. Now we move on to in-plane modelling in 2-D. The 3-D elastodynamic problem can be reduced to 2-D by assuming axisymmetry of a lateral heterogeneity, with its structural variation confined to the source-receiver plane. Under this assumption, several axisymmetric methods have been developed (Igel & Weber 1995; Jahnke *et al.* 2008; Li *et al.* 2014; Nissen-Meyer *et al.* 2014), capable of unravelling the primary in-plane scattering effects at a substantially reduced computational cost. Many recent forward studies of small-scale mantle heterogeneities fall under this category (Rondenay *et al.* 2010; Thorne *et al.* 2013; Vanacore *et al.* 2016; Haugland *et al.* 2017, 2018). Its disadvantage is also clear: a localized 3-D heterogeneity will generate strong off-plane scattered energy (behaving as a secondary source), which in turn can be useful in constraining the lateral dimensions of the heterogeneity (Dalkolmo & Friederich 2000; Cottaar & Romanowicz 2012; Maguire *et al.* 2016; Leng *et al.* 2019). Meanwhile, even within the source-receiver plane, the axisymmetric assumption may cause considerable waveform discrepancies from a full 3-D solution (Leng *et al.* 2019). Some

ray-based corrections have been developed to remedy such discrepancies (Helberger & Ni 2005; Li *et al.* 2014), but available only for weak, long-wavelength inhomogeneity.

Elastic wave scattering by a small-scale heterogeneity in the mantle is essentially a cross-scale problem, that is, scattered wave generation is at a local scale (~ 10 km) but propagation is at a global scale (~ 1000 km). The technique of wavefield injection comes as a cross-scale framework to solve such a cross-scale problem. Under this framework, a subdomain that contains the small-scale heterogeneity is separated from the large, entire domain so that two different numerical methods or discretizations can be used to solve the wave equation inside and outside the subdomain in sequence. The wave propagator used inside the subdomain has to be sufficiently comprehensive (and thus computationally expensive) to capture any 3-D scattering effects in the near field, whereas the one outside is supposed to be highly efficient for wave propagation in the far field based on certain approximations (such as symmetry of the background model). The basic idea may date back to Alterman & Karal (1968), who incorporated a ‘source box’ into a finite difference grid to handle the cross-scale effect of a point source. Since then, such an approach has received broad applications and developments from the finite difference community, as reviewed by Opršal *et al.* (2009). It has often been referred to as the ‘hybrid method’ in the literature because it solves the total wavefield inside and the scattered wavefield outside the subdomain. To avoid any ambiguity (as AxiSEM3D by itself is a hybrid method), we always refer to such a methodology as ‘wavefield injection’, a term introduced by Robertsson & Chapman (2000).

At a continental to global scale, wavefield injection has been realized, combining a comprehensive local solver with a more efficient global solver, for instance, a 2-D FDM with the Kirchhoff method (Wen & Helberger 1998), a 3-D SEM with the direct solution method (Monteiller *et al.* 2012, 2015), a 3-D SEM with the frequency–wavenumber method (Tong *et al.* 2014a, b), a 3-D regional SEM with a 3-D global SEM (Clouzet *et al.* 2018) and a 3-D SEM or FDM with a database of precomputed 1-D Green’s functions (Pienkowska-Cote *et al.* forthcoming). A general theoretical framework along with a guideline for implementation in FDM and SEM can be found in Masson *et al.* (2013) and Masson & Romanowicz (2016). Most of these studies have been focused on strong scattering around the earthquake source or the receivers, with the subdomain located right beneath the Earth’s surface. Using a 2-D SEM, Lin *et al.* (2019) have demonstrated the potential of applying such a concept to deep Earth scattering, who have also provided an inclusive review on the existing implementations in forward modelling and inversion.

It is emphasized that, though wavefield injection is motivated by promoting the computational efficiency for a cross-scale problem, it belongs to the genre of exact wavefield methods with no compromise on the accuracy of solution. From an inverse perspective, applying wavefield injection does not impose any additional limitation on an inverse problem restricted in a localized region of the Earth’s interior (Masson & Romanowicz 2017; Clouzet *et al.* 2018). Further, we clarify two distinctions. First, wavefield injection is distinct from the perturbation theory: though both involve a sequential computation of the incident and the scattered wavefields, the former honours the exact wave scattering physics and thus preserves any non-linear multiscattering effects. Also, it is distinct from another kind of hybrid approaches that couple two computational domains or discretizations during the time loop of an ‘all-in-one’ simulation, for example the ‘sandwich’ approach that couples SEM and normal mode (Capdeville *et al.* 2003a, b) and AxiSEM3D that couples the

spectral element and the pseudospectral discretizations (Leng *et al.* 2016, 2019).

In this paper, we combine wavefield injection and AxiSEM3D to compute the 3-D scattered wavefield originating from a small-scale heterogeneity in the Earth’s mantle at a high frequency up to 1 Hz. Different from any existing implementations of wavefield injection that have involved two different wave propagators, we use AxiSEM3D to solve both the incident and the scattered wavefields. Our approach, however, is not just another implementation with some different wave propagator but a stand-alone, physics-based approach that can maximally exploit the spatiotemporal sparsity of both the incident and the scattered wavefields. This relies on the wavefield adaptivity of AxiSEM3D (Leng *et al.* 2016, 2019), that is, its computational cost can be optimized to the azimuthal complexity of a 3-D wavefield; for instance, it naturally degenerates to an axisymmetric SEM (AxiSEM, Nissen-Meyer *et al.* 2014) for a 1-D earth model and can be a few orders of magnitude faster than a conventional 3-D SEM for a smooth, long-wavelength mantle model. In this sense, the role of wavefield injection can be understood as the decomposition of the total wavefield into an incident and a scattered part, both of low azimuthal complexity (but under two different coordinate systems) for efficient solutions with AxiSEM3D. The most distinguishing feature of our approach is the capability to incorporate a 3-D background model at a high frequency, such as a tomographic mantle model with 3-D crustal structures. Besides, our approach does not involve domain truncation at the local scale and thus preserves any higher-order interactions between the heterogeneity and the background medium. Such interactions can also be honoured by an exact boundary condition (van Manen *et al.* 2007; Masson & Romanowicz 2017).

The rest of this paper is structured as follows. Our approach will be elaborated in the following methodological section, including its theory, implementation and characteristics; notably, we derive a traction-free scheme that can largely facilitate any SEM-based implementation of wavefield injection. In Section 3, we verify our implementation by a nearly error-free benchmark solution against a conventional 3-D SEM. Sections 4 and 5 contain two deep Earth applications, both accomplished at a 1 Hz dominant frequency. In Section 4, we study a spherical heterogeneity in the mid-lower mantle, where we use the 3-D full-wave solution to assess the error of the two approximate solutions obtained, respectively, by the perturbation theory and 2-D in-plane modelling. In Section 5, we carry out a comprehensive case study about a ULVZ, incorporating 3-D crustal structures on the receiver side as part of the background model; we zoom into the scattering process around the ULVZ through wavefield animations, which help us to identify the origin of all scatter phases in the seismograms; next, we study the scattering effects of the 3-D crust and finally use the synthetic data to image the ULVZ by teleseismic migration. Section 6 summarizes the features of our approach and the findings from the numerical examples and ends up with some further discussions related to wavefield extrapolation and source- or receiver-side heterogeneities.

2 METHOD

In this section, we elaborate the theory, implementation and characteristics of our approach. We start from a brief introduction to the fundamentals of AxiSEM3D, with emphasis on the azimuthal complexity of a 3-D wavefield. Next, we revisit the theory of wavefield injection from a new perspective; it realizes the decomposition of the total wavefield into an incident and a scattered part, both of which

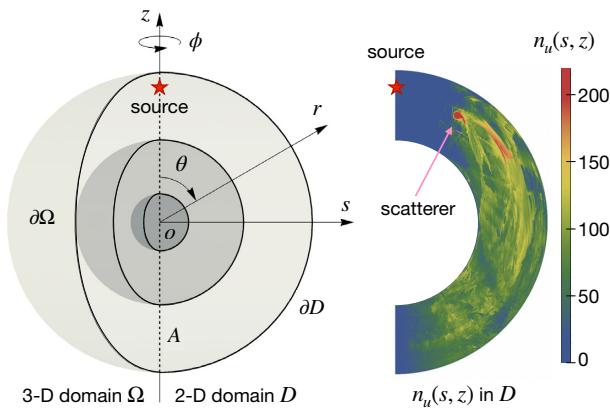


Figure 1. Dimension reduction in AxiSEM3D, adapted from Leng *et al.* (2016). In the source-centred coordinate system (s, ϕ, z) , we reduce the 3-D computational domain Ω to a 2-D meridian domain D through a Fourier series characterization of the ϕ -dimension. At different locations in D , the Fourier series may contain different number of terms, that is, $n_u = n_u(s, z)$, determined by the azimuthal complexity of the wavefield. In this example, we show $n_u(s, z)$ for a PREM model including a spherical heterogeneity (with a radius of 160 km and a 50 per cent velocity reduction, located at 30° in distance and 1200 km in depth), obtained by wavefield scanning (Leng *et al.* 2019) at a 5 s period and with a 1-hr record length.

have low azimuthal complexity and can thus be solved efficiently with AxiSEM3D. Based on this reformulation, we design a new traction-free scheme for implementing wavefield injection in any 2-D or 3-D SEM. Finally, we realize such a scheme in AxiSEM3D and formulate and measure its computational cost.

2.1 AxiSEM3D

AxiSEM3D is a numerical method to solve wave propagation in 3-D geometry and media, a hybrid of SEM and Fourier-spectral method. Consider an earth model with volume Ω , density ρ and elasticity tensor \mathbf{C} . Subject to a point moment tensor \mathbf{M} at location \mathbf{r}_s and a stress-free boundary condition on the surface ($\partial\Omega$), the equations of motion can be written as the following weak form:

$$\int_{\Omega} (\rho \partial_t^2 \mathbf{u} \cdot \mathbf{w} + \nabla \mathbf{u} : \mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3 = \mathbf{M} : \nabla \mathbf{w}(\mathbf{r}_s), \quad (1)$$

where $\mathbf{u}(\mathbf{r}; t)$ and $\mathbf{w}(\mathbf{r})$ denote, respectively, the displacement field and an arbitrary test function.

As shown in Fig. 1, in a source-centred cylindrical coordinate system, that is, $\mathbf{u} = \mathbf{u}(s, \phi, z; t)$, Leng *et al.* (2016) parametrize the azimuthal dimension (ϕ -dimension) by Fourier series, that is,

$$\mathbf{u}(s, \phi, z; t) = \sum_{|\alpha| \leq n_u} \mathbf{u}^\alpha(s, z; t) \exp(i\alpha\phi), \quad (2)$$

where $\mathbf{u}^\alpha(s, z; t)$ are the Fourier coefficients of $\mathbf{u}(s, \phi, z; t)$, defined in the 2-D meridian domain D . Applying the same parametrization to the material properties and the moment tensor, we can reduce the 3-D weak form, eq. (1), to an algebraic system of coupled 2-D weak forms in D :

$$\begin{aligned} & \sum_{|\alpha| \leq n_u} \int_D \rho^{-(\alpha+\beta)} \partial_t^2 \mathbf{u}^\alpha \cdot \mathbf{w}^\beta s \, ds dz \\ & + \sum_{|\alpha| \leq n_u} \int_D \nabla_D \mathbf{u}^\alpha : \mathbf{C}^{-(\alpha+\beta)} : \nabla_D \mathbf{w}^\beta s \, ds dz \\ & = \mathbf{M}^{-\beta} : \nabla_D \mathbf{w}^\beta(\mathbf{r}_s), \quad \forall \beta \in [-n_u, n_u] \cap \mathbb{Z}, \end{aligned} \quad (3)$$

where ∇_D denotes the gradient operator in D (see eq. (26) in Leng *et al.* 2016). The right-hand side term, $\mathbf{M}^{-\beta} : \nabla_D \mathbf{w}^\beta(\mathbf{r}_s)$, vanishes unless $|\beta| = 0, 1$ or 2 , respectively, corresponding to the monopole, dipole and quadrupole radiation patterns emanating from a point moment tensor [see eq. (49)–(51) in Leng *et al.* 2016].

Eq. (3) is equivalent to eq. (1) but with collapsed ϕ -dimension, which enables us to solve $\mathbf{u}^\alpha(s, z; t)$ in 2-D as an alternative to solving $\mathbf{u}(s, \phi, z; t)$ in 3-D. It contains $(2n_u + 1)$ weak forms coupled by the Fourier coefficients of the material properties, ρ^γ and \mathbf{C}^γ , as formally expressed by the two summations over α . Leng *et al.* (2016) developed a 2-D SEM to solve $\mathbf{u}^\alpha(s, z; t)$ from eq. (3), named as AxiSEM3D. Leng *et al.* (2019) extended the above theory to aspherical geometry to accommodate undulating vertical discontinuities (such as the Earth's surface and Moho) based on the particle relabelling transformation (Al-Attar & Crawford 2016; Al-Attar *et al.* 2018).

The number of terms in the Fourier series parametrization, n_u in eq. (2), plays a central role in AxiSEM3D. In theory, it can be a function of both time and position in D , that is, $n_u = n_u(s, z; t)$; in practice, however, we only use a time-independent one, $n_u = n_u(s, z)$, because varying it with time demands dynamic domain repartitioning in the context of high-performance computing (HPC). At a given location, n_u can be bounded from above such that the wavelength resolution in the ϕ -direction matches that in D , namely,

$$n_u^{\text{sup}}(s, z) = \frac{2\pi s}{\lambda(s, z)} \times N_D \times \frac{1}{2} = \frac{\pi N_D s}{\lambda(s, z)}, \quad (4)$$

where $\lambda(s, z)$ denotes the spatial resolution taken as the local wavelength (S -wavelength in solid or P -wavelength in fluid), and thus $2\pi s/\lambda$ the number of wavelengths along the circle generated by revolving the point (s, z) , and N_D the number of gridpoints in D to resolve one wavelength in the s - or z -direction, a constant normally ranging from 6 to 8 in SEM, and the rearward 1/2 stems from the equivalence of one complex term in the Fourier series and two grid points along the circle. Given the Earth model or $\lambda(s, z)$, n_u^{sup} scales with s , the distance to the axis. Owing to the coupling of the Fourier modes, solving an eq. (3) with $n_u(s, z) = n_u^{\text{sup}}(s, z)$ will be more expensive than solving eq. (1) in 3-D. To make the Fourier parametrization or eq. (2) meaningful, we require

$$\int_D n_u(s, z) \, ds dz \ll \int_D n_u^{\text{sup}}(s, z) \, ds dz. \quad (5)$$

In other words, the 3-D wavefield must have low azimuthal complexity by nature or, as we describe in this paper, must be *quasi-axisymmetric* in the chosen coordinate system.

As a matter of fact, such quasi-axisymmetry does exist commonly for global Earth structures. In the simplest case, if the Earth model is axisymmetric, namely, $\rho^\gamma = 0$ and $\mathbf{C}^\gamma = 0$ for $|\gamma| > 0$, eq. (3) will become decoupled:

$$\begin{aligned} & \int_D \left(\rho^0 \partial_t^2 \mathbf{u}^{-\beta} \cdot \mathbf{w}^\beta + \nabla_D \mathbf{u}^{-\beta} : \mathbf{C}^0 : \nabla_D \mathbf{w}^\beta \right) s \, ds dz \\ & = \mathbf{M}^{-\beta} : \nabla_D \mathbf{w}^\beta(\mathbf{r}_s), \quad \forall \beta \in \{0, 1, 2\}. \end{aligned} \quad (6)$$

It contains only three independent weak forms, respectively, with unknowns \mathbf{u}^0 , \mathbf{u}^1 and \mathbf{u}^2 , or $n_u(s, z) = 2$. It is equivalent to the non-tensorial weak forms by Nissen-Meyer *et al.* (2007) and strong forms by Li *et al.* (2014). Next, as closely examined by Leng *et al.* (2016, 2019), $n_u \sim 10^2$ for the state-of-the-art tomographic mantle models, leading to a speedup from 10^2 to 10^3 in reference to a conventional 3-D SEM below a 5 s period. The wave physics behind this is that the weak, long-wavelength inhomogeneity cannot substantially alter the radiation patterns dominated by

the point moment tensor and in-plane scattering. Such a speedup may drop by one order of magnitude with a 3-D crust (Crust 1.0, Laske *et al.* 2013) because n_u will increase to $\sim 10^3$ above the penetration depth of the surface waves and the time step will be diminished by the thin oceanic plates (Leng *et al.* 2019). In short, by varying n_u in D , the computational cost of AxiSEM3D can be adapted to the azimuthal complexity of the 3-D wavefield, which is dominated by the complexity of the 3-D model. Such adaptivity distinguishes AxiSEM3D from a conventional 3-D method.

A strong, small-scale heterogeneity in the mantle poses a more outstanding challenge. An incident wave, upon encountering the heterogeneity, initiates a scattered wave that resembles a delta function in the ϕ -direction, leading to a large n_u in the near field of the heterogeneity. Such a scattered wave further propagates in all directions from the heterogeneity, forming a complicated $n_u(s, z)$ in the far field. Such scattering and propagation effects are well illustrated by the $n_u(s, z)$ shown in Fig. 1, obtained with a PREM earth model (Dziewonski & Anderson 1981) including a spherical heterogeneity in the mid-lower mantle. Being small in the far field (mostly smaller than 150), this $n_u(s, z)$ demands a computational cost similar to that for a global tomographic mantle model. However, its acquisition is not straightforward. Leng *et al.* (2019) developed a technique called ‘wavefield scanning’ to obtain such an $n_u(s, z)$, accomplished by the following three steps: (i) start a simulation with a sufficiently large Fourier expansion order, $n_u^{\text{start}}(s, z)$; (ii) during the time loop, monitor the required order, $n_u^{\text{scan}}(s, z)$, by checking the convergence of the spectrum $|\mathbf{u}^q(s, z)|$; (iii) at the end of the simulation, store the maximum $n_u^{\text{scan}}(s, z)$ over time. This stored $n_u^{\text{scan}}(s, z)$ can be re-applied to a similar earth model, for example, the field in Fig. 1 to a smaller or weaker heterogeneity at the same location. Nevertheless, global wavefield scanning becomes unavailable at a high frequency near 1 Hz because, as the frequency increases, n_u^{start} approaches to n_u^{sup} in the near field.

In brief, a strong, localized heterogeneity behaves more like a secondary source, emitting scattered waves in all directions. Such localization of energy, either at an earthquake source or near a heterogeneity, can be efficiently modelled in AxiSEM3D if it is located on the axis. The reason is that, no matter how complex the wavefield is, n_u is always bounded from above by n_u^{sup} , which scales with s in eq. (4). Having this in mind, one may naturally come up with a scheme of ‘decompose-and-rotate’, as illustrated in Fig. 2. Let \mathbf{u}_0 denote the incident wavefield that emanates from the earthquake source but propagates in the background model (without the heterogeneity) and $\delta\mathbf{u}$ the scattered wavefield defined by the difference $(\mathbf{u} - \mathbf{u}_0)$. Evidently, \mathbf{u}_0 and $\delta\mathbf{u}$ are both quasi-axisymmetric but under two different coordinate systems: \mathbf{u}_0 is quasi-axisymmetric about the earthquake and $\delta\mathbf{u}$ about the heterogeneity, that is, their own sources of energy. Therefore, both of them can be efficiently solved with AxiSEM3D under the two coordinate systems as long as the decomposition, $\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}$, can be exactly achieved—this is where the technique of wavefield injection comes into play.

2.2 Wavefield injection

The long-established theory of wavefield injection has been formulated from different perspectives, for example, Opršal *et al.* (2009) in a loose and understandable manner using the wave operator and Masson & Romanowicz (2016) and Lin *et al.* (2019) based on the representation theorems. Masson & Romanowicz (2016) have

shown that the representation theorems can naturally emerge from the domain decomposition and have established an equivalence between the different approaches for implementing wavefield injection. Here we present the theory in the weak forms. Our formulation is rigorous and concise, with emphasis on a natural construction of the hybrid wavefield and its quasi-axisymmetry under the heterogeneity-centred coordinate system.

Let us consider a general 3-D elastic medium with volume Ω , which encompasses a small-scale heterogeneity with volume H , as shown in Fig. 3. Note that H can be located right below the surface ($\partial H \cap \partial \Omega \neq \emptyset$). Given the material properties ρ and \mathbf{C} , a point moment tensor \mathbf{M} at \mathbf{r}_s and a stress-free boundary condition on $\partial \Omega$, the weak form for this problem is given by eq. (1). Now we choose a closed surface, denoted Γ_{inj} , to divide Ω into two subdomains, the far-field subdomain Ω_F and the near-field one Ω_N , with $H \in \Omega_N$, as shown in Fig. 3. We refer to Γ_{inj} as the *injection boundary*. The smallest Γ_{inj} we may choose is ∂H , but it is more convenient to use a regularly shaped one slightly larger than ∂H for an irregular heterogeneity. By introducing the traction \mathbf{t} exposed on Γ_{inj} , we can rebuild eq. (1) upon the two subdomains as

$$\begin{cases} \int_{\Omega_F} (\rho \partial_t^2 \mathbf{u} \cdot \mathbf{w} + \nabla \mathbf{u} : \mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ = \int_{\Gamma_{\text{inj}}} \mathbf{t} \cdot \mathbf{w} \, d\mathbf{r}^2 + \mathbf{M} : \nabla \mathbf{w}(\mathbf{r}_s), \end{cases} \quad (7a)$$

$$\begin{aligned} & \int_{\Omega_N} (\rho \partial_t^2 \mathbf{u} \cdot \mathbf{w} + \nabla \mathbf{u} : \mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ &= - \int_{\Gamma_{\text{inj}}} \mathbf{t} \cdot \mathbf{w} \, d\mathbf{r}^2. \end{aligned} \quad (7b)$$

Here we assume $\mathbf{r}_s \in \Omega_F$, but having \mathbf{r}_s in Ω_N makes no difference other than moving the source term $\mathbf{M} : \nabla \mathbf{w}(\mathbf{r}_s)$ from eqs (7a) to (7b). Note that neither eq. (7a) nor (7b) can be solved independently because \mathbf{t} on Γ_{inj} is introduced as an extra unknown. The summation of the two goes back to eq. (1).

Next, we consider the background medium, whose density and elasticity tensor are denoted, respectively, by ρ_0 and \mathbf{C}_0 , with $\rho_0 = \rho$ and $\mathbf{C}_0 = \mathbf{C}$ outside H . Respectively in line with eqs (1) and (7), the weak forms for the background medium subject to the same source and boundary conditions can be written as

$$\int_{\Omega} (\rho_0 \partial_t^2 \mathbf{u}_0 \cdot \mathbf{w} + \nabla \mathbf{u}_0 : \mathbf{C}_0 : \nabla \mathbf{w}) \, d\mathbf{r}^3 = \mathbf{M} : \nabla \mathbf{w}(\mathbf{r}_s), \quad (8)$$

and

$$\begin{cases} \int_{\Omega_F} (\rho_0 \partial_t^2 \mathbf{u}_0 \cdot \mathbf{w} + \nabla \mathbf{u}_0 : \mathbf{C}_0 : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ = \int_{\Gamma_{\text{inj}}} \mathbf{t}_0 \cdot \mathbf{w} \, d\mathbf{r}^2 + \mathbf{M} : \nabla \mathbf{w}(\mathbf{r}_s), \end{cases} \quad (9a)$$

$$\begin{aligned} & \int_{\Omega_N} (\rho_0 \partial_t^2 \mathbf{u}_0 \cdot \mathbf{w} + \nabla \mathbf{u}_0 : \mathbf{C}_0 : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ &= - \int_{\Gamma_{\text{inj}}} \mathbf{t}_0 \cdot \mathbf{w} \, d\mathbf{r}^2, \end{aligned} \quad (9b)$$

where the solution \mathbf{u}_0 has been referred to as the *incident wavefield*. Note that we have inserted $\rho_0 = \rho$ and $\mathbf{C}_0 = \mathbf{C}$ outside H into eq. (9a).

Subtracting eq. (9a) from eq. (7a)¹, we obtain

$$\begin{aligned} & \int_{\Omega_F} (\rho \partial_t^2 \delta \mathbf{u} \cdot \mathbf{w} + \nabla \delta \mathbf{u} : \mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ &= \int_{\Gamma_{\text{inj}}} \mathbf{t} \cdot \mathbf{w} \, d\mathbf{r}^2 - \int_{\Gamma_{\text{inj}}} \mathbf{t}_0 \cdot \mathbf{w} \, d\mathbf{r}^2, \end{aligned} \quad (10)$$

¹In general, addition or subtraction cannot be applied to two weak forms even they are formally additive. Such addition further requires that the two problems have the same Dirichlet (or displacement) boundary condition such that the arbitrary test function \mathbf{w} can always be chosen to be the same in the two weak forms. This requirement is satisfied by eqs (9a) and (7a), between which the only difference comes from the model parameters.

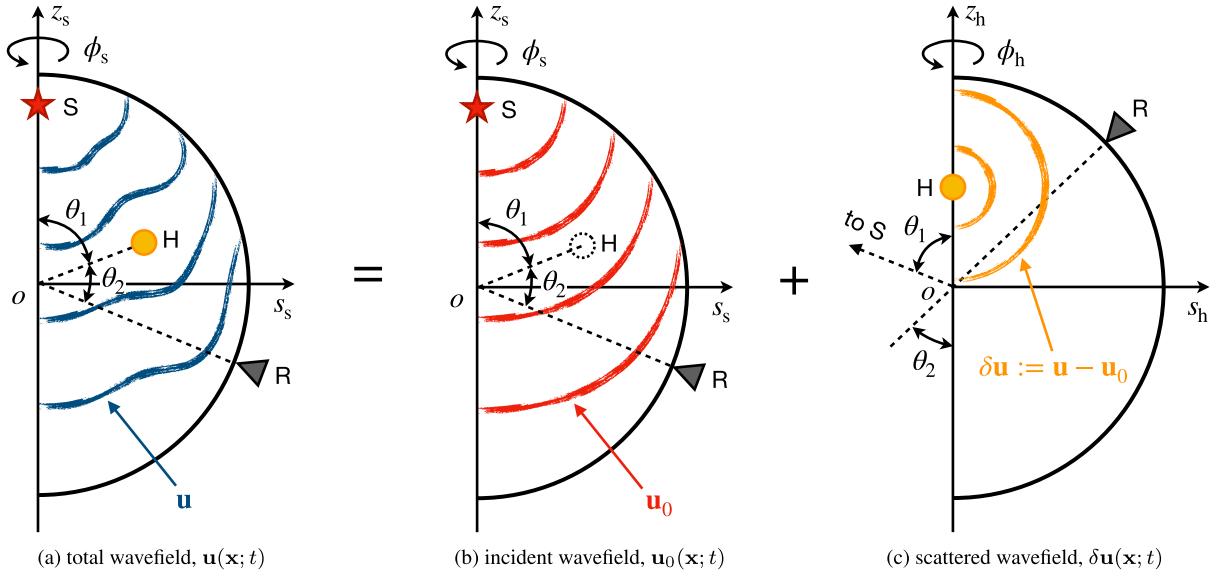


Figure 2. Decomposition of a total wavefield into an incident and a scattered wavefield, $\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}$. Excited by the earthquake source on the background model, the incident wavefield \mathbf{u}_0 is quasi-asymmetric in the source-centred coordinate system (s_s, ϕ_s, z_s). The scattered wavefield $\delta\mathbf{u}$, as defined by the difference $\mathbf{u} - \mathbf{u}_0$, is quasi-axisymmetric in the heterogeneity-centred coordinate system (s_h, ϕ_h, z_h). Once such a decomposition can be realized, AxiSEM3D can efficiently solve \mathbf{u}_0 and $\delta\mathbf{u}$, respectively, under the two different coordinate systems.

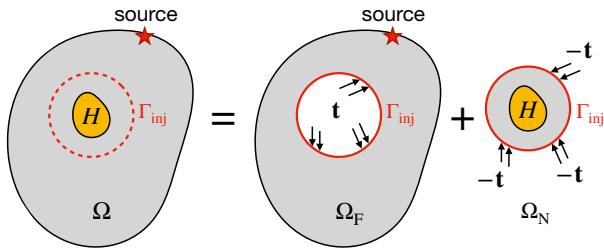


Figure 3. Sketch of wavefield injection. A heterogeneity H is included in Ω , a bounded 3-D domain of elastic medium. An injection boundary Γ_{inj} enclosing H can be chosen to divide Ω into two subdomains, exposing the traction \mathbf{t} on Γ_{inj} . Note that H can be located adjacent to the domain boundary $\partial\Omega$ and the source can be placed inside Γ_{inj} .

where

$$\delta\mathbf{u} \stackrel{\text{def}}{=} \mathbf{u} - \mathbf{u}_0, \quad (11)$$

which has been referred to as the *scattered wavefield*. In view of the difference between eqs (7a) and (10), it is natural to construct a *hybrid wavefield*, denoted \mathbf{u}_{hy} , which represents two different physical wavefields in Ω_F and Ω_N :

$$\mathbf{u}_{\text{hy}} \stackrel{\text{def}}{=} \begin{cases} \delta\mathbf{u}, & \mathbf{r} \in \Omega_F; \\ \mathbf{u}, & \mathbf{r} \in \Omega_N. \end{cases} \quad (12)$$

Replacing eq. (7a) with (10), we obtain the weak forms for \mathbf{u}_{hy} :

$$\int_{\Omega_F} \left(\rho \partial_t^2 \mathbf{u}_{\text{hy}} \cdot \mathbf{w} + \nabla \mathbf{u}_{\text{hy}} : \mathbf{C} : \nabla \mathbf{w} \right) d\mathbf{r}^3 = \int_{\Gamma_{\text{inj}}} \mathbf{t} \cdot \mathbf{w} d\mathbf{r}^2 - \int_{\Gamma_{\text{inj}}} \mathbf{t}_0 \cdot \mathbf{w} d\mathbf{r}^2, \quad (13a)$$

$$\int_{\Omega_N} \left(\rho \partial_t^2 \mathbf{u}_{\text{hy}} \cdot \mathbf{w} + \nabla \mathbf{u}_{\text{hy}} : \mathbf{C} : \nabla \mathbf{w} \right) d\mathbf{r}^3 = - \int_{\Gamma_{\text{inj}}} \mathbf{t} \cdot \mathbf{w} d\mathbf{r}^2, \quad (13b)$$

or, cancelling \mathbf{t} on Γ_{inj} by adding the above two,

$$\int_{\Omega} \left(\rho \partial_t^2 \mathbf{u}_{\text{hy}} \cdot \mathbf{w} + \nabla \mathbf{u}_{\text{hy}} : \mathbf{C} : \nabla \mathbf{w} \right) d\mathbf{r}^3 = - \int_{\Gamma_{\text{inj}}} \mathbf{t}_0 \cdot \mathbf{w} d\mathbf{r}^2. \quad (14)$$

Different from eqs (1) and (8), both complete, respectively, for the solutions of \mathbf{u} and \mathbf{u}_0 , eq. (14) is incomplete for that of \mathbf{u}_{hy} . This is because, in general, Dirichlet boundary conditions cannot be automatically satisfied by a displacement-based weak form; instead, they must be imposed upon the weak form as an additional constraint. Eqs (1) and (8) are complete because \mathbf{u} and \mathbf{u}_0 involve no Dirichlet boundary condition (the stress-free surface is a Neumann boundary condition honoured by the weak forms). In addition to eq. (14), however, \mathbf{u}_{hy} must satisfy the following Dirichlet boundary condition on Γ_{inj} :

$$\mathbf{u}_{\text{hy}}|_{\mathbf{r} \in \Omega_N \cap \Gamma_{\text{inj}}} - \mathbf{u}_{\text{hy}}|_{\mathbf{r} \in \Omega_F \cap \Gamma_{\text{inj}}} = \mathbf{u}_0|_{\mathbf{r} \in \Gamma_{\text{inj}}}. \quad (15)$$

Such a condition enforces \mathbf{u}_{hy} to be discontinuous on Γ_{inj} by a prescribed gap of \mathbf{u}_0 , as a consequence of its definition by eq. (12). Combing eqs (14) and (15), we conclude that \mathbf{u}_{hy} is the wavefield excited by two surface sources, that is, \mathbf{u}_0 and \mathbf{t}_0 on Γ_{inj} , exerted on the 3-D perturbed medium.

Thus far, we have established a complete theory for wavefield injection, which suggests a two-stage approach as an alternative to solving the all-in-one problem given by eq. (1). In the first stage, one solves eq. (8) for the incident wavefield \mathbf{u}_0 , recording \mathbf{u}_0 and \mathbf{t}_0 on Γ_{inj} and \mathbf{u}_0 at the receiver locations \mathbf{r}_r . This stage only involves wave propagation in the background medium. In the second stage, one solves eqs (14) and (15) for the hybrid wavefield \mathbf{u}_{hy} , injecting the recorded \mathbf{u}_0 and \mathbf{t}_0 on Γ_{inj} into the perturbed medium and recording \mathbf{u}_{hy} or $\delta\mathbf{u}$ at \mathbf{r}_r . In the end, the total wavefield \mathbf{u} can be reconstructed at \mathbf{r}_r by $\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}$.

Now we come back to deep Earth scattering and investigate the azimuthal complexity of \mathbf{u}_0 and $\delta\mathbf{u}$. Here we assume that the background mantle model be characterized by weak, long-wavelength lateral heterogeneity. This assumption is consistent with our current knowledge about the Earth's mantle from seismic tomography. In the case that the background model has a scattering strength comparable to that of the small-scale heterogeneity, the problem should be solved in an all-in-one fashion rather than by wavefield

injection. Under this assumption, the incident wavefield \mathbf{u}_0 is quasi-axisymmetric in the source-centred coordinate system, as illustrated in Fig. 2(b). This can be explained from an energy perspective: localization of energy occurs only at the earthquake source, that is, the right-hand side of eq. (8), which is placed on the axis in the source-centred coordinate system. The quasi-axisymmetry of \mathbf{u}_0 has also been verified by Leng *et al.* (2016, 2019). The hybrid wavefield \mathbf{u}_{hy} , on the other hand, is quasi-axisymmetric in the heterogeneity-centred coordinate system, as illustrated in Fig. 2c. The reasons are similar but slightly more sophisticated: first, the two surface sources [the right-hand side of eqs (14) and (15)] and the strong 3-D structures [the left-hand side of eq. (13b)] are both confined to the near-field subdomain Ω_N ; secondly, once placed on the axis, Ω_N can be described by a small n_u^{sup} owing to its small lateral extent, that is, s in eq. (4). Therefore, as proposed in Section 2.1 and illustrated in Fig. 2, wavefield injection realizes the decomposition of \mathbf{u} into \mathbf{u}_0 and $\delta\mathbf{u}$ in the far field, both of which can be efficiently solved with AxiSEM3D, respectively, under the source- and heterogeneity-centred coordinate systems.

It must be emphasized again that the above two-stage approach, that is, solving first eq. (8) and then eqs (14) and (15), yields the exact solution identical to the all-in-one solution by eq. (1). Also, this approach does not care about the location of the heterogeneity, that is, whether it is near the source, in the deep interior or beneath the receivers. However, because of the 3-D structures in H , the solution of \mathbf{u}_{hy} or the second stage necessitates a 3-D global solver (even the large background model is spherically symmetric or 1-D). If the global solver can only handle a 1-D background model, as those used by most existing injection approaches, the second stage has to be conducted upon a truncated domain slightly larger than Ω_N using a local 3-D solver, and a third stage, known as ‘wavefield extrapolation’ (Masson *et al.* 2013; Masson & Romanowicz 2016), must follow to obtain $\delta\mathbf{u}$ at the receivers. Such a three-stage approach becomes inexact due to the introduction of an absorbing boundary for domain truncation. It is noted that this absorbing boundary can be replaced by an ‘exact’ boundary condition to preserve the higher-order interactions between the heterogeneity and the background model (van Manen *et al.* 2007; Masson & Romanowicz 2017). This exact boundary condition requires a precomputed database of the Green’s functions based on the background model, with the point source placed throughout a vicinity of the local domain boundary. Assume that there are N gridpoints on the boundary; for a general 3-D background model, the database must be computed with a cost of $O(N)$ and demands a storage of $O(N^2)$, both difficult to realize in 3-D; however, it becomes plausible for a 1-D background model, with the computation and storage reduced to $O(M)$ and $O(M \times N)$, respectively, where M is the number of different depths of the N gridpoints. In the scenario of wavefield extrapolation, AxiSEM3D can be adopted as either the global or the local solver. Related topics will be further discussed in Sections 6.1 and 6.2.

2.3 Implementation in SEM

This subsection aims for implementing wavefield injection in any SEM in 2-D or 3-D. Based on the general scheme developed here, we shall elaborate on our implementation in AxiSEM3D in the following subsection.

Regardless of the used solvers, the only difficulty in implementing wavefield injection is to record and inject the two surface sources on Γ_{inj} , that is, the right-hand side of eqs (14) and (15). In a straight course, the traction source can be implemented as is based on the

following stress formula,

$$-\int_{\Gamma_{\text{inj}}} \mathbf{t}_0 \cdot \mathbf{w} d\mathbf{r}^2 = -\int_{\Gamma_{\text{inj}}} \mathbf{n} \cdot \boldsymbol{\sigma}_0 \cdot \mathbf{w} d\mathbf{r}^2, \quad (16)$$

where $\boldsymbol{\sigma}_0$ and \mathbf{n} denote, respectively, the incident stress tensor and the unit norm on Γ_{inj} . To follow this path, the global solver has to compute and record $\boldsymbol{\sigma}_0$ on Γ_{inj} (e.g. Monteiller *et al.* 2012, using the direct solution method). Masson *et al.* (2013) have shown the option of directly recording \mathbf{t}_0 based on the discretization of the surface integral in the representation theorem, which is essentially the same as recording $\boldsymbol{\sigma}_0$. If the local solver is an FDM (e.g. Masson *et al.* 2013; Opršal *et al.* 2009), the complete description of a wavefield requires both displacement and stress on a staggered grid; in this case, using eq. (16) becomes mandatory. Here we only consider SEM, which is a pure displacement method.

Directly recording and injecting $\boldsymbol{\sigma}_0$ or \mathbf{t}_0 on Γ_{inj} may lead to three complexities. First, many seismogram-driven global solvers are incapable of computing stress by its nature; additional development work is required and the accuracy for stress may not match that for displacement. Secondly, the surface integral on the right-hand side of eq. (16) can be complicated by an irregular shape of Γ_{inj} . Finally, as a continuous Galerkin technique (Peter *et al.* 2011), SEM cannot naturally handle the displacement gap on Γ_{inj} ; honouring such discontinuity by increasing the number of degrees of freedom at those Gauss–Lobatto–Legendre (GLL) points on Γ_{inj} will necessitate low-level changes to the SEM architecture, especially in an HPC environment.

These difficulties can be solved by our traction-free scheme described in Appendix A. This scheme is completely displacement-based, which allows us to record and inject only \mathbf{u}_0 in the vicinity of Γ_{inj} instead of both \mathbf{u}_0 and \mathbf{t}_0 (or $\boldsymbol{\sigma}_0$) on Γ_{inj} . Secondly, it avoids computing any surface terms, either Neumann or Dirichlet; therefore, the geometry of Γ_{inj} can be arbitrary. Finally, it features an extremely simple implementation in SEM.

2.4 Implementation in AxiSEM3D

Following the traction-free scheme established in Fig. A2, we have implemented wavefield injection in AxiSEM3D. Nothing is special except that the recording and the injection of the incident wavefield must be conducted in terms of Fourier series, respectively, in the source-centred and the heterogeneity-centred coordinate system. The technical details will be elaborated in Appendix B. The complete process from mesh generation to post-processing will be demonstrated in the following section. In this subsection, we focus on determining the Fourier expansion order $n_u(s, z)$ for the hybrid wavefield, a unique and critical step for using AxiSEM3D as the underlying solver. Based on this $n_u(s, z)$, we provide an equation to estimate the computational cost of an injection simulation.

2.4.1 Fourier expansion order of \mathbf{u}_{hy}

As explained in Section 2.1, the hybrid wavefield \mathbf{u}_{hy} is naturally quasi-axisymmetric in the heterogeneity-centred coordinate system (s_h, ϕ_h, z_h). Here we show how AxiSEM3D exploits such quasi-axisymmetry via a small $n_u(s_h, z_h)$ for \mathbf{u}_{hy} based on the wave scattering physics.

In an injection simulation with AxiSEM3D, the heterogeneity H is placed around the axis and the injection boundary Γ_{inj} becomes an open curve with both its endpoints located on the axis, as shown

in Fig. A1. This Γ_{inj} divides the 2-D meridian domain D into a far-field subdomain D_F and a near-field subdomain D_N , which form, respectively, the meridians of Ω_F and Ω_N . To determine a $n_u(s_h, z_h)$ sufficiently large for an exact description of \mathbf{u}_{hy} , we must take into account three factors: (i) the azimuthal complexity of the injection source, that is, \mathbf{u}_0 in the vicinity of Γ_{inj} , (ii) the near-field scattering of \mathbf{u}_0 by the heterogeneity and (iii) the far-field scattering of $\delta\mathbf{u}$ by the background model.

Let us start from a 1-D background model. In this case, the following $n_u(s_h, z_h)$ can be proved sufficient:

$$n_u^{\text{het}}(s_h, z_h) = \begin{cases} n_u^{\text{sup}}(s_h, z_h), & (s_h, z_h) \in D_N; \\ \max_{(s'_h, z'_h) \in \Gamma_{\text{inj}}} n_u^{\text{sup}}(s'_h, z'_h), & (s_h, z_h) \in D_F. \end{cases} \quad (17)$$

In the near field, we adopt a *full* azimuthal resolution by using the theoretical upper bound $n_u^{\text{sup}}(s_h, z_h)$ in eq. (4); this is because, first, \mathbf{u}_0 loses its quasi-axisymmetry when expanded in (s_h, ϕ_h, z_h) and, second, the 3-D structure inside H can be arbitrary. In the far field, however, because the 1-D background model cannot generate any off-plane scattering effects [or no mode coupling as in eq. (6)], n_u will never increase outside H ; therefore, throughout D_F , we can constantly use the maximum n_u^{sup} over Γ_{inj} . In brief, $n_u^{\text{het}}(s_h, z_h)$ is completely determined by the heterogeneity in both D_N and D_F (so we label it by ‘het’ in the superscript). Thanks to the small lateral size of H , n_u^{het} is remarkably smaller than n_u^{sup} in the vast far field, which reflects the quasi-axisymmetry of \mathbf{u}_{hy} [recall eq. (5)] and thus the efficiency of AxiSEM3D.

In the presence of a 3-D background model, a $n_u^{\text{bm}}(s_h, z_h)$ will be added to $n_u^{\text{het}}(s_h, z_h)$ in D_F to account for the scattering of $\delta\mathbf{u}$ by the background model. This $n_u^{\text{bm}}(s_h, z_h)$ can be determined following the same way as determining the $n_u(s_s, z_s)$ for \mathbf{u}_0 in the first stage, for example, by wavefield scanning with a fictitious earthquake source placed at the heterogeneity, as elaborated in Leng *et al.* (2019) and briefed in Section 2.1. Different from n_u^{het} that is constant in D_F , n_u^{bm} must be (sometimes highly) heterogeneous to unleash the performance of AxiSEM3D (see e.g. Leng *et al.* 2019). Here we emphasize the frequency dependency of these two n_u fields. As verified by Leng *et al.* (2016, 2019), n_u^{bm} barely increases with frequency for the state-of-the-art tomographic mantle models, whereas n_u^{het} scales with frequency because λ appears on the denominator in eq. (4). At some high frequency, therefore, the sum of the two will still be dominated by n_u^{het} , which implies that the computation costs with and without a 3-D background model can be similar (or at least on the same order of magnitude).

2.4.2 Computational cost

Given an arbitrary 3-D heterogeneity in a 1-D background model, the computation cost for an injection simulation with a n_u^{het} given by eq. (17) can be explicitly written as

$$c_{\text{inj}} = C_{\text{bm}} \times C_{\text{het}} \times \omega^3 \times \left(\omega \max_{\Gamma_{\text{inj}}} \frac{s}{v_s} \right), \quad (18)$$

where C_{bm} and C_{het} are two model-dependent constants, ω the frequency, s the distance to axis and v_s the shear velocity. Note that v_s is determined by the background model because our traction-free scheme requires $\Gamma_{\text{inj}} \cap \partial H = \emptyset$. eq. (18) shows that the heterogeneity affects the computational cost only via its lateral size (or s), irrespective of its shape and perturbation strength.

We first explain the scaling characteristics of eq. (18). The factor ω^3 accounts for the two in-plane spatial dimensions and the time dimension, and $\omega \max_{\Gamma_{\text{inj}}} \frac{s}{v_s}$ for the azimuthal dimension (because

n_u^{het} in D_F is proportional to this term). Note that the computational cost is governed by the constant n_u^{het} in the vast far field in eq. (17), whereas the difference due to the small near field is negligible. Such a scaling behaviour is distinct from what has been observed for a weak, long-wavelength inhomogeneity. As verified by Leng *et al.* (2016, 2019), AxiSEM3D has a computational cost scaling with ω^3 for a tomographic mantle model, so it can be significantly more efficient than a 3-D SEM at a high frequency. For a small-scale heterogeneity, however, its computational cost scales with ω^4 , the same as that of a 3-D SEM. Instead of frequency, the speedup now stems from the small lateral size of the heterogeneity.

Now we explain the two constants and report their measurements on Archer, the UK National Supercomputing Service. The constant C_{bm} depends only on the background model and has a unit of CPU-hr³. Adopting a mesh resolution of 1.5 elements (or 6 GLL points) per wavelength and a 30-min record length (using the explicit second-order Newmark time scheme), our global-scale tests show that $C_{\text{bm}} \approx 14\,000$ CPU-hr³ for a visco-elastic, anisotropic PREM, including a 25 per cent parallelization penalty due to unbalanced I/O operations (reading in \mathbf{u}_0 on the injection elements and writing seismograms at the receivers). For surface wave modelling with the heterogeneity located in the crust or the uppermost mantle, the number of elements per wavelength should be increased to 2.0, whereby C_{bm} will be increased by 1.5–2 times. The constant C_{het} is dimensionless, which reflects any mesh refinements required by the velocity structure or geometry of the heterogeneity; $C_{\text{het}} = 1$ if no refinement is needed. There are two typical situations where mesh refinements become necessary: first, the shear velocity of the heterogeneity is smaller than that of the background medium (because the element size is determined by the local S -wavelength); secondly, the boundary of the heterogeneity, which is not an interface in the background model, is required to be exactly honoured by the mesh. For instance, to accommodate a ULVZ with a 30 km height and a 50 per cent v_s reduction, one should refine the mesh generated based on PREM by first adding a radial discontinuity at 30 km above the CMB and using a halved mesh size between CMB and the new discontinuity; such refinements lead to $C_{\text{het}} \approx 1.3$. Mesh refinements are automatically performed by our mesher.

To visualize the efficiency suggested by eq. (18), let us consider a spherical heterogeneity with a 50 km radius, located in the mantle where $v_s = 7$ km s⁻¹. As estimated by eq. (18), the computational cost at 1 Hz for a 30-min record length approaches 100 000 CPU-hr, expensive but readily achievable with modern HPC; such a cost rapidly decreases to ~ 6250 CPU-hr at 0.5 Hz, which may enable a parameter study with a good number of simulations; finally, it plunges to ~ 400 CPU-hr at 0.25 Hz, which could even empower a full-waveform inversion algorithm. It must be noted that the computational cost given by eq. (18) has been derived and measured using the constant n_u^{het} in D_F , as given by eq. (17). Such a constant n_u^{het} guarantees that the scattered wavefield is solved exactly throughout the Earth’s interior. This is the worst case in terms of computational cost. In practice, n_u^{het} can be further optimized, for example, using wavefield scanning (Leng *et al.* 2019), taking into account propagation effects such as structural complexity, record length and path geometry. In this paper, however, we ignore any case-dependent optimizations and always report the highest computational cost for an exact global solution.

Using a 3-D background model will increase the above computational cost to different extents. Such an increase cannot be

summarized by a simple equation like eq. (18) because of the heterogeneous distribution of $n_u^{\text{bm}}(s_h, z_h)$ (see Leng *et al.* (2019) for various examples obtained by wavefield scanning). As analysed in Section 2.4.1, the different frequency-scaling behaviours of n_u^{bm} and n_u^{het} suggest that, at the low-frequency end, the computational cost will be dominated by n_u^{bm} or the complexity of the background model while, at the high-frequency end, by n_u^{het} or the lateral size of the heterogeneity. Both these limit cases will appear in the numerical examples.

3 VERIFICATION

In this subsection, we verify our injection approach by a benchmark solution against a discretized 3-D SEM, SPECFEM3D_GLOBE (SPECFEM, Komatitsch & Tromp 2002). We also conduct a reproduction test to measure the numerical errors originating from wavefield injection.

3.1 Benchmark setup

We consider a spherical heterogeneity located in the mid-lower mantle on top of a global tomographic model S40RTS (Ritsema *et al.* 2011), as detailed in Fig. 4. The dominant period is 10 s (NEX = 432 in SPECFEM). In order to generate sufficiently strong scattering effects (so that they cannot be modelled by any approximate approaches), we use an extraordinarily strong ($\delta v_s = 50$ per cent) and large ($R = 250\text{ km} \approx 4\lambda_s$) heterogeneity. Neither the SPECFEM nor the AxiSEM3D mesh can exactly honour the spherical shape of the heterogeneity. A sharp spherical boundary will be smoothed by a spatial interpolation over the spectral elements it cuts through and thus interpreted differently in SPECFEM and AxiSEM3D. To completely avoid the error caused by model misinterpretation, we let the material perturbations gradually decrease to zero across an outer shell with a 50 km thickness, which is slightly smaller than the local S -wavelength but can still be well sampled by the 4–5 GLL points distributed within it. Such a thick, gradual boundary is adopted only for benchmark purpose and will be dropped in the other numerical examples.

For the earthquake source, we use a fully axisymmetric moment tensor ($M_{rr} = 2$, $M_{tt} = M_{pp} = -1$, $M_{rt} = M_{tp} = M_{pr} = 0$) so that any SH-energy observed in the transverse displacement originates from the 3-D structure. The source depth is 600 km. We adopt a 30 min seismogram length to cover most of the significant scattered phases. We compare the seismograms at more than 3000 receivers distributed uniformly on the surface, covering an epicentral distance from 10° to 150° and an azimuth from 0° to 90° . For each pair of seismograms obtained by SPECFEM and wavefield injection with AxiSEM3D, we quantify their goodness-of-fit by the time-frequency misfits (Kristeková *et al.* 2009). Similar to Chaljub *et al.* (2015), we calculate the scores of goodness-of-fit by $10^{\exp[-\max(|\text{EM}|, |\text{PM}|)]}$, where EM and PM denote, respectively, the globally normalized envelope and phase misfits; a score of 10 means a perfect match.

The reference solution can be obtained from two simulations with SPECFEM, one for \mathbf{u}_0 (S40RTS without the heterogeneity) and the other for \mathbf{u} (S40RTS with the heterogeneity), and $\delta\mathbf{u}$ is then computed by $(\mathbf{u} - \mathbf{u}_0)$. Wavefield injection with AxiSEM3D is accomplished by the following steps:

(i) *Mesh generation*: generate a mesh based on PREM, as shown in Fig. 4, which will be used for both the incident simulation and the injection simulation; no mesh refinement is required because the

shear velocity perturbation is positive (see Section 2.4.2 for details about mesh refinements).

(ii) *Injection boundary*: under (s_h, ϕ_h, z_h) , select an injection boundary Γ_{inj} to enclose the heterogeneity and then identify the injection elements and the injection GLL points as instructed by Fig. A1; the selected Γ_{inj} in Fig. 4 consists of two arcs (top and bottom) and one line (right), leading to 30 injection elements in total.

(iii) *n_u^{het} for \mathbf{u}_{hy} and the cardinal points*: determine n_u^{het} at all the GLL points on the injection elements using eq. (4); next, locate the cardinal points associated with these GLL points in (s_s, ϕ_s, z_s) ; the maximum n_u^{het} on Γ_{inj} occurs at its top-right-hand corner, as labelled A in the zoomed view; at this point, $s \approx 358\text{ km}$ and $v_s \approx 6.5\text{ km s}^{-1}$, so the maximum n_u^{het} is calculated as $\frac{\pi \times 6 \times 358\text{ km}}{6.5\text{ km/s} \times 10\text{ s}} \approx 104$.

(iv) *Incident simulation*: propagate the wave emanating from the earthquake source upon the background model under (s_s, ϕ_s, z_s) , recording \mathbf{u}_0 at both the cardinal points and the receivers; this is a normal AxiSEM3D simulation with its $n_u(s_s, z_s)$ determined by the background model; for this problem, we simply use a constant one, $n_u(s_s, z_s) = 300$, which is much larger than the one obtained by wavefield scanning on S40RTS (Leng *et al.* 2019).

(v) *Processing \mathbf{u}_0* : perform the coordinate transformation from $\mathbf{u}_0(s_s^J, \phi_s^J, z_s^J)$ to $\mathbf{u}_0(s_h, \phi_h^J, z_h)$ at each cardinal point and then compute the Fourier coefficients $\mathbf{u}_0^\alpha(s_h, z_h)$ from $\mathbf{u}_0(s_h, \phi_h^J, z_h)$ at each GLL point by a fast Fourier transform.

(vi) *Injection simulation*: back into (s_h, ϕ_h, z_h) and following the workflow in Fig. A2, inject $\mathbf{u}_0(s_h, z_h)$ into the perturbed 3-D model as a volume source over the injection elements, recording \mathbf{u}_{hy} or $\delta\mathbf{u}$ at the receivers; here we use $n_u(s_h, z_h) = 404$ in the far field (as $n_u^{\text{het}} = 104$ and $n_u^{\text{bm}} = 300$).

(vii) *Reconstructing \mathbf{u}* : transform $\delta\mathbf{u}(s_h, \phi_h, z_h)$ to $\delta\mathbf{u}(s_s, \phi_s, z_s)$ at each receiver and reconstruct the total wavefield by $\mathbf{u}_0 + \delta\mathbf{u}$.

3.2 Benchmark results

The agreement between SPECFEM and wavefield injection with AxiSEM3D turns out nearly exact. Among all the three components at all the receivers, the lowest score of goodness-of-fit is about 9.6 out of 10. Fig. 5 displays the seismograms at a receiver located at 60° in epicentral distance and 20° in azimuth. It is shown that, for both \mathbf{u} (left-hand column) and $\delta\mathbf{u}$ (middle column), the seismograms obtained by wavefield injection (dashed red) are basically identical to those by SPECFEM (solid black).

Let us look into the scattered phases. The selected time window in Fig. 5 reveals two significant scattered energy packets, sP2S+S2S² and ScS2sP. First, the arrival time of S2S is close to that of S because the heterogeneity is located approximately halfway between the earthquake and the receiver. Secondly, because the heterogeneity is strong and large, the amplitude of sP2S+S2S is only one order of magnitude lower than that of S in the radial and vertical components (comparing Figs 5a to b and g to h); in the transverse component (comparing Figs 5d to e), their amplitudes are mostly the same because there is no incident SH-energy. Finally, the amplitude of

²In this paper, we name a scattered phase as A2B, where A and B are, respectively, the incident and the outgoing phases. Such a ray-based nomenclature implies Mie scattering, a scattering regime where the structural scale is smaller than the wavelength (Wu & Aki 1985). Our approach is aimed at the general scattering regime where the heterogeneity can be much larger than the wavelength. In this general case, both converted phase names and traveltimes give inexact descriptions of a scattered wave; still, these concepts are helpful in understanding the scattered waveforms.

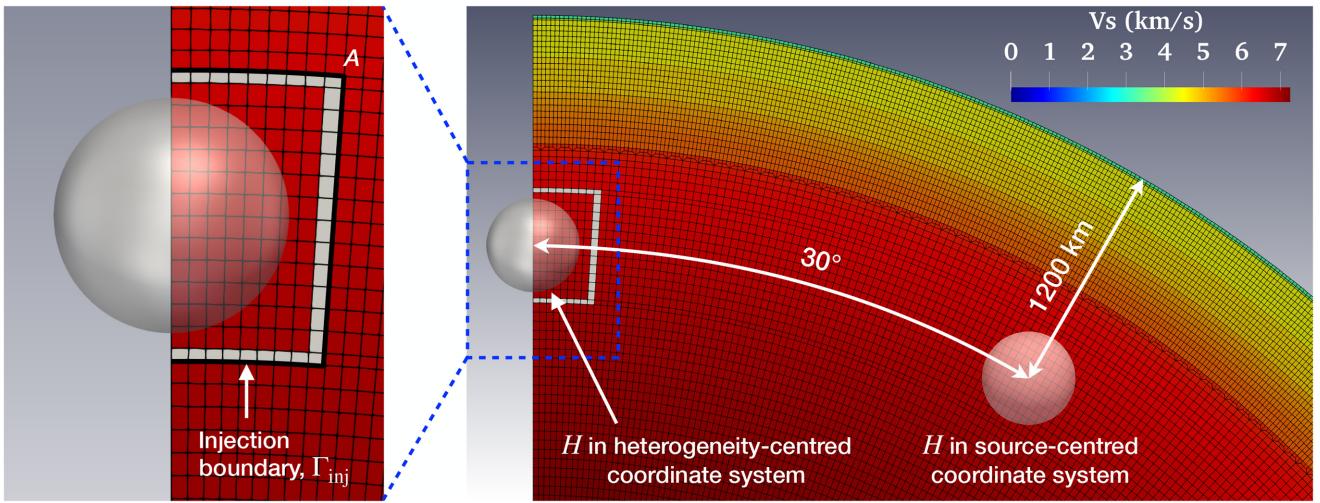


Figure 4. Spectral element mesh for simulating wave scattering by a spherical heterogeneity in the mid-lower mantle. We use PREM to generate the mesh at a 10 s period and use S40RTS as the background model for wave propagation. The heterogeneity is located 1200 km in depth and 30° in epicentral distance, with an outer radius of 250 km ($\approx 4\lambda_s$). Within an inner radius of 200 km, the perturbations to v_s , v_p and ρ are, respectively, 50, 20 and 10 per cent, which smoothly decrease to zero at 250 km. This mesh will be used for both the incident simulation under the source-centred coordinate system and the injection simulation under the heterogeneity-centred coordinate system, with the heterogeneity located, respectively, at 30° and on the axis. The left-hand panel zooms into the dashed box on the right-hand side, highlighting the injection boundary Γ_{inj} (in thick curves) and the injection elements (in white colour).

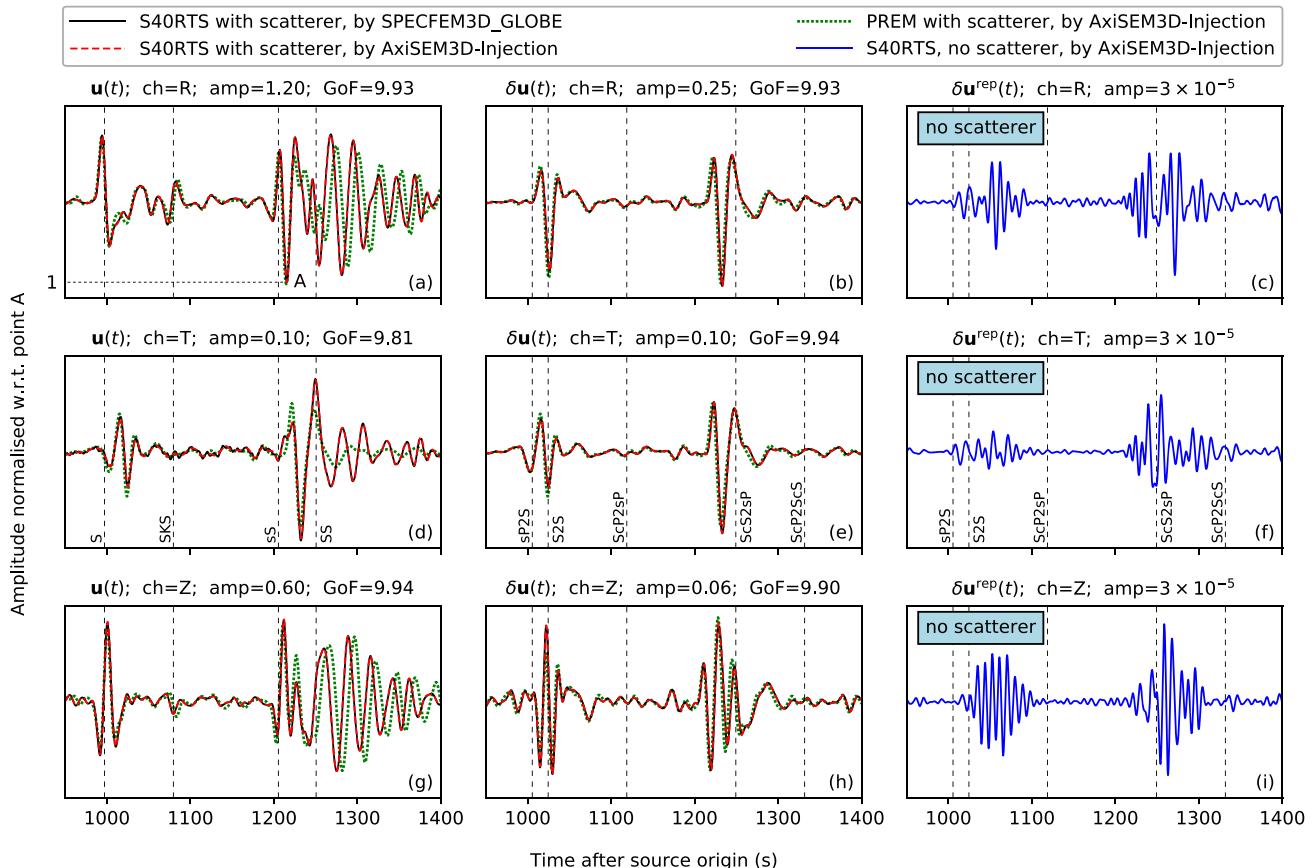


Figure 5. Verification by a comparison to a 3-D SEM and the reproduction test. The problem is described in Fig. 4, simulated at a 10 s period. We display the synthetic seismograms at a receiver located at 60° in epicentral distance and 20° in azimuth. The three columns from left-hand side to right-hand side show, respectively, the total wavefield $\mathbf{u}(t)$, the scattered wavefield $\delta\mathbf{u}(t)$ and the reproduction error $\delta\mathbf{u}^{rep}(t)$ (the computed $\delta\mathbf{u}(t)$ without the small-scale heterogeneity). The three rows from top to bottom show, respectively, the radial (R), transverse (T) and vertical (Z) components. All the seismograms are normalized by the amplitude at point A in (a); after such normalization, the amplitude range of each plot is given by 'amp' in the title. The scores of goodness-of-fit (GoF) are calculated for S40RTS, using the waveforms computed, respectively, by SPECFEM3D_GLOBE (solid black) and AxiSEM3D-Injection (dashed red). The traveltimes of the scattered phases are calculated by ray tracing, assuming point scattering at the centre of the heterogeneity.

ScS2sP is slightly larger than that of sP2S+S2S; this is because the scattering angle³ of ScS2sP (24.6°) turns out smaller than those of sP2S (35.2°) and S2S (38.8°), as the receiver is located 20° away from the source-heterogeneity plane. It is noted that three other scattered phases, S2ScS, ScS2S and sP2ScS, arrive at the receiver mostly at the same time as ScS2sP does, but we can ascertain that the energy packet is dominated by ScS2sP because the scattering angles of the other three are all close to 80° .

Further, to assess the influence of the 3-D background model, we replace S40RTS with PREM and repeat the simulations. As shown in Fig. 5, the waveform differences between using S40RTS and PREM are small for both S and S2S; this is because they travel mainly in the mid-lower mantle where S40RTS turns out less heterogeneous in the source-receiver plane. The differences become larger for ScS2sP, whose incident branch extends into the lowermost mantle and encounters some long-wavelength slow anomalies. The largest differences are found near SS, which travels primarily within the upper mantle and the transition zone and comes across a strong fast anomaly near its bounce point. In general, a tomographic mantle model alters a scattered wave in a smooth manner, the same as it alters an incident wave. For a global-scale application, whether to use a 3-D background model depends on various factors such as the source-heterogeneity-receiver geometry, the scattered phases of interest and the scattering strength of the small-scale heterogeneity relative to the background model.

Finally, we compare the computational cost. The all-in-one simulation with SPECFEM took 1069 CPU-hr (0.55 hr on 1944 cores) for the 30-min record length (the same for PREM and S40RTS). The injection simulations with AxiSEM3D took 490 CPU-hr (0.68 hr on 720 cores) and 86 CPU-hr (0.12 hr on 720 cores), respectively, using S40RTS and PREM as the background model. First, wavefield injection with AxiSEM3D has acquired a speedup around 12 for PREM. Such a speedup is fundamental because it is frequency-independent (recall that both the costs scale with ω^4). More significantly, this speedup will become increasingly attractive as the heterogeneity shrinks laterally [recall that c_{inj} scales with s in eq. (18)]; for instance, it may reach ~ 80 when the maximum s on Γ_{inj} decreases to 50 km. Besides, using S40RTS as the background model has multiplied the computational cost by a factor of 6, which can be remarkably decreased by using an optimized $n_u^{\text{bm}}(s_h, z_h)$ instead of the large constant one, $n_u^{\text{bm}} = 300$. However, even we keep using $n_u^{\text{bm}} = 300$, such a difference will decrease with frequency.

3.3 Reproduction test

A reproduction test is an injection simulation in the absence of the heterogeneity (Opršal *et al.* 2009), which is supposed to yield $\delta\mathbf{u} = 0$ everywhere. However, wavefield injection inevitably introduces some systematic numerical errors that are generally larger than the random floating-point errors. These systematic errors, as termed the reproduction errors, mainly come from the calculation and storage of \mathbf{u}_0 near Γ_{inj} : first, a spatial interpolation is required to bridge the two different representations (or discretizations) of \mathbf{u}_0 in the two stages (e.g. through the cardinal points in AxiSEM3D); secondly, to cut the storage for \mathbf{u}_0 , a temporal downsampling may become

³A scattering angle is the angle between the incident and the outgoing wavevectors. In Mie scattering, the amplitude of a scattered wave decreases with its scattering angle (Wu & Aki 1985). Similar to converted phase names and traveltimes, scattering angles also give an inexact description but useful in analysing the amplitude of a scattered phase.

necessary at a high frequency (e.g. the sampling rate determined by the simulation time step can reach about 100 for PREM at 1 Hz). A reproduction test is indispensable to guarantee that the level of these reproduction errors are safely below that of the scattered wavefield.

As shown in the right-hand column of Fig. 5, the reproduction errors for our benchmark problem are on the order of magnitude of 10^{-5} relative to the incident waves. Such an error level again verifies our implementation, as we have compiled the solver with single precision, which has a floating-point error of $\sim 10^{-6}$ (ISO C standard). Concerning a smaller and weaker heterogeneity, however, the scattered wave can drop below this floating-point error. To avoid that, the solver must be compiled with double precision, which has a floating-point error of $\sim 10^{-15}$ (ISO C standard); a scattered wave below this error level is totally undetectable.

4 COMPARISON WITH APPROXIMATE SOLUTIONS

This section delivers a high-frequency application of our approach. As reviewed in the introduction, two approximations have been playing a fundamental role in seismological investigations of mantle heterogeneities, the perturbation theory and 2-D in-plane modelling. In this section, we compare these two approximate solutions with the exact solution obtained by wavefield injection with AxiSEM3D.

4.1 Perturbation theory

Similar to wavefield injection, the perturbation theory starts from solving the incident wavefield \mathbf{u}_0 associated with a background model (ρ_0, \mathbf{C}_0) , as governed by eq. (8). It then aims at establishing an explicit relation between the wavefield perturbation $\delta\mathbf{u}$ and the model perturbations $(\delta\rho, \delta\mathbf{C})$ in the neighbourhood of \mathbf{u}_0 and (ρ_0, \mathbf{C}_0) .

Subtracting eq. (8) from (1) and applying eq. (11) to the result, one can obtain a weak form of the scattered wavefield $\delta\mathbf{u}$:

$$\begin{aligned} & \int_{\Omega} (\rho_0 \partial_t^2 \delta\mathbf{u} \cdot \mathbf{w} + \nabla \delta\mathbf{u} : \mathbf{C}_0 : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ &= - \int_H (\delta\rho \partial_t^2 \delta\mathbf{u} \cdot \mathbf{w} + \nabla \delta\mathbf{u} : \delta\mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ & \quad - \int_H (\delta\rho \partial_t^2 \mathbf{u}_0 \cdot \mathbf{w} + \nabla \mathbf{u}_0 : \delta\mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3, \end{aligned} \quad (19)$$

where the model perturbations, $\delta\rho = \rho - \rho_0$ and $\delta\mathbf{C} = \mathbf{C} - \mathbf{C}_0$, vanish outside the heterogeneity H . The above weak form, along with its equivalent formulations such as its strong form (Hudson & Heritage 1981) or representation theorem (Dalkolmo & Friederich 2000), has usually been used as the starting point to establish a perturbation theory. Such a weak form exhibits two differences from the original one in eq. (1): on the left-hand side, the wave operator in Ω is determined by the background model instead of the perturbed one and, on the right-hand side, two volume sources over H replace the moment tensor. Despite such apparent differences, it does not simplify the solution because $\delta\mathbf{u}$ appears on both its left and right-hand sides.

If the scattering strength of the heterogeneity is so weak that $|\delta\mathbf{u}| \ll |\mathbf{u}_0|$ in H , we can make eq. (19) explicit by ignoring the first volume integral on its right-hand side:

$$\begin{aligned} & \int_{\Omega} (\rho_0 \partial_t^2 \delta\mathbf{u} \cdot \mathbf{w} + \nabla \delta\mathbf{u} : \mathbf{C}_0 : \nabla \mathbf{w}) \, d\mathbf{r}^3 \\ & \approx - \int_H (\delta\rho \partial_t^2 \mathbf{u}_0 \cdot \mathbf{w} + \nabla \mathbf{u}_0 : \delta\mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3. \end{aligned} \quad (20)$$

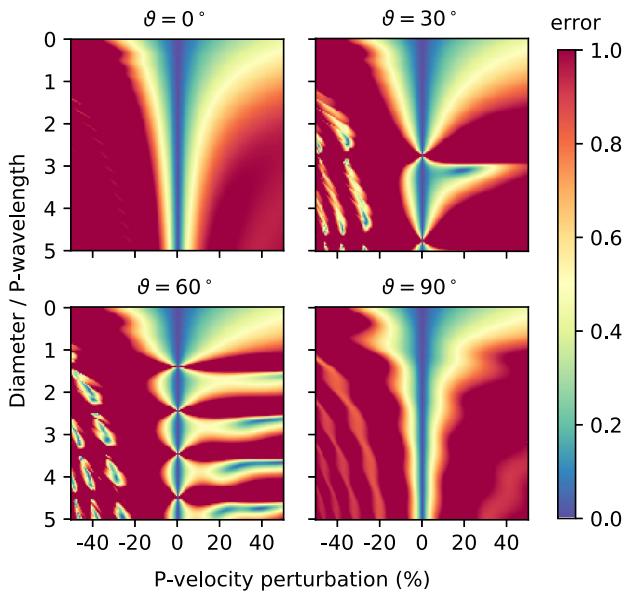


Figure 6. Error of the Born approximation as a function of perturbation strength, structure-wavelength ratio and scattering angle (θ). We show the errors for P2P scattering based on the analytical solution of Ying & Truell (1956), assuming a spherical inclusion in an unbounded homogeneous isotropic medium ($v_p = 10 \text{ km s}^{-1}$, $v_s = 5 \text{ km s}^{-1}$, $\rho = 4 \text{ kg m}^{-3}$). These errors are calculated by $|(\Psi_{\text{Born}} - \Psi_{\text{exact}})/\Psi_{\text{exact}}|$, where Ψ is a scalar potential such that the scattered P -wave in the far field is determined by $\delta\mathbf{u} = -\nabla\Psi$ [see eq. (10) of Ying & Truell 1956]. Similar patterns can be shown for P2S, S2P and S2S. The limit $d/\lambda_p = 0$ stands for Rayleigh scattering and $d/\lambda_p \leq 1$ for Mie scattering.

This is commonly known as the *Born approximation*. The solution now becomes straightforward because the model perturbations no longer appear in the wave operator; instead, they contribute only through the source: a vector force field $\delta\rho\partial_t^2\mathbf{u}_0$ and a moment tensor field $\nabla\mathbf{u}_0 : \delta\mathbf{C}$. Most remarkably, it establishes a linear relation between $\delta\mathbf{u}$ and $(\delta\rho, \delta\mathbf{C})$ so that each perturbed point in H contributes independently to the scattered wavefield by means of a simple superposition. Such linearity has laid the cornerstone of many inverse techniques such as finite-frequency tomography (Dahlen *et al.* 2000), the random media theory (Shearer 2015), stacking and migration (Rost & Thomas 2002) and each update in gradient-based full-waveform inversion (e.g. Tape *et al.* 2009; Bozdağ *et al.* 2016).

The Born approximation can be accurate only if the heterogeneity is sufficiently small and weak. An comprehensive theoretical study has been carried out by Hudson & Heritage (1981). Here we assess its accuracy based on the exact solution of Ying & Truell (1956) for P2P scattering by a spherical heterogeneity in an unbounded homogeneous isotropic medium. A similar assessment can be found in Korneev & Johnson (1993), but our results will be more compact and readable. Fig. 6 shows the error of the Born approximation as a function of the perturbation strength ($-50\% \leq \delta v_p \leq 50\%$), the structure-wavelength ratio ($0 \leq d/\lambda_p \leq 5$) and the scattering angle ($0^\circ \leq \theta \leq 90^\circ$). Such a parameter space should be able to cover all known kinds of heterogeneities in the Earth's mantle. The patterns shown in Fig. 6 reveal three remarkable characteristics. First, given a weak perturbation (e.g. $|\delta v_p| < 5\%$), the accuracy can remain relatively high even the diameter has grown considerably larger than the wavelength; this result supports the utilization of sensitivity kernels in global tomography for inverting large-scale mantle structures. On the contrary, a strong perturbation (e.g. $|\delta v_p|$

> 20 per cent) will lead to a persistently large error even at the limit of Rayleigh scattering by an infinitesimal scatterer ($d/\lambda_p = 0$); therefore, the Born approximation is inadequate for a strong mantle heterogeneity such as a ULVZ; only within the regime of Mie scattering ($d/\lambda_p \leq 1$), the long-wave approximation (Gubernatis 1979) can be applied to correct the Born solution for an arbitrarily strong perturbation strength. Finally, a non-zero scattering angle can make the pattern highly oscillating, as dominated by a series of singular points along the axis of $\delta v_p = 0$.

With such an overall pattern of error in mind, we now examine the waveform differences between the Born and the exact solutions. We continue with the benchmark problem described in Section 3.1, decreasing the radius of the heterogeneity to 30 km and increasing the frequency to 1 Hz. Such a size and frequency give rise to a structure-S-wavelength ratio greater than 9 ($\approx \frac{30 \text{ km} \times 2}{6.5 \text{ km/s} \times 1 \text{ s}}$), so the heterogeneity is 'large' in the sense of wavelength. We keep using the strong velocity perturbations, that is, $\delta v_s = 50$ per cent and $\delta v_p = 20$ per cent, but no longer adopt a smeared outer boundary. In short, the heterogeneity is both strong and large, with complex scattering phenomena all along its sharp boundary. Note that we do not use another independent method to compute the Born solution; instead, we compute the two exact solutions, respectively, for $\delta\lambda = 1$ per cent and $\delta\mu = 1$ per cent by wavefield injection; these two 'unit' solutions can be scaled by any desired perturbation strength based on the linearity of eq. (20). Besides, we let $\delta\rho = 0$ to apply the long-wave approximation by Gubernatis (1979). The narrowest Γ_{inj} in the mesh to enclose the heterogeneity has a maximum s around 34 km, filling the far field with $n_u^{\text{het}} = 99$ [$\approx \frac{\pi \times 6 \times 34 \text{ km}}{6.5 \text{ km/s} \times 1 \text{ s}}$, eq. (4)] and leading to a measured computational cost of 71270 CPU-hr (11.60 hr on 6144 cores) for a 30-min record length.

Fig. 7 shows the scattered waveforms at the same receiver as in Fig. 5, located at 60° in epicentral distance and 20° in azimuth. On the top, we show a half-hour time window where we manage to identify most of the scattered phases. Below that, we zoom into three narrow windows, varying the perturbation strength and the frequency. Each column shares the same time window and each row the same perturbation strength and frequency. Below we analyse the waveform differences based on the zoomed windows:

- (i) *Strong and large heterogeneity*: the first row shows the waveforms obtained with $\delta v_s = 50$ per cent at 1 Hz. The Born approximation basically fails for such a strong and large heterogeneity. Clearly, the Born solution exhibits more individual scattered phases than the exact solution does, such as the two separated S2P phases in Fig. 7(a); this is understandable because different parts of the heterogeneity behave as independent scatterers in the Born approximation.
- (ii) *Weak and large heterogeneity*: in the second row, we reduce δv_s to 5 per cent and, as expected, the waveform differences significantly decrease. Still, such differences are visible because the heterogeneity is one order of magnitude larger than the wavelength.
- (iii) *Strong and small heterogeneity*: in the last row, we filter the waveforms in the first row with a low-pass at 10 s, so the size of heterogeneity becomes slightly smaller than the wavelength. Even so, the Born approximation still cannot correctly predict the amplitudes of the scattered waveforms because of the strong perturbation strength. Such amplitude differences can be well corrected by the long-wave approximation by Gubernatis (1979). Such a correction can be briefly summarized into three steps: first, in the incident simulation, we store the strain at the centre of the heterogeneity, as denoted \mathbf{e}^0 by Gubernatis (1979) (it would be \mathbf{e}_0 in our notation); next, we use his eq. (4.4) and (4.10) to compute the Eshelby's

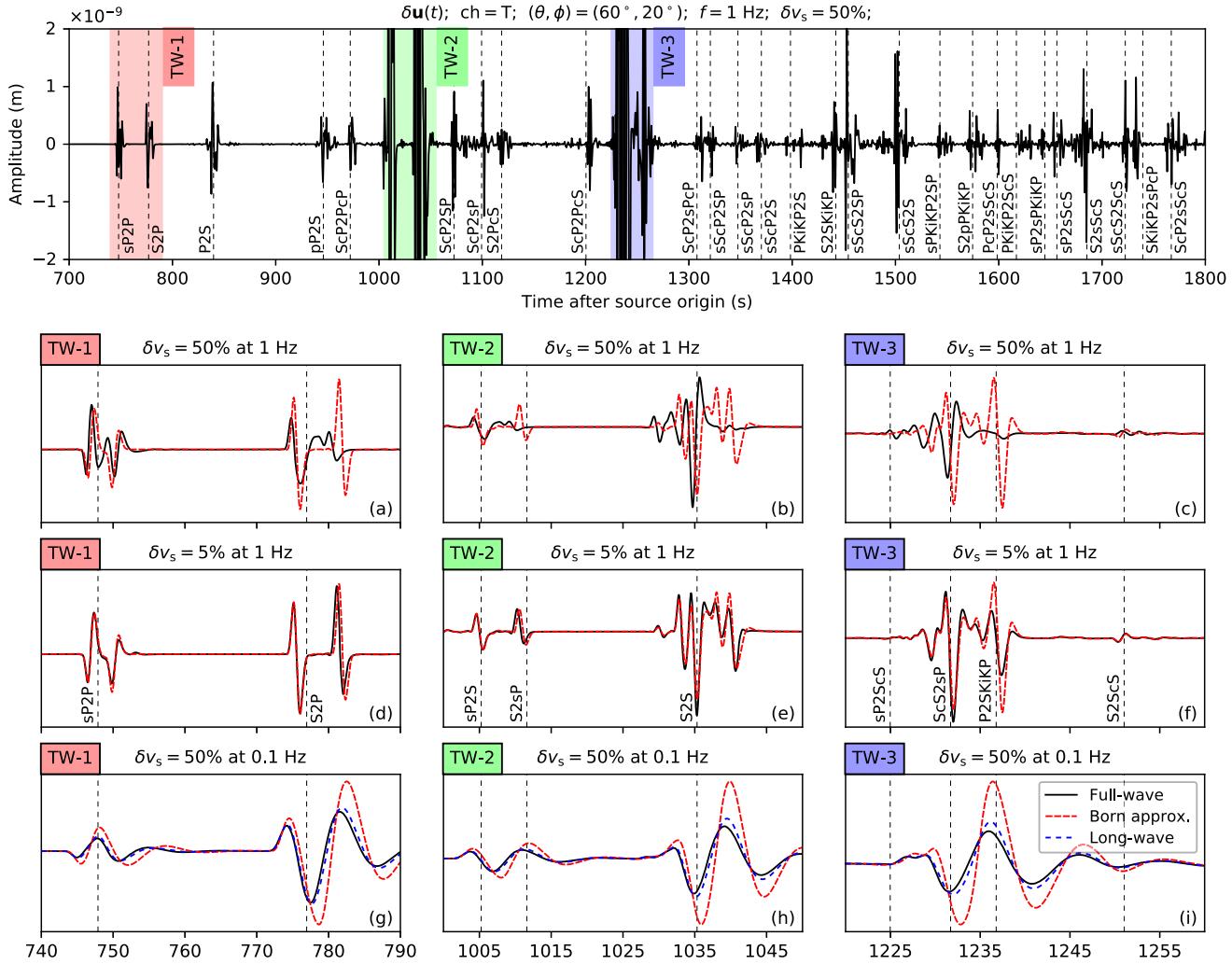


Figure 7. Waveform comparisons between 3-D modelling and the perturbation theory. The problem is same as described in Fig. 4 except that the heterogeneity has a 30 km radius (with sharp boundary) and a vanished $\delta\rho$ and that the dominant frequency is 1 Hz. The full-length waveform on the top shows the transverse component of the scattered displacement at a receiver located at 60° in epicentral distance and 20° in azimuth. Three time windows are magnified below, as labelled TW-1 (left-hand column), TW-2 (middle column) and TW-3 (right-hand column). The three rows from top to bottom show, respectively, the comparisons for $\delta v_s = 50$ per cent at 1 Hz, $\delta v_s = 5$ per cent at 1 Hz and $\delta v_s = 50$ per cent at 0.1 Hz. The waveforms labelled ‘Long-wave’ are obtained by correcting the Born solutions using the long-wave approximation by Gubernatis (1979). The traveltimes of the scattered phases are calculated by ray tracing, assuming point scattering at the centre of the heterogeneity.

eigenstrain ϵ^σ at the static limit; finally, in view of the difference between his eq. (3.8) and (3.11) and that $\delta\rho = 0$, the Born waveforms can be corrected by a scaling matrix $(\delta\mathbf{C} : \epsilon^0)^{-1} \cdot (\delta\mathbf{C} : \epsilon^\sigma)$ in the frequency domain.

4.2 2-D in-plane modelling

Different from the perturbation theory, the 2-D in-plane approximation reduces the dimensionality of the model while preserving the wave scattering physics. By confining structural variations within the source–receiver plane, it prescribes model axisymmetry about the source and thereby substantially change the computability at high frequencies. Taking a global-scale problem at 1 Hz for example, 2-D in-plane modelling (e.g. with AxiSEM, Nissen-Meyer *et al.* 2014) can lead to a speedup of 4–5 orders of magnitude with respect to full 3-D modelling. Because of such efficiency, it has

been frequently used in recent forward studies of small-scale mantle heterogeneities (e.g. Rondenay *et al.* 2010; Thorne *et al.* 2013; Vanacore *et al.* 2016; Haugland *et al.* 2017).

We work on the same problem as described in Section 4.1. In axisymmetric modelling, this heterogeneity becomes a 3-D torus with a 2585.5 km major radius (i.e. the distance from its centre to the axis) and a 30 km minor radius. The waveforms modelled in 2-D and 3-D are compared on two record sections, one within and the other orthogonal to the source-heterogeneity plane.

We first examine the in-plane record section, as shown in Fig. 8(a). The 2-D simulation correctly predicts most of the scattered phases, but the amplitudes turn out much larger, especially for the distant receivers (note that the amplitudes of the 3-D waveforms are scaled by a factor of three). These large amplitude differences result from the fact that the lateral dimension of the sphere is much smaller than that of the torus. Therefore, when using the amplitudes of

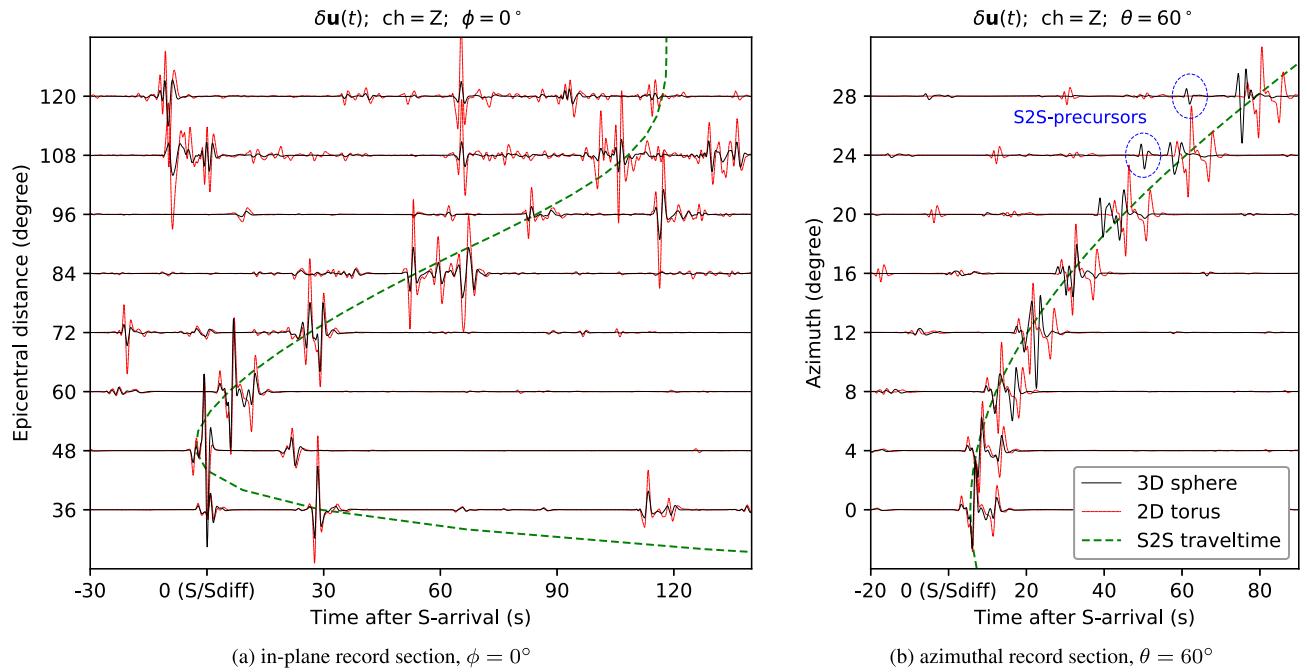


Figure 8. Waveform comparisons between 3-D modelling and 2-D in-plane modelling. The problem is the same as in Fig. 7 ($\delta v_s = 50$ per cent at 1 Hz) except that the spherical heterogeneity becomes a torus in 2-D. We show the vertical component of the scattered displacement on two record sections: (a) the in-plane record section ($\phi = 0^\circ$) and (b) an off-plane record section with $\theta = 60^\circ$. The seismograms are aligned at the incident S-wave (or Sdiff for some distant receivers). In both (a) and (b), the amplitudes of the 3-D waveforms are multiplied by a factor of 3 for visualization purpose. In (b), two additional corrections are applied: first, the 2-D waveforms are shifted by the difference between the two S2S traveltimes, respectively, in 3-D and 2-D (otherwise they will look exactly the same at all the azimuths) and, secondly, the amplitudes of the 3-D waveforms are scaled by another two factors to further account for the differences in geometric spreading and scattering angle. The traveltimes of the scattered phases are calculated by ray tracing, assuming point scattering at the centre of the heterogeneity.

scattered waveforms modelled in 2-D to constrain the perturbation strength of a heterogeneity, one must be aware of the geometric discrepancy introduced by the in-plane assumption. For instance, such discrepancy can be small for a subducted slab (Haugland *et al.* 2017) but rather large for a ULVZ (Vanacore *et al.* 2016) or a plume (Maguire *et al.* 2016).

Now we move on to the off-plane record section located at $\theta = 60^\circ$, as shown in Fig. 8(a). Here we focus on the S2S phase. A direct comparison will be silly because the 2-D waveforms do not change with azimuth (note that we use an axisymmetric moment tensor). To make the comparison more meaningful, we first attempt to correct for the path difference of S2S in 2-D and 3-D modelling based on the ray theory: first, we delay the 2-D waveforms by the differential traveltimes of S2S ($t_{3\text{-D}} - t_{2\text{-D}}$); secondly, in addition to a constant factor of three, we scale the 3-D waveforms by two more factors, one for geometric spreading (based on the distance from heterogeneity to receiver) and the other for the scattering angle [based on the ‘shape factor’ given by eq. (4.8) in Gubernatis (1979) or eq. (38) in Wu & Aki (1985)]. Despite such ray-based corrections, the waveform differences between 2-D and 3-D modelling still increase rapidly with azimuth: the two waveforms become mostly uncorrelated after $\phi = 8^\circ$. Indeed, the S2S phase contains two sub-phases, travelling, respectively, as P and S within the heterogeneity. In 3-D, these two subphases take off at different locations on the surface of heterogeneity and such a difference increases with the azimuth. After $\phi = 20^\circ$, these two subphases start to arrive separately, forming an apparent precursor of the ray-based S2S phase, as marked in Fig. 8(a).

Consequently, we can infer from this numerical example that, for a heterogeneity with a small lateral extent, the only robust measurement one can directly take from 2-D in-plane modelling is the phase information at a receiver within or sufficiently close to the source-heterogeneity plane.

5 AN APPLICATION TO A ULVZ

In this section, we carry out a 3-D study of a ULVZ at a 1 Hz dominant frequency. We adopt an earthquake source located in northern Chile and compute the seismograms across the USArray. The ULVZ is located off the east coast of Costa Rica, roughly halfway between the epicentre and the geographic centre of the United States. It is 15 km in height and 60 km in radius and with a strong 50 per cent shear velocity reduction. The problem is detailed in Fig. 9.

We incorporate Crust 1.0 (Laske *et al.* 2013) underneath the USArray to investigate the crustal effects on the ULVZ phases. This 3-D crustal model has a $1^\circ \times 1^\circ$ horizontal resolution and includes up to six vertical layers from the surface to the Moho (three for sediment and three for crystalline crust). It also includes topography on both the surface and the Moho: the undulation on the Moho comes mostly in the opposite direction of that on the surface and can be two to three times larger than that on the surface. Wave scattering by the ULVZ is modelled by wavefield injection while Crust 1.0 is incorporated as part of the background model. Similar to SPECFEM3D_GLOBE, we exactly honour the topography on both the surface and the Moho by stretching the spectral

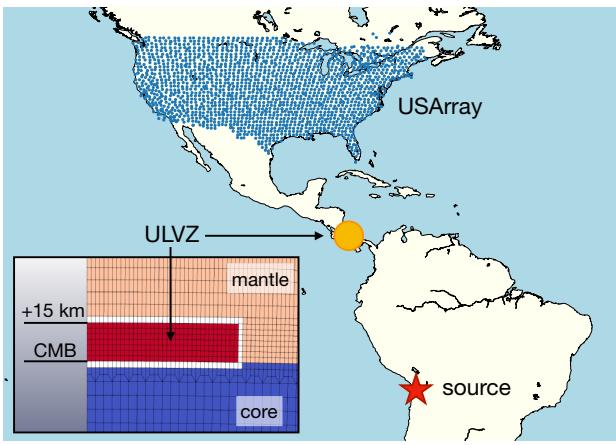


Figure 9. A 3-D forward study of a ULVZ. The epicentre is located at 20.02°S, 69.23°W and 94.5 km in depth (200506132244A, northern Chile). The seismic array contains 1843 stations from the transportable virtual network ‘US-TA’ of the USArray. We incorporate 3-D crustal structures (Crust 1.0; Laske *et al.* 2013) underneath the array. The ULVZ has a cylindrical shape with a height of 15 km and a radius of 60 km, located at 10.11°N, 82.34°W above the CMB. Within the ULVZ, the S- and P-velocities are reduced, respectively, by 50 and 20 per cent and the density increased by 10 per cent. The spectral element mesh is generated based on PREM (no crust) at a 1 Hz frequency, with the lowermost mantle refined to accommodate the 50 per cent S-velocity reduction. The boundary of the ULVZ is exactly honoured by the mesh. The elements coloured in red form the ULVZ and those in white are the injection elements.

element mesh (by means of the particle relabelling transformation in AxiSEM3D, Leng *et al.* 2019) while leaving the intermediate layer boundaries cutting through the elements. Note that we do not include any crustal structure on the source side to isolate the receiver-side effects. Besides, we need a reference solution for PREM without Crust 1.0; in this solution, we also remove the two 1-D crustal layers in PREM (by extending the top of the mantle to the surface) so as to exclude all crustal effects.

Because $\delta v_s = -50$ per cent, the mesh generated based on PREM must be doubled in the lowermost 15 km of the mantle; in other words, the highest dominant frequency actually reaches 2 Hz for PREM-based mesh generation. To keep the mesh conformal, such refinement will be extended globally in the horizontal (θ) direction and to the mid-lower mantle in the vertical (r) direction, leading to $C_{\text{het}} \approx 1.3$ in eq. (18). Note that our mesh fully honours the boundary of the ULVZ, including its top, bottom and side faces. The narrowest Γ_{inj} in the mesh to enclose the ULVZ has a maximum s around 62 km (see the injection elements coloured in white in Fig. 9) and $v_s \approx 7.3 \text{ km s}^{-1}$ on the CMB, leading to $n_u^{\text{het}} = 160 [\approx \frac{\pi \times 6 \times 62 \text{ km}}{7.3 \text{ km s}^{-1} \times 1 \text{ s}}, \text{eq. (4)}]$ in the far field. Within the 3-D crust, we use $n_u^{\text{bm}} = 1500$ (Leng *et al.* 2019), which means $n_u = n_u^{\text{het}} + n_u^{\text{bm}} = 1660$.

Using PREM as the background model, the measured computational cost is 152,712 CPU-hr (21.21 hr on 7200 cores) for a 30-min record length, which increases to 188,352 CPU-hr (19.62 hr on 9600 cores) after adding Crust 1.0. Because the farthest receiver is about 52° away from the ULVZ, these measured costs can be reduced by 2/3 simply with a truncated computational domain at $\theta = 60^\circ$.

5.1 Scattering by the ULVZ

We first investigate the scattering process in the near field of the ULVZ. Fig. 10 shows a series of 3-D wavefield snapshots in the near

field since the S-wave incidence, animating the radial component of the total displacement. These snapshots can be better understood with the help of the supplemented animation. Let us start with the incident waves. At 590 s, a moment when the S wave front nearly touches the ULVZ, Fig. 10(a) displays a Y-shaped pattern composed of three continuous wave fronts: S, ScS and SK (note that ScP does not exist at this distance); in Fig. 10(b), S and ScS form the two parallel wave fronts in the ‘TOP’ cross section, which merge into one on the CMB. The complete scattering process can be roughly divided into the following three stages:

(i) *The wave-trapping stage* — after the incidence (590 s) and before the S wave front approaches the rear corner of the ULVZ (608 s), the scattering is dominated by the top and the bottom faces. The incident S wave is first reflected and refracted by the top face; the down-going refracted waves then encounter the CMB where part of the energy is transmitted to the outer core and the rest reflected back to the top face; such top-bottom reflections and refractions (accompanied by a deflection along the circular side face) repeat until the S wave front approaches the rear corner. In short, the ULVZ behaves as a ‘leaking’ wave trap in this stage.

(ii) *The point-source stage* — the circular side face starts to dominate as the S wave front enters the rear quarter of the ULVZ. It focuses the trapped waves at the rear corner where they eventually escape from the ULVZ; in other words, the rear corner of the ULVZ emits the trapped energy to the far field as a point source. We track two phases on the CMB in Fig. 10(b), as marked, respectively, by ④ and ⑤: a comparison between 602 and 608 s clearly exhibits the focusing of these two phases, whereas, at 614 s, ④ has escaped from the rear corner and ⑤ has been mostly focused and ready to escape; dissimilar to ④, both ④ and ⑤ are highly irregular after the takeoff, as shown in Fig. 10(a) at 620 s, caused by the extraordinarily strong 3-D effects around the rear corner (see the supplemented animation);

(iii) *The back-propagation stage* — while most of the trapped energy escapes from the ULVZ at its rear corner, the rest reflected back to the ULVZ continues to propagate towards its front corner. Once again, such back-propagation is dominated first by the top and the bottom faces as a wave trap and then by the front corner as a point source. The back-propagation can hardly be perceived from Fig. 10 but is visible in the supplemented animation. This stage seems less significant for probing a ULVZ because the back-scattered energy may be too small to be observed on the Earth’s surface; still, it might suggest a hunt for such small energy by some array methods (Rost & Thomas 2002).

Next, we analyse the scattered waveforms computed for the USArray, as shown in Fig. 11. The time window in Fig. 11(a) is dominated by two outstanding phase clusters: (P+pP+sP)2S and (S+pS+sS)2S. The phases within the two clusters maintain the same moveout (i.e. their stacking patterns are parallel to one another), as their traveltimes only come from the incident branch. We zoom into the second cluster in Fig. 11(b), where the three scattered phases become distinguishable: S2S, pS2S and sS2S (the incidence of pS and sS are annotated in Fig. 10a at 620 s). The S2S phase consists of four subphases at a shorter distance, as marked by ① to ④ in Fig. 11(b), which correspond to the four wave fronts departing from the top face of the ULVZ during the wave-trapping stage, as marked by ④ in Fig. 10 at 608 s. As the distance increases, these four subphases gradually merge because of the transition of scattering regime from wave-trapping to point-source. Though we know that a subphase observed at a longer distance originates from

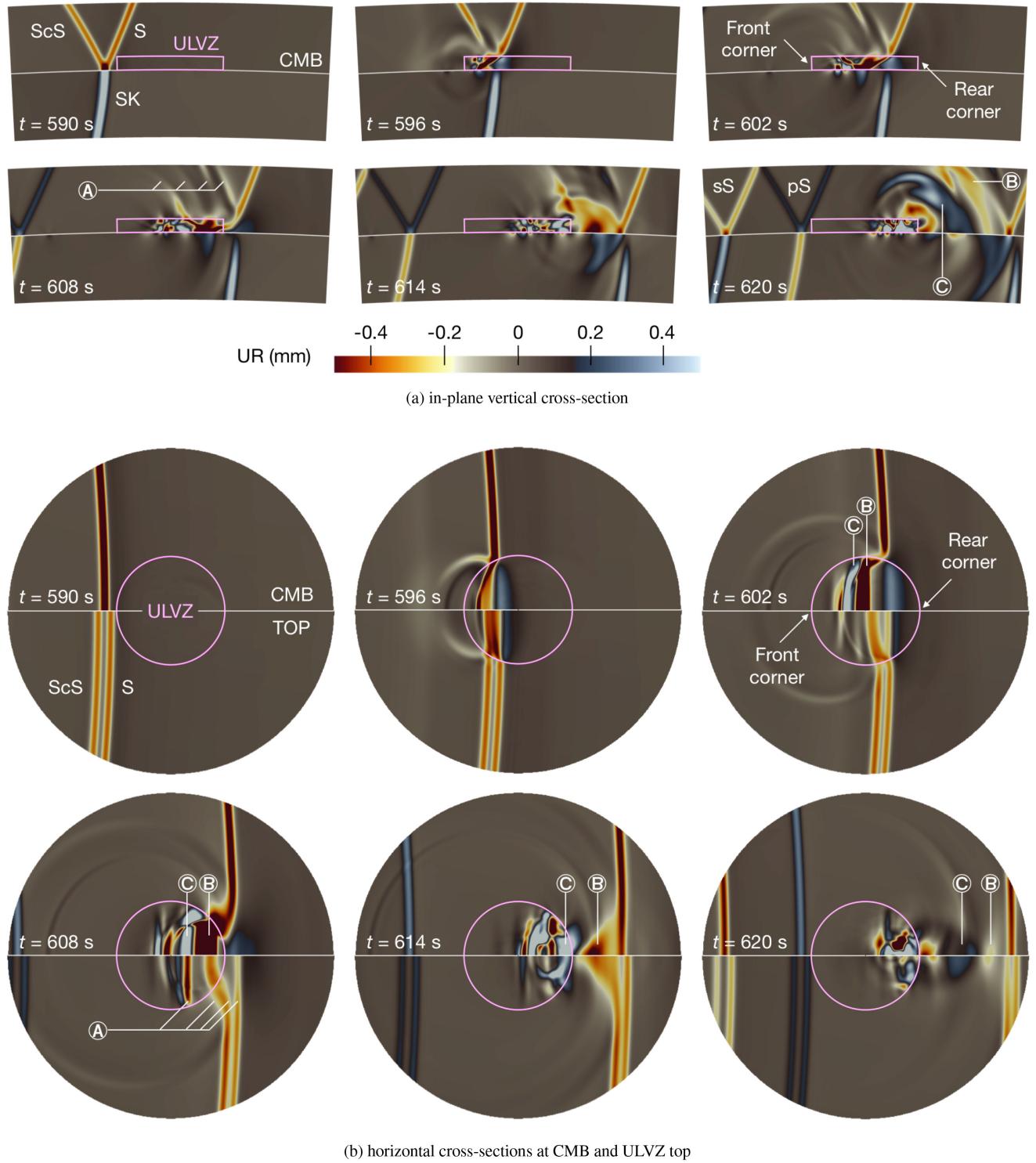


Figure 10. Wavefield snapshots around a ULVZ since the incidence of S wave. The problem is described in Fig. 9, simulated at a 1 Hz dominant frequency. In (a), we show the in-plane vertical cross-section determined by the epicentre and the ULVZ centre. In (b), we show two horizontal cross-sections centred at the ULVZ, one on the CMB (upper semicircle) and the other on the top of the ULVZ (lower semicircle). The radial component of the total displacement is visualized, which is discontinuous on the CMB (because the CMB is a solid–fluid discontinuity). The wave fronts marked \textcircled{A} , \textcircled{B} and \textcircled{C} will be examined in the text. The complete animation can be found in the supplementary materials.

the rear corner of the ULVZ, it is difficult to exactly identify their origin in the near-field snapshots because the wave fronts emanating from the rear corner are highly irregular, such as \textcircled{B} and \textcircled{C} in Fig. 10.

In general, the scattering process in the near field is highly sensitive to the location and geometry of a ULVZ. Varying the location and geometry, in turn, can lead to distinct waveform patterns in the far field, which usually involves generation and suppression

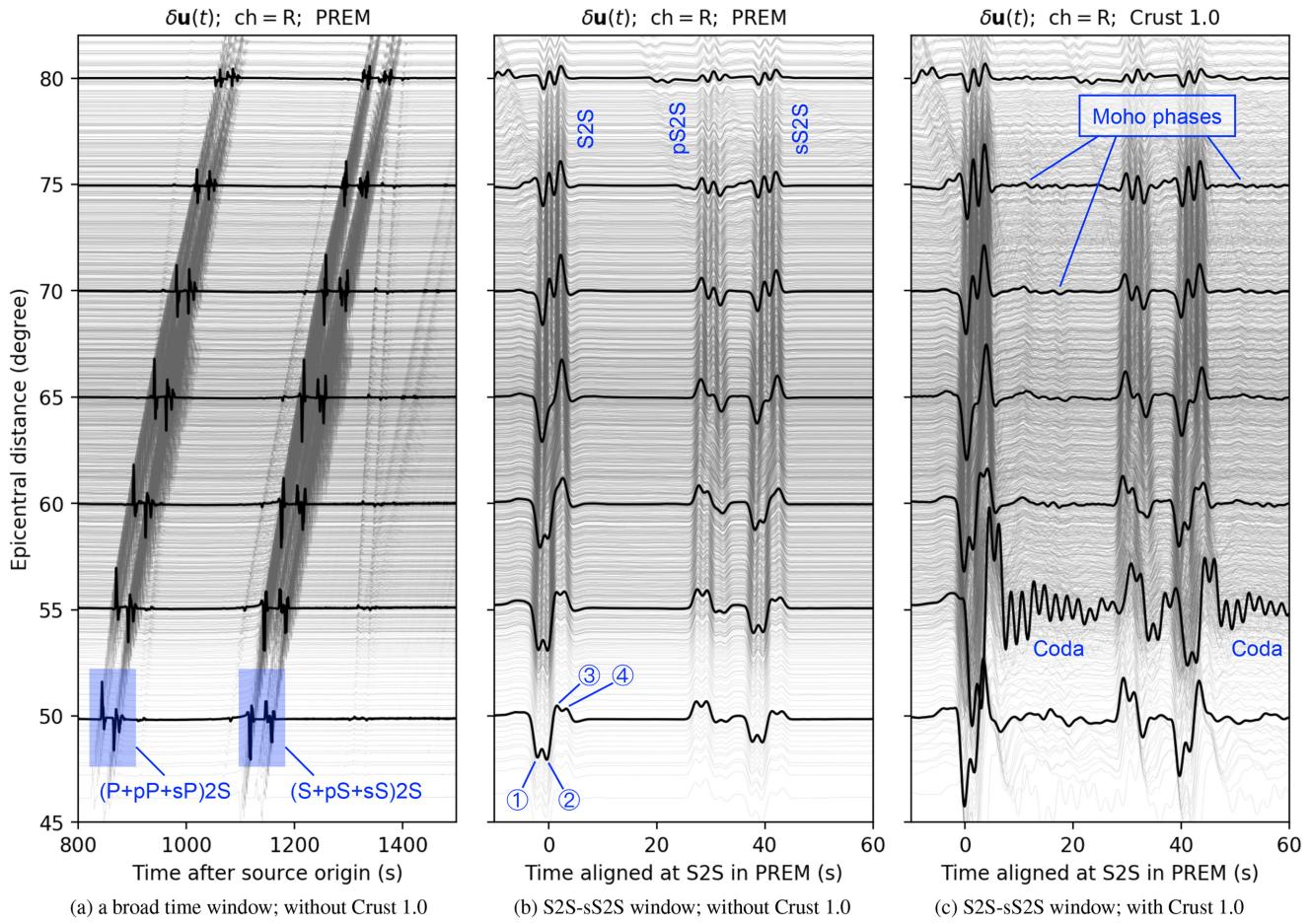


Figure 11. Synthetic scattered waveforms originating from a ULVZ and computed for the USAArray. The problem is described in Fig. 9, simulated at a 1 Hz dominant frequency. We show the radial component of the scattered displacement. The broad time window in (a) contains many scattered phases, among which the largest two clusters are $(P+pP+sP)2S$ and $(S+pS+sS)2S$. The window of $(S+pS+sS)2S$ is magnified in (b) and (c), respectively, without and with Crust 1.0. The strong coda near 55° in (c) are caused by the thin and slow sediment layers in Crust 1.0; such strong coda also appear at some other distances, and we only highlight one. The same amplitude scale is used across this figure. The traveltimes of the scattered phases are calculated by ray tracing, assuming point scattering at the centre of the ULVZ.

of scattered phases. A profound understanding of such non-linear scattering effects (such as above) should be fundamental to data interpretation and utilization, for which high-frequency 3-D modelling is the best approach (or probably the only approach under certain scenarios). A parameter study for a 3-D ULVZ is beyond the scope of this paper and left to a future one.

5.2 Effects of 3-D crust

Now we investigate the 3-D crustal effects. Fig. 11(c) shows the waveforms of $(S+pS+sS)2S$, the same as Fig. 11(b) but with Crust 1.0. There are primarily three crustal effects. First and foremost, the waveform patterns of the ULVZ phases are basically preserved, but they are notably delayed and amplified by the slow crustal layers, especially by the sediment layers. Secondly, strong coda are generated at some of the stations, mostly located in the sedimentary basins, whose amplitudes are similar to those of the ULVZ phases. Finally, quite a few Moho phases are generated; however, they are less likely to be observed in data because of their small amplitudes (note that the ULVZ phases are already much smaller than the main phases).

By cross-correlating the waveforms in Figs 11(b) and (c), we can quantify the traveltime delay caused by Crust 1.0 at each station. For the S2S window, such delays are mapped in Fig. 12(a), which exhibits some of the large-scale geological features of the United States. Using such modelling results, we can evaluate the accuracy of the ray-based traveltime correction, a common practice in seismic tomography (chap. 13, Nolet 2008). Fig. 12(b) shows the map obtained by ray tracing in Crust 1.0; this map complies with Fig. 12(a) in many regions, especially in the sedimentary basins (e.g. those in California, Louisiana and Florida), but significantly overestimates the delay around the Rocky Mountains. As shown in Fig. 12(c), such large errors in the mountains can be well remedied by ignoring the sediment layers in Crust 1.0; however, doing so will overturn the accuracy in the sedimentary basins. Combining Figs 12(b) and (c), we figure out that the discrepancy between waveform modelling and ray tracing stems from the sediment layers thinner than the wavelength; these layers will be poorly sampled by the spectral element mesh and thus cannot effectively alter the waveforms. By ignoring these subwavelength layers in Crust 1.0, we manage to render ray tracing consistent with waveform modelling, as shown in Fig. 12(d). This example justifies the necessity of considering finite-frequency

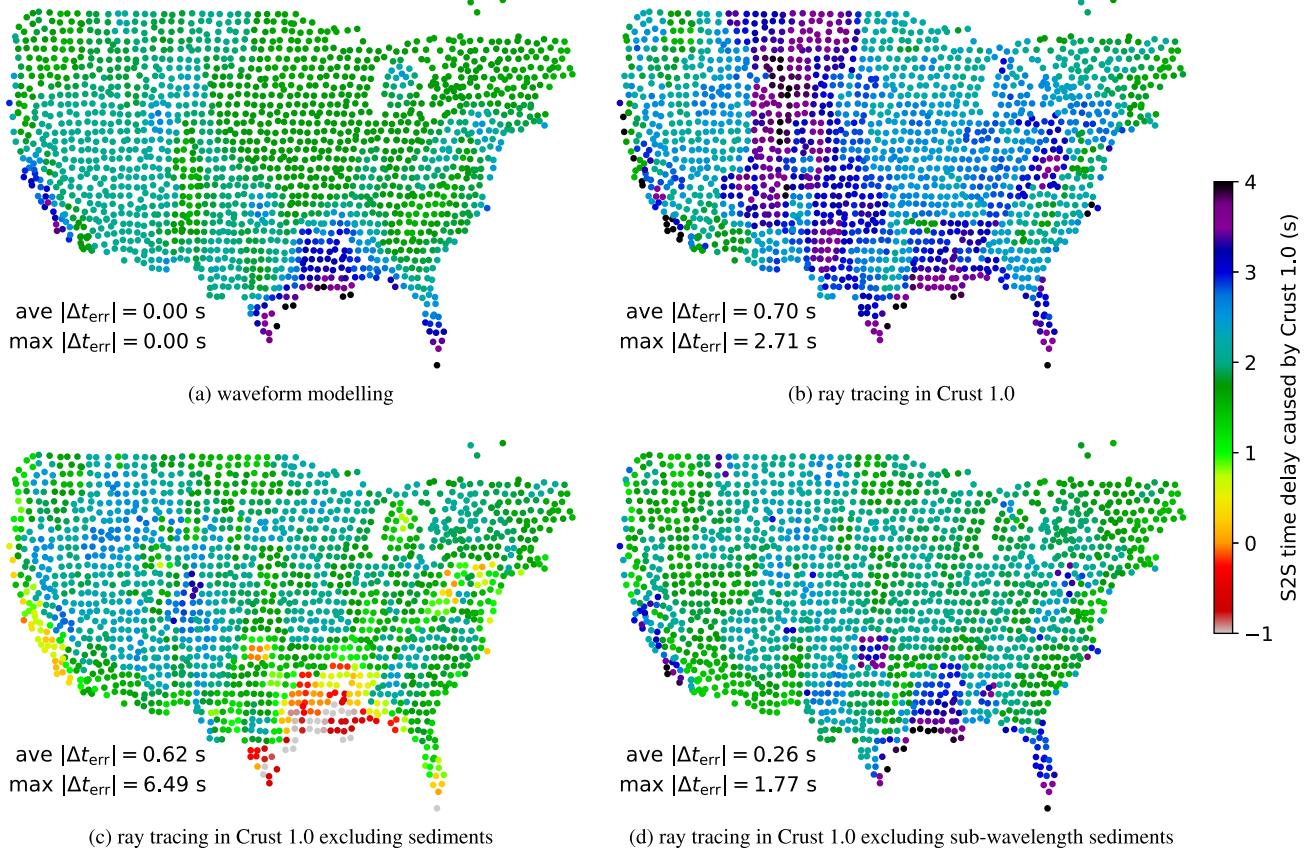


Figure 12. Traveltime delays of S2S phases caused by Crust 1.0. In (a), the delays are obtained by cross-correlating the synthetic waveforms in Fig. 11(c) against those in Fig. 11(b), taking a time window between ± 10 s. From (b) to (d), the delays are calculated by 1-D ray tracing in PREM with Crust 1.0 (using the vertical profile beneath the station as a global 1-D crust). In (b), all the layers in Crust 1.0 are incorporated, including up to three sediment layers. We exclude all the sediment layers in (c) and those thinner than one wavelength in (d). We compare the delays in (b)–(d) with those in (a) and denote the difference by Δt_{err} ; the maximal and average Δt_{err} over all the stations are indicated.

effects when correcting traveltimes for crustal structures (Dahlen 2005).

5.3 Imaging by migration

As the final part of this case study, we use our synthetic waveforms to image the ULVZ based on a simple migration algorithm. We use the scattered waveforms computed with Crust 1.0, as shown in Fig. 11(c), and assume S2S to be the only scattering mechanism. It means that any wiggles outside the S2S window (around the origin in Fig. 11c) will behave as noise, including the other ULVZ phases (such as pS2S) and the crustal phases and coda. Given any target point \mathbf{r}_m in the mantle, we compute its migration energy $E(\mathbf{r}_m)$ by

$$E(\mathbf{r}_m) = \sum_{k=1}^N \int_{-T}^T |\delta u(\mathbf{r}_r^k; \mathbf{r}_s \rightarrow \mathbf{r}_m + \mathbf{r}_m \rightarrow \mathbf{r}_r^k + t)|^2 dt, \quad (21)$$

where N denote the number of stations in the array, \mathbf{r}_r^k the location of the k th station, $t_{\mathbf{r}_1 \rightarrow \mathbf{r}_2}$ an operator that computes the S-wave traveltimes between \mathbf{r}_1 and \mathbf{r}_2 and T the dominant period of the scattered waveforms. The above equation should be the simplest algorithm to create a migration image: it first calculates the theoretical traveltimes of S2Ss, that is, $t_{\mathbf{r}_s \rightarrow \mathbf{r}_m} + t_{\mathbf{r}_m \rightarrow \mathbf{r}_r^k}$ for all \mathbf{r}_m and \mathbf{r}_r^k , and then the L^2 norms of the waveforms within a window $[-T, T]$ centred at S2S (using a boxcar window function). The only variable in eq. (21)

is the period T , for which we test 5, 2 and 1 s; a low-pass filter at T is first applied to the synthetic data.

The results are shown in Fig. 13. The obtained image converges to the real location and geometry of the ULVZ as the frequency increases. In particular, a hollow structure has been formed when $T = 1$ s, which means that migration has started to resolve the structural details at such a period. In view of that our migration algorithm simply assumes Rayleigh scattering (even less realistic than Born scattering), such a result looks satisfactory. The images also display strong path effects (essentially the isochronal artefacts, Kito & Korenaga 2010), that is, they are stretched along the ray paths in the in-plane cross-section. Migration with real data is much more challenging than with synthetic data and requires more sophisticated data processing; for example, the low signal-to-noise ratio of global seismic data may necessitate a non-linear stacking algorithm (e.g. Rost & Thomas 2002; Korenaga 2013) and stochastic test of reproducibility (Korenaga 2015).

6 DISCUSSIONS

We develop a new approach to simulate 3-D wave propagation in the presence of a strong, localized heterogeneity in the Earth's mantle at a high frequency. Distinct from approximate methods such as the perturbation theory, our approach preserves the full physics of wave scattering and delivers the exact elastodynamic wavefield originated from the heterogeneity. It is built upon two constituents: a 3-D wave

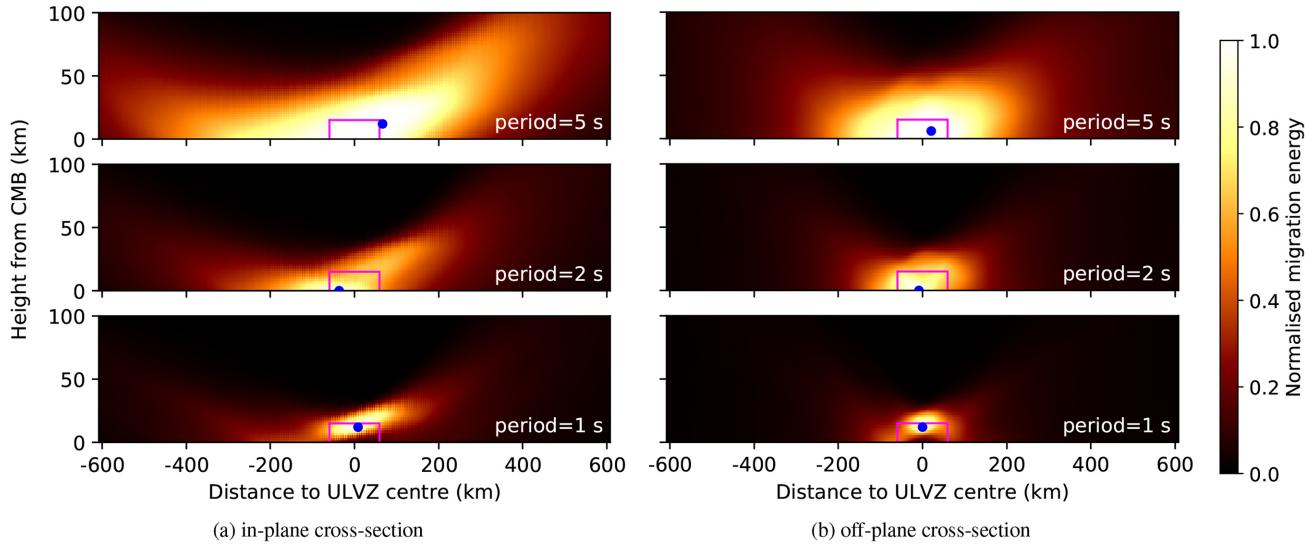


Figure 13. ULVZ geometry imaged by migration. We use the scattered waveforms in Fig. 11c, computed with Crust 1.0 and corrected using the traveltime delays in Fig. 12d. We test three low-pass filters, respectively, down to 5, 2 and 1 s. For migration, we assume S2S-scattering but do not mask any other scattered phases (including crustal phases and coda). The obtained images for the in-plane cross-section are shown in (a) and those for the orthogonal off-plane cross-section in (b). The small box indicates the real ULVZ boundary (note that the vertical axis is stretched) and the spot within the box the centroid of the ULVZ yielded by migration.

propagator called AxiSEM3D and a cross-scale technique known as wavefield injection. Given a background model, wavefield injection can realize the exact decomposition of the total wavefield into an incident and a scattered part, that is, $\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}$, which can be solved with two sequential simulations. Such a decomposition is significant in the scenario of deep Earth scattering because both \mathbf{u}_0 and $\delta\mathbf{u}$ are physically *quasi-axisymmetric*, that is, they have limited azimuthal variability, respectively, about the earthquake source and the small-scale heterogeneity. With adaptability to the azimuthal complexity of a wavefield, AxiSEM3D can efficiently solve \mathbf{u}_0 and $\delta\mathbf{u}$ under the two different coordinate systems. For a simple and flexible implementation in SEM, we design a new traction-free scheme that is (i) purely displacement-based, (ii) agnostic to the geometry of the injection boundary and (iii) encapsulated in the computation of the stiffness forces. Based on this scheme, we implement wavefield injection in AxiSEM3D and verify our implementation with a comprehensive, nearly error-free benchmark solution against a full 3-D SEM.

Our approach is more than another implementation of wavefield injection with some different solver. It follows a unique, physics-driven strategy where the adaptability of AxiSEM3D plays the leading role—an apparent sign is that it uses the same solver for both \mathbf{u}_0 and $\delta\mathbf{u}$. As a distinguishing feature, it allows for a 3-D background model (while remaining at a high frequency), such as a tomographic mantle or crustal model, preserving any far-field propagation effects such as attenuation and full anisotropy (Tesoniero *et al.* 2020). Further, because it does not require the third stage of wavefield extrapolation (see Section 6.1 for further discussions) and thus a domain truncation around the injection boundary, it yields the exact solution (i.e. identical to the all-in-one solution) without ignoring any higher-order interactions between the interior and the exterior of the injection boundary.

On top of these features, the major appeal of our approach lies in its computational efficiency. A 1 Hz dominant frequency is readily achievable for a global-scale problem with strong deep Earth scattering using modern HPC (on an order of magnitude of 10^5 CPU-hr).

Such efficiency stems from the quasi-axisymmetry of the scattered wavefield, embodied in AxiSEM3D through a Fourier expansion order field or $n_u(s, z)$ that is limited by the small lateral size of the heterogeneity. Given a 3-D heterogeneity in a 1-D background Earth model, $n_u(s, z)$ can be determined by eq. (17), whose computational cost is given by eq. (18). Such a computational cost scales with the lateral size of the heterogeneity (e.g. the cost is the same for a ULVZ and a deep-rooted plume with the same lateral size). For a 100 km heterogeneity within PREM, our measurements reveal a speedup of two orders of magnitude with respect to a convectional 3-D SEM; such a speedup is frequency-independent and inversely proportional to the lateral size of the heterogeneity. eq. (17) is an upper bound of $n_u(s, z)$ that guarantees an exact wavefield solution throughout the Earth's interior, so our cost measurements represent the highest end. Using a 3-D background model increases the computational cost to different extents, normally within an order of magnitude unless the heterogeneity is exceedingly small. At the low- and high-frequency ends, the computational cost is dominated, respectively, by the background model and the heterogeneity.

Two global-scale numerical experiments are conducted at 1 Hz. In the first one, we consider a spherical heterogeneity in the mid-lower mantle to evaluate two types of approximate solutions. We show that the Born approximation cannot deliver correct waveforms for a strong and large (measured in wavelength) heterogeneity, whereas its non-linear extension can correct the waveforms for a strong perturbation strength but available only for a subwavelength heterogeneity. We also show that, if the heterogeneity has a small lateral extent, the only robust measurement that can be taken from 2-D in-plane modelling is the phase information at a receiver within or sufficiently close to the source-heterogeneity plane. In the second experiment, we study a 3-D ULVZ, incorporating 3-D crustal structures on the receiver side as part of the background model. We zoom into the scattering process in the near field, through which the origin of each scattered phase is identified. Concerning 3-D crustal structures, a comparison between our modelling results and ray tracing suggests that the finite-frequency effects should be considered in

the crustal correction of traveltimes, particularly in the presence of thin sediment layers.

6.1 Wavefield extrapolation

In global seismology, many previous implementations of wavefield injection have adopted a global solver that can only take in a spherically symmetric or 1-D background model (e.g. Monteiller *et al.* 2012; Tong *et al.* 2014a; Pienkowska-Cote *et al.* forthcoming) for the sake of computability at a high frequency. As indicated by eq. (14), the hybrid wavefield \mathbf{u}_{hy} must be solved within the perturbed 3-D model, which cannot be accomplished by a 1-D global solver. Previous studies of this kind have had to introduce a third stage known as ‘wavefield extrapolation’ (Masson *et al.* 2013; Masson & Romanowicz 2016). In this case, the second stage will be conducted within a 3-D domain truncated by an absorbing boundary Γ_{abs} enclosing Γ_{inj} , using a comprehensive 3-D local solver. Next, in the third stage, the recorded $\delta\mathbf{u}$ and $\delta\mathbf{t}$ on Γ_{inj} will be ‘extrapolated’ within the background model to obtain $\delta\mathbf{u}$ at any far-field receivers using the 1-D global solver. The theory for wavefield extrapolation is perfectly reciprocal to that for wavefield injection in terms of a role-reversal between the background and the perturbed model.

Notwithstanding the 1-D simplification, the results of wavefield extrapolation still become inexact because Γ_{abs} (even if it is a perfectly matched layer) blocks off any interactive scattering effects between the heterogeneity and the background model. Such higher-order interaction could become notable when a strong heterogeneity is located close to a major radial discontinuity of the background model, such as the Moho and the CMB. As a remedy, this Γ_{abs} can be replaced by an exact boundary condition (van Manen *et al.* 2007; Masson & Romanowicz 2017), technically accessible for a 1-D background model. The overall performance can be remarkably improved because of the truncated 3-D domain and the high efficiency of the 1-D global solver, particularly attractive for local-scale inversion with consideration for global-scale propagation effects (Masson & Romanowicz 2017; Clouzet *et al.* 2018). Further, global wave propagation (in the first and the third stages) can be replaced with a precomputed database of the Green’s functions, for example Pienkowska-Cote *et al.* (forthcoming) using Instaseis (van Driel *et al.* 2015) as the database; a precomputed database can be more efficient than wave propagation if the number of receivers is not exceedingly large. In this three-stage workflow, AxiSEM3D can be used as either the global or the local solver. Serving as the local solver, AxiSEM3D will be significantly more efficient than a conventional 3-D solver if the small-scale heterogeneity is axisymmetric by itself, as commonly encountered in the study of a ULVZ or a plume.

6.2 Heterogeneity on source or receiver side

The above three-stage workflow will reduce to a two-stage one if the heterogeneity is located near the source ($\mathbf{r}_s \in H$) or beneath the receivers ($\mathbf{r}_s \in H$), as the first stage will disappear in the former case and the third in the latter. Though this paper is aimed for small-scale heterogeneities in the mantle, our implementation naturally encompasses these two special cases. In our approach, 3-D crustal structures can be either incorporated as part of the background model or handled as a local heterogeneity by wavefield injection. The choice mainly depends on the resolution and scale of the crustal model. For instance, one may include Crust 1.0 as the background

model and impose a high-resolution regional model on the source or the receiver side as a local heterogeneity.

ACKNOWLEDGEMENTS

We thank Shihao Yuan for sharing his experience and code for implementing wavefield injection, and Mike Thorne, Surya Pachhai, Sebastian Rost for discussions about scattering by a ULVZ. We thank the reviewers, Daniel Frost and Yder Masson, for their comprehensive comments and useful suggestions, and David Al-Attar for detailed comments in his review of the predecessor of this paper. We thank the developers of Obspy (Krischer *et al.* 2015) and SPECFEM3D_GLOBE. The computations have been undertaken on ARCHER, the UK National Supercomputing Service. This work is sponsored by US NSF and UK NERC under joint grant NE/R012199/1 and by US NSF under grant EAR-1610612. KL receives supports from Wolfson College, University of Oxford as a junior research fellow and the Key Laboratory of Transportation Tunnel Engineering, Southwest Jiaotong University as an overseas lecturer.

REFERENCES

- Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D.A., Knepley, M.G. & Fichtner, A., 2019. Modular and flexible spectral-element waveform modelling in two and three dimensions, *Geophys. J. Int.*, **216**(3), 1675–1692.
- Al-Attar, D. & Crawford, O., 2016. Particle relabelling transformations in elastodynamics, *Geophys. J. Int.*, **205**(1), 575–593.
- Al-Attar, D., Crawford, O., Valentine, A.P. & Trampert, J., 2018. Hamilton’s principle and normal mode coupling in an aspherical planet with a fluid core, *Geophys. J. Int.*, **214**(1), 485–507.
- Alterman, Z. & Karal, F., Jr, 1968. Propagation of elastic waves in layered media by finite difference methods, *Bull. seism. Soc. Am.*, **58**(1), 367–398.
- Auer, L., Boschi, L., Becker, T., Nissen-Meyer, T. & Giardini, D., 2014. Savani: a variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *J. geophys. Res.*, **119**(4), 3006–3034.
- Becker, T.W. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, **3**(1), doi:10.1029/2001GC000168.
- Bina, C.R., Stein, S., Marton, F.C. & Van Ark, E.M., 2001. Implications of slab mineralogy for subduction dynamics, *Phys. Earth planet. Inter.*, **127**(1–4), 51–66.
- Boyd, J.P., 2001. *Chebyshev and Fourier Spectral Methods*, Courier Corporation.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatsitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.*, **207**(3), 1739–1766.
- Capdeville, Y., Chaljub, E. & Montagner, J.P., 2003a. Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models, *Geophys. J. Int.*, **152**(1), 34–67.
- Capdeville, Y., To, A. & Romanowicz, B., 2003b. Coupling spectral elements and modes in a spherical Earth: an extension to the ‘sandwich’ case, *Geophys. J. Int.*, **154**(1), 44–57.
- Chaljub, E. *et al.*, 2015. 3-D numerical simulations of earthquake ground motion in sedimentary basins: testing accuracy through stringent models, *Geophys. J. Int.*, **201**(1), 90–111.
- Christensen, U.R. & Hofmann, A.W., 1994. Segregation of subducted oceanic crust in the convecting mantle, *J. geophys. Res.*, **99**(B10), 19 867–19 884.
- Clouzet, P., Masson, Y. & Romanowicz, B., 2018. Box tomography: first application to the imaging of upper-mantle shear velocity and radial anisotropy structure beneath the North American continent, *Geophys. J. Int.*, **213**(3), 1849–1875.

- Cottaar, S. & Romanowicz, B., 2012. An unusually large ULVZ at the base of the Mantle near Hawaii, *Earth planet. Sci. Lett.*, **355**, 213–222.
- Dahlen, F., 2005. Finite-frequency sensitivity kernels for boundary topography perturbations, *Geophys. J. Int.*, **162**(2), 525–540.
- Dahlen, F., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes – I. theory, *Geophys. J. Int.*, **141**(1), 157–174.
- Dalkolmo, J. & Friederich, W., 2000. Born scattering of long-period body waves, *Geophys. J. Int.*, **142**(3), 876–888.
- Defant, M.J. & Drummond, M.S., 1990. Derivation of some modern arc magmas by melting of young subducted lithosphere, *Nature*, **347**(6294), 662.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference earth model, *Phys. Earth planet. Inter.*, **25**(4), 297–356.
- Earle, P.S. & Shearer, P.M., 2001. Distribution of fine-scale mantle heterogeneity from observations of Pdiff coda, *Bull. seism. Soc. Am.*, **91**(6), 1875–1881.
- French, S.W. & Romanowicz, B., 2015. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, **525**(7567), 95.
- Frost, D.A., Garnero, E.J. & Rost, S., 2018. Dynamical links between small- and large-scale mantle heterogeneity: seismological evidence, *Earth planet. Sci. Lett.*, **482**, 135–146.
- Fukao, Y. & Obayashi, M., 2013. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity, *J. geophys. Res.*, **118**(11), 5920–5938.
- Garnero, E.J. & Helmberger, D.V., 1995. A very slow basal layer underlying large-scale low-velocity anomalies in the lower mantle beneath the Pacific: evidence from core phases, *Phys. Earth planet. Inter.*, **91**(1–3), 161–176.
- Garnero, E.J. & McNamara, A.K., 2008. Structure and dynamics of Earth's lower mantle, *Science*, **320**(5876), 626–628.
- Garnero, E.J., McNamara, A.K. & Shim, S.-H., 2016. Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nat. Geosci.*, **9**(7), 481.
- Gebraad, L., Boehm, C. & Fichtner, A., 2019. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, EarthArXiv, 10 Sept. 2019. Web, <https://doi.org/10.31223/osf.io/qftn5>.
- Gritto, R., Korneev, V.A. & Johnson, L.R., 1995. Low-frequency elastic-wave scattering by an inclusion: limits of applications, *Geophys. J. Int.*, **120**(3), 677–692.
- Gubernatis, J., 1979. Long-wave approximations for the scattering of elastic waves from flaws with applications to ellipsoidal voids and inclusions, *J. Appl. Phys.*, **50**(6), 4046–4058.
- Haugland, S.M., Ritsema, J., Kaneshima, S. & Thorne, M.S., 2017. Estimate of the rigidity of eclogite in the lower mantle from waveform modeling of broadband S-to-P wave conversions, *Geophys. Res. Lett.*
- Haugland, S.M., Ritsema, J., van Keken, P.E. & Nissen-Meyer, T., 2018. Analysis of PKP scattering using mantle mixing simulations and axisymmetric 3D waveforms, *Phys. Earth planet. Inter.*, **276**, 226–233.
- Hedlin, M.A., Shearer, P.M. & Earle, P.S., 1997. Seismic evidence for small-scale heterogeneity throughout the Earth's mantle, *Nature*, **387**(6629), 145.
- Helmberger, D., Ni, S., Wen, L. & Ritsema, J., 2000. Seismic evidence for ultralow-velocity zones beneath Africa and eastern Atlantic, *J. geophys. Res.*, **105**(B10), 23 865–23 878.
- Helmberger, D.V. & Ni, S., 2005. Approximate 3D body-wave synthetics for tomographic models, *Bull. seism. Soc. Am.*, **95**(1), 212–224.
- Hong, T. & Sen, M.K., 2009. A new MCMC algorithm for seismic waveform inversion and corresponding uncertainty analysis, *Geophys. J. Int.*, **177**(1), 14–32.
- Hudson, J. & Heritage, J., 1981. The use of the Born approximation in seismic scattering problems, *Geophys. J. Int.*, **66**(1), 221–240.
- Hwang, Y.K., Ritsema, J., van Keken, P.E., Goes, S. & Styles, E., 2011. Wavefront healing renders deep plumes seismically invisible, *Geophys. J. Int.*, **187**(1), 273–277.
- Igel, H., 2017. *Computational Seismology: A Practical Introduction*, Oxford Univ. Press.
- Igel, H. & Weber, M., 1995. SH-wave propagation in the whole mantle using high-order finite differences, *Geophys. Res. Lett.*, **22**(6), 731–734.
- Igel, H., Nissen-Meyer, T. & Jahnke, G., 2002. Wave propagation in 3D spherical sections: effects of subduction zones, *Phys. Earth planet. Inter.*, **132**(1–3), 219–234.
- Jahnke, G., Thorne, M.S., Cochard, A. & Igel, H., 2008. Global SH-wave propagation using a parallel axisymmetric spherical finite-difference scheme: application to whole mantle scattering, *Geophys. J. Int.*, **173**(3), 815–826.
- Kaneshima, S., 2016. Seismic scatterers in the mid-lower mantle, *Phys. Earth planet. Inter.*, **257**, 105–114.
- Kaneshima, S. & Helffrich, G., 1998. Detection of lower mantle scatterers northeast of the Marianna subduction zone using short-period array data, *J. geophys. Res.*, **103**(B3), 4825–4838.
- Karato, S.-I., 1997. On the separation of crustal component from subducted oceanic lithosphere near the 660 km discontinuity, *Phys. Earth planet. Inter.*, **99**(1–2), 103–111.
- Kellogg, L. & Turcotte, D., 1987. Homogenization of the mantle by convective mixing and diffusion, *Earth planet. Sci. Lett.*, **81**(4), 371–378.
- Kito, T. & Korenaga, J., 2010. Cross-correlation weighted migration: towards high-resolution mapping of mantle heterogeneities, *Geophys. J. Int.*, **181**(2), 1109–1127.
- Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation – I. validation, *Geophys. J. Int.*, **149**(2), 390–412.
- Korenaga, J., 2013. Teleseismic migration with dual bootstrap stack, *Geophys. J. Int.*, **196**(3), 1706–1723.
- Korenaga, J., 2015. Constraining the geometries of small-scale heterogeneities: a case study from the Mariana region, *J. geophys. Res.*, **120**(11), 7830–7851.
- Korneev, V.A. & Johnson, L.R., 1993. Scattering of elastic waves by a spherical inclusion – II. limitations of asymptotic solutions, *Geophys. J. Int.*, **115**(1), 251–263.
- Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. Obspy: a bridge for seismology into the scientific python ecosystem, *Comput. Sci. Discov.*, **8**(1), 014003.
- Kristeková, M., Kristek, J. & Moczo, P., 2009. Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophys. J. Int.*, **178**(2), 813–825.
- Laske, G., Masters, G., Ma, Z. & Pasquenos, M., 2013. Update on CRUST1.0 – a 1-degree global model of Earth's crust, in *Geophys. Res. Abstr.*, vol. 15, p. 2658, EGU General Assembly Vienna, Austria.
- Leng, K., Nissen-Meyer, T. & van Driel, M., 2016. Efficient global wave propagation adapted to 3-D structural complexity: a pseudospectral/spectral-element approach, *Geophys. J. Int.*, **207**(3), 1700–1721.
- Leng, K., Nissen-Meyer, T., Van Driel, M., Hosseini, K. & Al-Attar, D., 2019. AxiSEM3D: broad-band seismic wavefields in 3-D global Earth models with undulating discontinuities, *Geophys. J. Int.*, **217**(3), 2125–2146.
- Li, D., Helmberger, D., Clayton, R.W. & Sun, D., 2014. Global synthetic seismograms using a 2-D finite-difference method, *Geophys. J. Int.*, **197**(2), 1166–1183.
- Li, M., McNamara, A.K., Garnero, E.J. & Yu, S., 2017. Compositionally distinct ultra-low velocity zones on Earth's core-mantle boundary, *Nat. Commun.*, **8**(1), 177.
- Lin, C., Monteiller, V., Wang, K., Liu, T., Tong, P. & Liu, Q., 2019. High-frequency seismic wave modelling of the deep earth based on hybrid methods and spectral-element simulations: a conceptual study, *Geophys. J. Int.*, **219**(3), 1948–1969.
- Maguire, R., Ritsema, J., van Keken, P.E., Fichtner, A. & Goes, S., 2016. P-and S-wave delays caused by thermal plumes, *Geophys. J. Int.*, **206**(2), 1169–1178.
- Margerin, L. & Nolet, G., 2003. Multiple scattering of high-frequency seismic waves in the deep Earth: PKP precursor analysis and inversion for mantle granularity, *J. geophys. Res.*, **108**(B11), doi:10.1029/2003JB002455.
- Masson, Y. & Romanowicz, B., 2016. Fast computation of synthetic seismograms within a medium containing remote localized perturbations: a numerical solution to the scattering problem, *Geophys. J. Int.*, **208**(2), 674–692.

- Masson, Y. & Romanowicz, B., 2017. Box tomography: localized imaging of remote targets buried in an unknown medium, a step forward for understanding key structures in the deep earth, *Geophys. J. Int.*, **211**(1), 141–163.
- Masson, Y., Cupillard, P., Capdeville, Y. & Romanowicz, B., 2013. On the numerical implementation of time-reversal mirrors for tomographic imaging, *Geophys. J. Int.*, **196**(3), 1580–1599.
- Moczo, P., Kristek, J., Vavrycuk, V., Archuleta, R.J. & Halada, L., 2002. 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities, *Bull. seism. Soc. Am.*, **92**(8), 3042–3066.
- Monteiller, V., Chevrot, S., Komatitsch, D. & Fuji, N., 2012. A hybrid method to compute short-period synthetic seismograms of teleseismic body waves in a 3-D regional model, *Geophys. J. Int.*, **192**(1), 230–247.
- Monteiller, V., Chevrot, S., Komatitsch, D. & Wang, Y., 2015. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method, *Geophys. J. Int.*, **202**(2), 811–827.
- Montelli, R., Nolet, G., Dahlen, F. & Masters, G., 2006. A catalogue of deep mantle plumes: new results from finite-frequency tomography, *Geochem. Geophys. Geosyst.*, **7**(11), doi:10.1029/2006GC001248.
- Nissen-Meyer, T., Dahmen, F. & Fournier, A., 2007. Spherical-earth Fréchet sensitivity kernels, *Geophys. J. Int.*, **168**(3), 1051–1066.
- Nissen-Meyer, T., van Driel, M., Stähler, S.C., Hosseini, K., Hempel, S., Auer, L., Colombi, A. & Fournier, A., 2014. AxisSEM: broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth*, **5**(1), 425.
- Nolet, G., 2008. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*, Cambridge Univ. Press.
- Opršal, I., Matyska, C. & Irikura, K., 2009. The source-box wave propagation hybrid methods: general formulation and implementation, *Geophys. J. Int.*, **176**(2), 555–564.
- Peacock, S.M., Rushmer, T. & Thompson, A.B., 1994. Partial melting of subducting oceanic crust, *Earth planet. Sci. Lett.*, **121**(1–2), 227–244.
- Peter, D. et al., 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.*, **186**(2), 721–739.
- Pienkowska-Cote, M., Nissen-Meyer, T., Monteiller, V. & Komatitsch, D., 2018. Global high-frequency 3-D wavefields by wavefield injection and extrapolation, *Geophys. J. Int.*, **20**, 13873.
- Ringwood, A. & Irfune, T., 1988. Nature of the 650-km seismic discontinuity: implications for mantle dynamics and differentiation, *Nature*, **331**(6152), 131.
- Ritsema, J., Deuss, A.A., Van Heijst, H. & Woodhouse, J., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltimes and normal-mode splitting function measurements, *Geophys. J. Int.*, **184**(3), 1223–1236.
- Robertsson, J.O. & Chapman, C.H., 2000. An efficient method for calculating finite-difference seismograms after model alterations, *Geophysics*, **65**(3), 907–918.
- Rondenay, S., Cormier, V.F. & Van Ark, E.M., 2010. SKS and SPdKS sensitivity to two-dimensional ultralow-velocity zones, *J. geophys. Res.*, **115**(B4), doi:10.1029/2009JB006733.
- Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.*, **40**(3), 2–1.
- Shearer, P., 2015. Deep earth structure: seismic scattering in the deep Earth, in *Treatise on Geophysics*, pp. 759–787, Elsevier.
- Shearer, P.M. & Earle, P.S., 2004. The global short-period wavefield modelled with a Monte Carlo seismic phonon method, *Geophys. J. Int.*, **158**(3), 1103–1117.
- Sigloch, K., McQuarrie, N. & Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.*, **1**(7), 458.
- Silver, P.G., Carlson, R.W. & Olson, P., 1988. Deep slabs, geochemical heterogeneity, and the large-scale structure of mantle convection: investigation of an enduring paradox, *Ann. Rev. Earth Planet. Sci.*, **16**(1), 477–541.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science*, **325**(5943), 988–992.
- Tesoniero, A., Leng, K., Long, M. & Nissen-Meyer, T., 2020. Full wave sensitivity of sk (k) s phases to arbitrary anisotropy in the upper and lower mantle, *Geophys. J. Int.*, **222**(1), 412–435.
- Thorne, M.S., Garnero, E.J., Jahnke, G., Igel, H. & McNamara, A.K., 2013. Mega ultra low velocity zone and mantle flow, *Earth planet. Sci. Lett.*, **364**, 59–67.
- Tong, P., Chen, C.-w., Komatitsch, D., Basini, P. & Liu, Q., 2014a. High-resolution seismic array imaging based on an SEM–FK hybrid method, *Geophys. J. Int.*, **197**(1), 369–395.
- Tong, P., Komatitsch, D., Tseng, T.-L., Hung, S.-H., Chen, C.-W., Basini, P. & Liu, Q., 2014b. A 3-D spectral-element and frequency-wave number hybrid method for high-resolution seismic array imaging, *Geophys. Res. Lett.*, **41**(20), 7025–7034.
- van Driel, M., Krischer, L., Stähler, S., Hosseini, K. & Nissen-Meyer, T., 2015. Instaseis: instant global seismograms based on a broadband waveform database, *Solid Earth Discuss.*, **7**(1), 957–1005.
- van Keken, P.E., Hauri, E.H. & Ballentine, C.J., 2002. Mantle mixing: the generation, preservation, and destruction of chemical heterogeneity, *Ann. Rev. Earth planet. Sci.*, **30**(1), 493–525.
- van Manen, D.J., Robertsson, J.O. & Curtis, A., 2007. Exact wave field simulation for finite-volume scattering problems, *J. acoust. Soc. Am.*, **122**(4), EL115–EL121.
- Vanacore, E., Rost, S. & Thorne, M., 2016. Ultralow-velocity zone geometries resolved by multidimensional waveform modelling, *Geophys. J. Int.*, **206**(1), 659–674.
- Wen, L. & Helmberger, D.V., 1998. A two-dimensional P-SV hybrid method and its application to modeling localized structures near the core-mantle boundary, *J. geophys. Res.*, **103**(B8), 17 901–17 918.
- Williams, Q. & Garnero, E.J., 1996. Seismic evidence for partial melt at the base of Earth's mantle, *Science*, **273**(5281), 1528–1530.
- Wu, R.-S. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, **50**(4), 582–595.
- Ying, C. & Truell, R., 1956. Scattering of a plane longitudinal wave by a spherical obstacle in an isotropically elastic solid, *J. Appl. Phys.*, **27**(9), 1086–1097.
- Yu, S. & Garnero, E.J., 2018. Ultralow velocity zone locations: a global assessment, *Geochem. Geophys. Geosyst.*, **19**(2), 396–414.
- Zhao, D., 2004. Global tomographic images of mantle plumes and subducting slabs: insight into deep earth dynamics, *Phys. Earth planet. Inter.*, **146**(1–2), 3–34.

SUPPORTING INFORMATION

Supplementary data are available at [GJI](#) online.

ulvz3D.mpeg

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

APPENDIX A: A TRACTION-FREE SCHEME FOR SEM

In this appendix, we establish our traction-free scheme for the implementation of wavefield injection in any 2-D or 3-D SEM. For simplicity, we describe our scheme based on a 2-D spectral element mesh, as shown in Fig. A1. The scheme works exactly the same way in 3-D. Given an injection boundary Γ_{inj} composed of several element edges (or faces in 3-D), we can categorize the elements into the following three types:

- Exterior elements — those outside Γ_{inj} , filling Ω_F and determined by the background model.

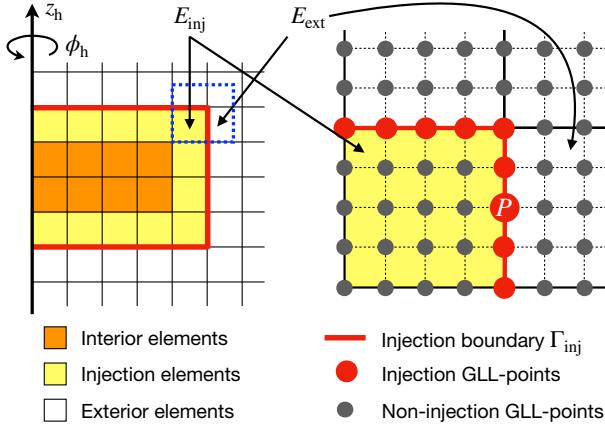


Figure A1. Implementation of wavefield injection in SEM. This illustrative 2-D mesh is designed for the solution of \mathbf{u}_{hy} in the second stage. It has a polynomial order of four, so each spectral element has $(4+1)^2$ GLL points. Given the injection boundary Γ_{inj} , we divide the spectral elements into three types: the exterior elements outside Γ_{inj} , the injection elements inside and adjacent to Γ_{inj} and the interior elements inside and non-adjacent to Γ_{inj} . A GLL point on Γ_{inj} is referred to as an injection GLL point. In AxiSEM3D, Γ_{inj} is normally an open curve with both its endpoints located on the axis, which generates a closed 3-D surface after being revolved around the axis; in this example, Γ_{inj} forms the meridian of a 3-D cylinder. The right-hand panel zooms into the dashed box on the left.

- (ii) Interior elements — those inside and non-adjacent to Γ_{inj} , filling an inner part of Ω_N ; arbitrary geometric and material perturbations can be imposed on these elements.
- (iii) Injection elements — those inside and adjacent to Γ_{inj} , filling an outer part of Ω_N that forms a ‘thickened’ or ‘smeared’ injection layer between Ω_N and Ω_F ; we prohibit the heterogeneity from extending to this layer.

Further, we refer to a GLL point on Γ_{inj} as an *injection GLL point*, which must be connected to by at least one injection and one exterior element.

Given such a mesh layout, one displacement-based scheme naturally emerges: instead of having both \mathbf{u}_0 and σ_0 recorded on Γ_{inj} , we may only record \mathbf{u}_0 on *all* the GLL points of the injection elements so that σ_0 on Γ_{inj} can be computed on-the-fly by the SEM. In short, we move the task of stress computation from the first stage to the second stage. The price is that the storage for recording \mathbf{u}_0 over the injection elements will be $3(p+1)/(3+6)$ times of that for recording both \mathbf{u}_0 and σ_0 on Γ_{inj} , where p is the polynomial order of the SEM (commonly $p=4$). Having resolved the difficulty of stress computation, this scheme makes a good start but remains insufficient, as the right-hand side of eq. (16) is still a surface integral, difficult to implement when Γ_{inj} has a complex geometry.

To convert the surface traction into a volume source, we resort to eq. (9b), which shows that the surface integral over Γ_{inj} equals to a volume integral over Ω_N , essentially a divergence theorem. Inserting eq. (9b) into (14), we obtain a modified weak form for \mathbf{u}_{hy} :

$$\int_{\Omega} (\rho \partial_t^2 \mathbf{u}_{\text{hy}} \cdot \mathbf{w} + \nabla \mathbf{u}_{\text{hy}} : \mathbf{C} : \nabla \mathbf{w}) \, d\mathbf{r}^3 = \int_{\Omega_N} (\rho_0 \partial_t^2 \mathbf{u}_0 \cdot \mathbf{w} + \nabla \mathbf{u}_0 : \mathbf{C}_0 : \nabla \mathbf{w}) \, d\mathbf{r}^3. \quad (\text{A1})$$

Consider an injection GLL point P connected to by an injection element E_{inj} and an exterior element E_{ext} , as shown in Fig. A1. The discretized form of eq. (A1) at point P can be written as (see chap.

7, Igel 2017, for the SEM discretization of a weak form)

$$\begin{aligned} & m_{\text{inj}}(P) \partial_t^2 \mathbf{u}(P) + \sum_{Q \in E_{\text{inj}}} \mathbf{K}_{\text{inj}}(P, Q) \cdot \mathbf{u}(Q) \\ & + m_{\text{ext}}(P) \partial_t^2 \delta \mathbf{u}(P) + \sum_{Q \in E_{\text{ext}}} \mathbf{K}_{\text{ext}}(P, Q) \cdot \delta \mathbf{u}(Q) \\ & = m_{\text{inj}}(P) \partial_t^2 \mathbf{u}_0(P) + \sum_{Q \in E_{\text{inj}}} \mathbf{K}_{\text{inj}}(P, Q) \cdot \mathbf{u}_0(Q), \end{aligned} \quad (\text{A2})$$

where $m_{\text{inj}}(P)$ and $m_{\text{ext}}(P)$ are the two partial masses of P contributed, respectively, by E_{inj} and E_{ext} , and $\mathbf{K}_{\text{inj}}(P, Q)$ a stiffness matrix of size 3×3 such that $\mathbf{K}_{\text{inj}}(P, Q) \cdot \mathbf{u}(Q)$ yields the contribution of E_{inj} to the stiffness force at P resulting from a displacement \mathbf{u} at GLL point Q , similarly for $\mathbf{K}_{\text{ext}}(P, Q)$. It is noted that, in eq. (A1), we have ρ and \mathbf{C} on the left-hand side and ρ_0 and \mathbf{C}_0 on the right-hand side, whereas we use the same m_{inj} and \mathbf{K}_{inj} on both sides of eq. (A2); this is because we do not allow any perturbation within an injection element. Besides, we have incorporated in eq. (A2) the definition of \mathbf{u}_{hy} by eq. (12), so the Dirichlet boundary condition in eq. (15) has been satisfied. No longer including a surface term (a term with \mathbf{n}), eq. (A2) still looks strenuous for implementation because two masses and two wavefields coexist at P . Using $\mathbf{u} = \mathbf{u}_0 + \delta \mathbf{u}$, we can merge the three mass terms (while keeping the three stiffness terms) to obtain

$$\begin{aligned} & m(P) \partial_t^2 \delta \mathbf{u}(P) + \sum_{Q \in E_{\text{inj}}} \mathbf{K}_{\text{inj}}(P, Q) \cdot \mathbf{u}(Q) \\ & + \sum_{Q \in E_{\text{ext}}} \mathbf{K}_{\text{ext}}(P, Q) \cdot \delta \mathbf{u}(Q) \\ & = \sum_{Q \in E_{\text{inj}}} \mathbf{K}_{\text{inj}}(P, Q) \cdot \mathbf{u}_0(Q). \end{aligned} \quad (\text{A3})$$

This equation reveals an extremely simple scheme.

By examining each stiffness term in eq. (A3), we now solve a modified hybrid wavefield that has a unique value at each GLL point:

$$\mathbf{u}_{\text{hy}}^* \stackrel{\text{def}}{=} \begin{cases} \delta \mathbf{u}, & \text{at a non-injection GLL point outside } \Gamma_{\text{inj}}; \\ \delta \mathbf{u}, & \text{at an injection GLL point;} \\ \mathbf{u}, & \text{at a non-injection GLL point inside } \Gamma_{\text{inj}}. \end{cases} \quad (\text{A4})$$

Compared to the originally defined \mathbf{u}_{hy} , the only difference is that \mathbf{u}_{hy}^* is forced to be $\delta \mathbf{u}$ on Γ_{inj} , as required by the only mass term in eq. (A3). From a computational viewpoint, \mathbf{u}_{hy}^* involves no discontinuity on Γ_{inj} , so the SEM architecture remains unchanged. Only two extra steps are required when computing the stiffness force of an injection element (such as E_{inj}): first, \mathbf{u}_0 must be added to \mathbf{u}_{hy}^* (or $\delta \mathbf{u}$) at its injection GLL points to reconstruct \mathbf{u}_{hy} (or \mathbf{u}) over E_{inj} , as required by the second term on the left-hand side of eq. (A3); secondly, we need to compute the stiffness force due to \mathbf{u}_0 (which has been recorded over E_{inj}) and, only at the injection GLL points, subtract it from the stiffness force due to \mathbf{u}_{hy} , as required by the right-hand side of eq. (A3). Nothing changes for the exterior or the interior elements. A complete workflow is shown in Fig. A2. Such a workflow is similar to the one proposed by Masson *et al.* (2013). The difference is that Masson *et al.* (2013) have suggested recording both the stiffness force and the displacement during the incident simulation whereas we only record the displacement and compute the stiffness force on the fly in the injection simulation. Unlike displacement, stiffness force is a discretization-dependent field, so a recorded stiffness force field can be injected only when the incident and the injection simulations use the same mesh.

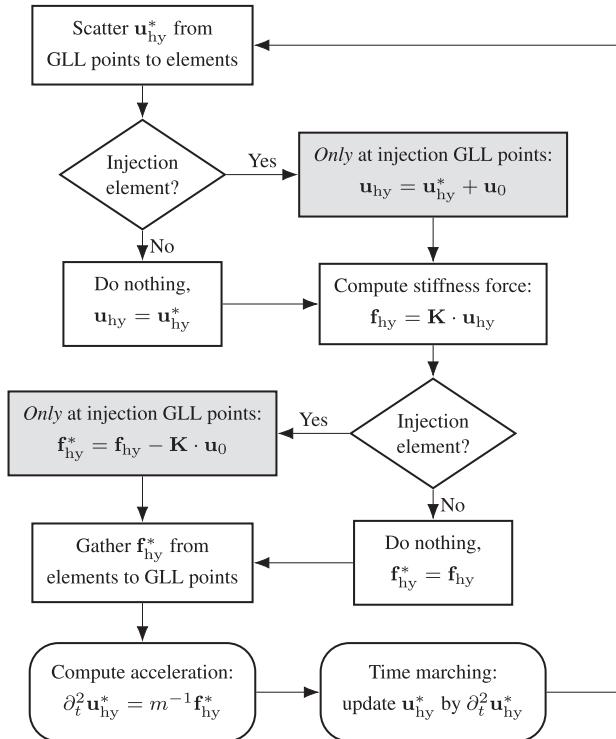


Figure A2. A traction-free scheme for implementing wavefield injection in SEM. The mesh setup is shown in Fig. A1. We assume that the incident wavefield \mathbf{u}_0 has been recorded over the injection elements (at both its injection and non-injection GLL points). The two shaded boxes indicate the two extra steps introduced by wavefield injection. The boxes with rounded corners indicate the operations on GLL points and the rest on elements.

In comparison with a traction- or stress-based scheme, our displacement-based one can benefit from improved flexibility and simplicity, as summarized below:

- (i) Requiring only \mathbf{u}_0 over the injection elements, it broadens the choice of the global solver or saves the effort of implementing stress calculation in the selected global solver.
- (ii) Agnostic of the geometry of Γ_{inj} , it allows for an arbitrary construction of the near-field subdomain; this can help enhancing model comprehensiveness and facilitating a compatible connection between the two cross-scale computational domains.
- (iii) The two extra steps introduced by wavefield injection are encapsulated in stiffness force calculation, minimizing the changes to the host SEM; in AxiSEM3D, for instance, these two steps are integrated into the ‘scattering’ and ‘gathering’ operations in Fig. A2, implemented by class inheritance with less than a dozen lines of extra code.

These advantages come at the expense of an expanded storage for recording \mathbf{u}_0 over the injection elements: with $p = 4$, for example, the storage will be increased by ~ 66 per cent in both 2-D and 3-D.

APPENDIX B: RECORDING OF THE INCIDENT WAVEFIELD

During an injection simulation with AxiSEM3D, the incident wavefield \mathbf{u}_0 must be injected as a Fourier series expanded under the heterogeneity-centred coordinate system (s_h, ϕ_h, z_h) , that is, $\mathbf{u}_0^\alpha(s_h, z_h), \alpha = 0, 1, \dots, n_u^{\text{het}}(s_h, z_h)$. However, these Fourier coefficients become undefined in the source-centred coordinate system (s_s, ϕ_s, z_s) under which \mathbf{u}_0 is computed and recorded. This issue can be easily solved by recording \mathbf{u}_0 at the ‘cardinal’ points associated with each GLL point over the injection elements.

Consider a generic GLL point located at (s, z) in D , which has a Fourier expansion order of n_u . This GLL point can be associated with $(2n_u + 1)$ cardinal points located at $(s, \phi^J, z), J = 0, 1, \dots, 2n_u$, where

$$\phi^J = \frac{2\pi J}{2n_u + 1}. \quad (\text{B1})$$

Given a generic 3-D function $f(s, \phi, z)$, it is straightforward to show that $f^\alpha(s, z), \alpha = 0, 1, \dots, n_u$ and $f(s, \phi^J, z), J = 0, 1, \dots, 2n_u$ are two equivalent azimuthal discretizations (chap. 5 and 10, Boyd 2001). The conversion between the two can be efficiently handled by a fast Fourier transform.

The following steps are required to prepare for an injection simulation with AxiSEM3D. Given a GLL point at which $\mathbf{u}_0^\alpha(s_h, z_h)$ is required, we first find its cardinal points (s_h, ϕ_h^J, z_h) and locate them in the source-centred coordinate system, that is, a coordinate transformation from (s_h, ϕ_h^J, z_h) to (s_s^J, ϕ_s^J, z_s^J) . Next, we compute and record $\mathbf{u}_0(s_s^J, \phi_s^J, z_s^J)$ by the incident simulation. Finally, we perform another coordinate transformation from $\mathbf{u}_0(s_s^J, \phi_s^J, z_s^J)$ back to $\mathbf{u}_0(s_h, \phi_h^J, z_h)$ and compute the required Fourier coefficients $\mathbf{u}_0^\alpha(s_h, z_h)$ by a fast Fourier transform. Compared with the two simulations, the computational cost for the processing of \mathbf{u}_0 is always negligible because the injection elements take up a tiny fraction of the computational domain.

The above process only requires recording the real-valued \mathbf{u}_0 at the given 3-D cardinal points (s_s^J, ϕ_s^J, z_s^J) , so the global solver can be any one capable of computing the displacement field in the background model. Using AxiSEM3D as the global solver, this process can be optimized to significantly reduce the storage of \mathbf{u}_0 . Instead of recording \mathbf{u}_0 at (s_s^J, ϕ_s^J, z_s^J) , we can directly record its Fourier coefficients in the source-centred coordinate system during the incident simulation, that is, $\mathbf{u}_0^\alpha(s_s, z_s), \alpha = 0, 1, \dots, n_u^*$, whereas $\mathbf{u}_0(s_s^J, \phi_s^J, z_s^J)$ can be computed on the fly in the injection simulation. Because \mathbf{u}_0 is quasi-axisymmetric in the source-centred coordinate system, the Fourier expansion order n_u^* will be small; for example, $n_u^* = 2$ for an axisymmetric background model, reducing the wavefield storage from 3-D to 2-D. Such a strategy is similar to Instaseis (van Driel *et al.* 2015), the difference being that we compute a database covering the heterogeneity instead of on the surface. For example, we have created a database for the study of ULVZs, which covers the lowermost 30 km of the entire mantle (so that the 3-D ULVZ can be placed at any location on the CMB and can have an arbitrary lateral shape) and a record length of 2400 s (with a sampling frequency of 20 per period). This database takes up a 2 TB storage at 1 Hz, which can be reduced to 250 GB at 0.5 Hz.