Optimized viscoelastic wave propagation for weakly dissipative media

Martin van Driel[1] and Tarje Nissen-Meyer[2]

[1]*Institute of Geophysics, ETH Zurich, Sonneggstrasse 5, 8092 Zurich, Switzerland. E-mail: Martin@vanDriel.de*
[2]*Department of Earth Sciences, University of Oxford, South Parks Road, Oxford, OX1 3AN, United Kingdom*

SUMMARY

The representation of viscoelastic media in the time domain becomes more challenging with greater bandwidth of the propagating waves and number of travelled wavelengths. With the continuously increasing computational power, more extreme parameter regimes become accessible, which requires the reassessment and improvement of the standard 'memory variable' methods to implement attenuation in time-domain seismic wave-propagation methods. In this paper, we propose a method to minimize the error in the wavefield for a fixed complexity of the anelastic medium. This method consists of defining an appropriate misfit criterion based on a first-order analysis of how errors in the discretized medium propagate into errors in the wavefield and a simulated annealing optimization scheme to find the globally optimal parametrization. Furthermore, we derive an analytical time-stepping scheme for the memory variables that encode the strain history of the medium. Then we develop the coarse grained memory variable approach for the spectral element method (SEM) and benchmark it using the 2.5-D code *AxiSEM* for global body waves up to 1 Hz. Showing very good agreement with a reference solution, it also leads to a speedup of a factor of 5 in the anelastic part of the code (factor 2 in total) in this 2.5-D approach. A factor of $\approx$15 (3 in total) can be expected for the 3-D case compared to conventional implementations.

Key words: Seismic attenuation; Computational seismology; Theoretical seismology; Wave propagation.

1 INTRODUCTION

Ongoing advances in supercomputer architecture and numerical methods enable the solution of the wave equation in increasingly extreme parameter regimes, such as higher frequency waves in larger modelling domains, having longer propagation distances in terms of the number of travelled wavelengths. Numerical errors as well as errors in physical approximations accumulate over these larger distances and require more precision both in the numerical solution and in the physical medium representation. We address the question of how accurate attenuation and the corresponding physical dispersion needs to be represented to accurately model seismic waves including attenuation with a focus on global body waves, although the treatment is more general and scale invariant. Even though the quality factor Q is itself poorly constrained by existing seismic studies, especially its 3-D structure and frequency dependence (Romanowicz & Mitchell 2007, and references therein), we consider it an essential pre-requisite to carefully evaluate the accuracy of the numerous numerical and physical approximations, before upscaling conventional methods of implementing attenuation to the more extreme regimes that modern computational seismology approaches.

While moving from purely elastic to viscoelastic media is easy in the frequency domain via the correspondence principle and introduction of complex-valued media properties, it is more involved in the time domain since the multiplication in the constitutive relation of the elastic medium needs to be replaced by a convolution. A first method to transform the stress–strain relation into a differential form using Padé approximations is introduced by Day & Minster (1984). Later, Emmerich & Korn (1987) and Carcione *et al.* (1988) suggested to improve this by approximating the medium properties with a discrete relaxation spectrum (Liu *et al.* 1976) and fitting the parameters used to describe this spectrum to the observed behaviour numerically. These methods are still common today (e.g. Komatitsch & Tromp 2002b; Graves & Day 2003; Kristek & Moczo 2003; Käser *et al.* 2007; Fichtner *et al.* 2009; Savage *et al.* 2010), a more complete summary and historical overview is given by Carcione (2007) and Moczo *et al.* (2014). In this study, we suggest several improvements to this scheme leading to better accuracy in the medium representation at zero extra cost.

Even if Q is large [on the global scale the minimum observed value is about 50, Gung & Romanowicz (2004)] and the effect of attenuation on the seismograms small, accounting for it typically leads to an increase of the computational costs by a factor of two to

four in both computation time and memory (e.g. Blanch *et al.* 1995; Käser *et al.* 2007). Day (1998) suggested a method called the coarse-grained memory variable approach to redistribute the medium properties on the subwavelength scale such that it behaves the same for the wavefield, but is computationally significantly less expensive. Day's method was originally developed for the acoustic wave equation and regular grid finite-difference schemes. It is generalized to the viscoelastic wave equation and improved by more appropriate averaging schemes by Day & Bradley (2001) and Graves & Day (2003). Kristek & Moczo (2003) further improved the scheme by introducing material-independent memory variables to avoid artificial averaging of material parameters at gridpoints where interpolation of the memory variables is necessary (e.g. in the context of heterogeneous finite-differences methods at material discontinuities or for thin layers). Ma & Liu (2006) apply the coarse-grained method to a low-order finite-element scheme on unstructured grids. However, their approach cannot directly be translated to higher-order methods such as the spectral element method (SEM), because in such schemes the elements are too large compared to the wavelength (less than four elements per wavelength). We propose to redistribute the medium properties within the elements on the high-order basis functions, which again is small scale compared to the wavelength.

On the global scale, high-frequency body waves are routinely observed at propagation distances of 1000–2000 wavelengths and more, a regime that is hardly accessible with current global 3-D solvers and computers (Carrington *et al.* 2008). While the methods we propose for parameter optimization are completely general and the coarse-grained memory variable approach applicable to all high-order finite-element methods, we use the axisymmetric SEM *AxiSEM* introduced by Nissen-Meyer *et al.* (2007a,b, 2008), further developed to include anisotropy by van Driel & Nissen-Meyer (2014) and published open source by Nissen-Meyer *et al.* (2014) as an example implementation to test our theoretical arguments. The efficiency of this 2.5-D approach allows very high-frequency simulations (currently up to 2 Hz on the global scale) that are still impossible to reach using full 3-D methods, so it provides a good basis to test common physical and numerical approximations.

Other axisymmetric approaches to global and local wave propagation have been presented (Alterman & Karal 1968; Igel & Weber 1995, 1996; Chaljub & Tarantola 1997; Furumura *et al.* 1998; Thomas *et al.* 2000; Takenaka *et al.* 2003; Toyokuni *et al.* 2005; Jahnke *et al.* 2008), but only recently Toyokuni & Takenaka (2006, 2012) generalized their method to include moment tensor sources, attenuation and the Earth's centre. These methods are all based on isotropic media and especially the finite-difference methods among them have to deal with large dispersion errors for interface-sensitive waves such as surface waves and diffracted waves (Igel & Weber 1995; Igel *et al.* 1995). We generalize *AxiSEM* to viscoelastic anisotropic axisymmetric media to overcome these issues. This enables the simulation of high-frequency body waves with a particularly high sensitivity to attenuation, travelling distances over thousands of wavelengths such as transmitted, reflected and diffracted core phases.

2 THEORY

Here, we introduce the concepts and our notation for time-domain modelling of the memory variable approach to viscoelastic dissipation. We start with the simple 1-D case (scalar instead of tensorial quantities) and generalize to the full 3-D problem subsequently.

2.1 Preliminaries

The most general linear stress–strain relation is the convolution, that is, the strain ϵ from all times can linearly influence the stress σ at time t:

$$\sigma(t) = \int_{-\infty}^{\infty} M(t-\tau) \cdot \epsilon(\tau) \, \mathrm{d}\tau$$

$$= \int_{-\infty}^{\infty} R(t-\tau) \cdot \dot{\epsilon}(\tau) \, \mathrm{d}\tau. \tag{1}$$

Here, $R(t)$ is the stress relaxation function with the modulus $M(t) = \dot{R}(t)$, that is, the stress response to a unit step in the strain. Assuming causality [strain in the future cannot affect the current stress state: $M(t) = 0$, $t < 0$], fading memory [more recent strain has a larger impact on the current state: $M(t)$ is monotonous and converges to 0 for $t \to \infty$] and solid behaviour in the limit of low frequencies [a constant strain causes a constant nonzero stress: $\lim_{t \to \infty} R(t) = M_R > 0$], the time-dependent modulus takes the general form shown in Fig. 1: zero for negative time and decaying to a constant positive value for positive times (Christensen 1982). This can be approximated with a discrete relaxation spectrum (Liu *et al.* 1976)

$$R(t) = \left[M_R + \delta M \sum_{j=1}^{N} a_j e^{-\omega_j t} \right] H(t), \tag{2}$$

with $N > 0$ single peaks of strength $a_j > 0$, $\sum_j a_j = 1$ located at the relaxation frequencies $\omega_j > 0$.

This frequency dependence can be interpreted using mechanical models with combinations of springs and dash-pots such as the generalized Maxwell or Zener bodies. As shown by Moczo & Kristek (2005), these two interpretations lead to different parametrization of the medium, but result in the same mechanical behaviour. The relation between the parameter sets in the Maxwell (a_j and ω_j) and Zener body ($\tau_{\epsilon j}$ and $\tau_{\sigma j}$, the strain and stress relaxation times) representation can be found by comparing eq. (8) in Blanch *et al.* (1995) and eq. (11) in Emmerich & Korn (1987):

$$a_j \frac{\delta M}{M_R} = \frac{\tau_{\epsilon j}}{\tau_{\sigma j}} - 1, \quad \omega_j = \frac{1}{\tau_{\sigma j}}. \tag{3}$$

In this paper, we will use the Maxwell body notation as introduced by Emmerich & Korn (1987), but the results can be directly applied to the Zener body formulation using the above relations. Using the
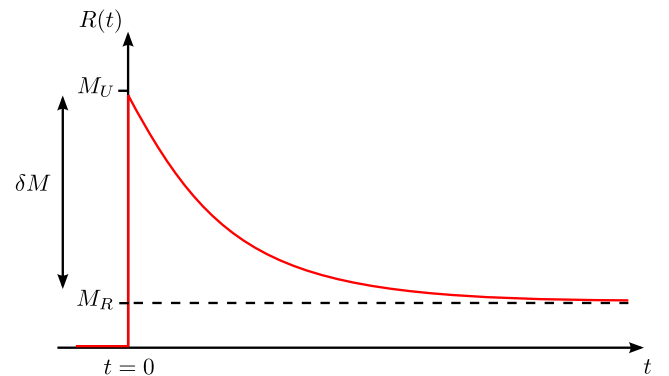

Figure 1. Stress relaxation function R for a viscoelastic medium in the time domain, that is, the stress response to a unit step in strain. M_U and M_R denote unrelaxed and relaxed modulus, respectively. Adapted from Emmerich & Korn (1987) and Christensen (1982).

discrete relaxation spectrum, the resulting stress–strain relation can be written as

$$\sigma(t) = M_U \epsilon(t) - \sum_{j=1}^{N} \zeta^j(t), \tag{4}$$

where influence of the strain history on the current state of the material is encoded in the 'memory variables' ζ^j. The memory variables obey the N differential 'memory variable equations'

$$\dot{\zeta}^j(t) + \omega_j \zeta^j(t) = a_j \omega_j \delta M \epsilon(t) \tag{5}$$

that are driven by the strain of the medium $\epsilon(t)$. The resulting frequency-dependent modulus and quality factor Q are (Emmerich & Korn 1987):

$$M(\omega) = M_R + \sum_j a_j \delta M \frac{i\omega}{i\omega + \omega_j}, \tag{6}$$

$$Q^{-1}(\omega) = \frac{\mathrm{Im}M}{\mathrm{Re}M} = \frac{\delta M}{M_R} \frac{\sum_j a_j \frac{\omega/\omega_j}{1+(\omega/\omega_j)^2}}{1 + \frac{\delta M}{M_R} \sum_j a_j \frac{(\omega/\omega_j)^2}{1+(\omega/\omega_j)^2}}. \tag{7}$$

Arbitrary frequency dependency of Q can hence be approximated by a sum of absorption bands (see Fig. 2), by tuning the $2N$ parameters of the discrete relaxation spectrum a_j and ω_j. This is one non-linear optimization problem for each set of Q and M in the model. It is important to note that this optimization is subject to the additional non-linear constraint $a_j > 0$ [equivalent $\tau_\epsilon > \tau_\sigma$ for the Zener model, compare Carcione (2007), eqs. 2.169 and 2.193].

An important consequence of causality are the Kramers–Kronig relations stating that the real and imaginary part of $M(\omega)$ are related by Hilbert transforms. The medium is therefore fully described by either the modulus or the phase velocity at a reference frequency
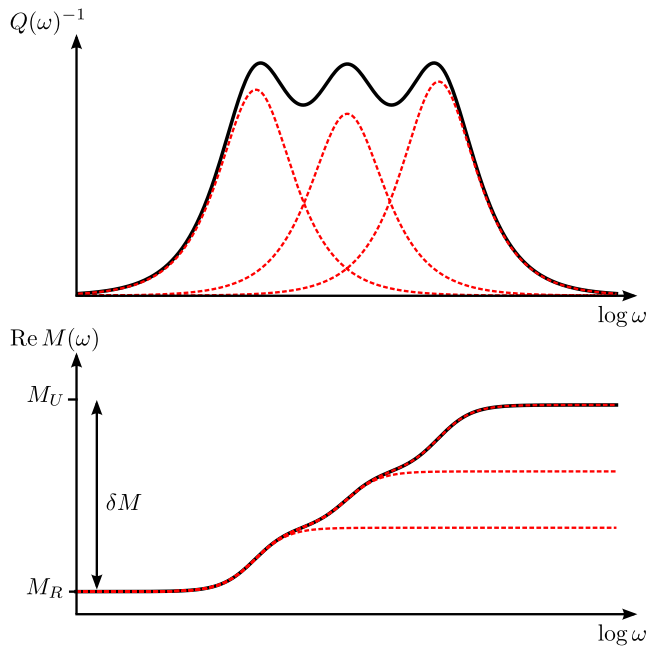

Figure 2. Inverse quality factor Q^{-1} and real part of the modulus M for a medium with three standard linear solids (see eqs 6 and 7). Dashed red lines indicate contributions from the individual linear solids, black solid lines the sum. Arbitrary frequency dependency (here: constant Q/logarithmic M in a limited frequency range) can be approximated by a sum of absorption bands. In the limit of large Q these take the form of Debye functions.

ω_r, and the quality factor in the frequency range of interest $Q(\omega)$, which in practice is often assumed to be constant or obey a power law.

Given the modulus or phase velocity at a reference frequency ω_r inside the frequency range where Q is optimized, δM, M_R and M_U can be found by evaluating eq. (6) at this frequency. As δM is unknown before the optimization, a change of variables is necessary defining new coefficients $y_j = \frac{\delta M}{M_R} a_j$. Then, with the y_j found by optimization, we can find

$$\gamma := \frac{1}{\sum_j y_j} \sum_j y_j \frac{\omega_j^2}{\omega_r^2 + \omega_j^2}, \tag{8}$$

$$\delta M = M(\omega_r) \Big/ \left(\frac{1}{\sum_j y_j} + 1 - \gamma \right), \tag{9}$$

$$M_U = M(\omega_r) + \gamma \delta M, \tag{10}$$

$$M_R = M(\omega_r) - (1 - \gamma)\delta M. \tag{11}$$

A basic question is then what criterion to use in order to find the parametrization of the medium for a numerical solution to the wave equation. We strive to optimize this procedure by means of the maximal error tolerance in the wavefield as defined by amplitude and phase error estimates in the next section.

2.2 Optimal Q parametrization

In this section, we analyse the influence of deviations in $M(\omega)$ and $Q(\omega)$ on the wavefield. As discussed, the optimization problem of finding the best set of $2N$ parameters is inherently non-linear (also for large Q due to the free choice of relaxation frequencies) and has a non-linear constraint. Taking this as given, choosing a non-linear optimization approach such as simulated annealing also enables the free choice of optimization criteria. The goal of this section is to define such a criterion for finding the material parametrization for the discrete relaxation spectrum that minimizes the error in the wavefield.

We analyse the performance of different medium parametrizations using the dissipation operator (Müller 1983) in the approximation for large Q as suggested by Emmerich & Korn (1987):

$$D(\omega) = \exp\left[i\omega T_r \left(1 - \sqrt{\frac{|M_r|}{M(\omega)}} \right) \right]. \tag{12}$$

Here T_r and $M_r = M(\omega_r)$ denote the traveltime and the modulus at the reference frequency ω_r. The anelastic response is then found by convolving the elastic response computed using the medium properties at the reference frequency with this dissipation operator. To evaluate the influence of errors in the representation of the medium, we separate amplitude and phase effects by writing this operator as

$$D(\omega) = A(\omega)e^{i\varphi(\omega)}. \tag{13}$$

In the case of large Q, these can be approximated with $M = M_1 + iM_2 = M_1(1 + i/Q)$ as:

$$A(\omega) \approx \exp\left(-\frac{1}{2} \frac{\omega T_r}{Q(\omega)} \sqrt{\frac{|M_r|}{M_1(\omega)}} \right) \tag{14}$$

for the amplitude and

$$\varphi(\omega) \approx \omega T_r \left(1 - \sqrt{\frac{|M_r|}{M_1(\omega)}} \right) \tag{15}$$

for the phase. The relative error in Q is typically an order of magnitude larger than the relative error in M_1. This can be seen from eqs (6) and (7): while Q^{-1} depends in first order on the frequency, M is a constant plus a small frequency-dependent term. Then, the first-order effect of amplitude and phase errors in the medium parametrization can be found as

$$\frac{\Delta A}{A} = \frac{1}{4\pi} \frac{\Delta Q}{Q} \frac{n_\lambda}{Q} \qquad (16)$$

and

$$\Delta\varphi = -\frac{1}{4\pi} \frac{\Delta M_1}{M_1} \cdot n_\lambda, \qquad (17)$$

where n_λ denotes number of propagated wavelengths and $\frac{\Delta Q}{Q}$ and $\frac{\Delta M_1}{M_1}$ are the relative errors in quality factor and real part of the modulus. As expected intuitively, the phase error is determined by the error in the real part of the modulus, which for large Q dominates the phase velocity, while the error in amplitude is determined by the error in Q.

Fig. 3 shows how the acceptable errors in the real part of the modulus and Q can be determined based on the phase and amplitude errors that are acceptable in a given application. For global high-frequency body waves ($t^* = T_r/Q = 4\,s$, $n_\lambda = 1000$) and an acceptable error of a few per cent in phase and amplitude, we con-

clude that Q should be approximated to well below 1 per cent and the modulus below 0.03 per cent. In contrast, the same seismic phase observed at lower frequencies, that is, smaller $n_\lambda = 100$ as typically used in full 3-D global simulations, is well represented with errors in Q and modulus that are 10 times larger.

2.2.1 Optimization criteria and variables

It is common practice (Emmerich & Korn 1987; Blanch *et al.* 1995; Komatitsch & Tromp 2002b; Graves & Day 2003; Kristek & Moczo 2003; Käser *et al.* 2007; Savage *et al.* 2010) to find the medium parametrization a_j, ω_j by choosing the relaxation frequencies ω_j *a priori*, mostly logarithmically spaced in the frequency range of interest. Then, the a_j can be found by sampling $Q(\omega)$ at a finite number of frequencies ω_k and solving an overdetermined inverse problem. We propose the following three-fold strategy to improve the parametrization: (1) a specific choice of the norm for the aforementioned inverse problem, (2) invert for ω_j instead of setting them *a priori* and (3) fit the medium more accurately at the higher frequencies.

As can be seen from eq. (14), variations in Q affect seismogram amplitudes exponentially. The logarithmic error in the amplitude is therefore of the same size when Q is multiplied or divided by a constant. Also, Q is always positive, which motivates usage of the logarithmic error for Q instead of the plain l2-norm. While in the limit of small deviations these two norms are asymptotically identical, the log-l2 norm emphasizes negative deviations from the optimal Q that are larger than a few per cent. Eqs (16) and (17) also show that high-frequency waves that have travelled more wavelengths are more sensitive to errors in the parametrization with the factor n_λ. This motivates a linear weighting of the frequencies and we minimize

$$\epsilon^2 = \int \left(\frac{\ln(Q_{\text{target}})}{\ln(Q_{\text{ls}})} \right)^2 \cdot w(\omega)\, d\ln\omega, \qquad (18)$$

where Q_{target} and Q_{ls} are the exact quality factor and its approximation by the linear solid from eq. (7), respectively. The weights $w(\omega)$ are set to 1 by most authors (compare citations in the first paragraph of this section), the linear frequency weighting we suggest is $w(\omega) = \omega$. For the error in the real part of the modulus we use the standard l2-norm, as the relative errors are very small anyway.

Fig. 4 visualizes the importance of including the relaxation frequencies in the optimization: for log-spaced fixed frequencies, the discretization of the medium does not converge towards the analytical behaviour, neither for increased N nor for smaller bandwidth for which the medium is optimized. This is in contrast to the finding by Savage *et al.* (2010), Fig. 2, whose results we are only able to reproduce when ignoring the constraint $a_j > 0$. For applications where fitting of Q better than 1 per cent is needed (this is low Q or many wavelength propagation as for global high-frequency body waves, compare Fig. 3), inversion for the relaxation frequencies is inevitable. This non-linear optimization problem with $2N$ parameters can effectively be solved by a simulated annealing approach within seconds using 10^5–10^7 iterations for $N \leq 6$, this was also suggested by Liu & Archuleta (2006).

Fig. 5 shows three examples of seismograms calculated with different medium representations that have the same numerical complexity in time-domain wave-propagation solvers due to the same number of memory variables. Both the reference solution and the approximations are calculated using the dissipation operator, the difference is only in $M(\omega)$. The first column represents the standard method of choosing the relaxation frequencies of the absorption
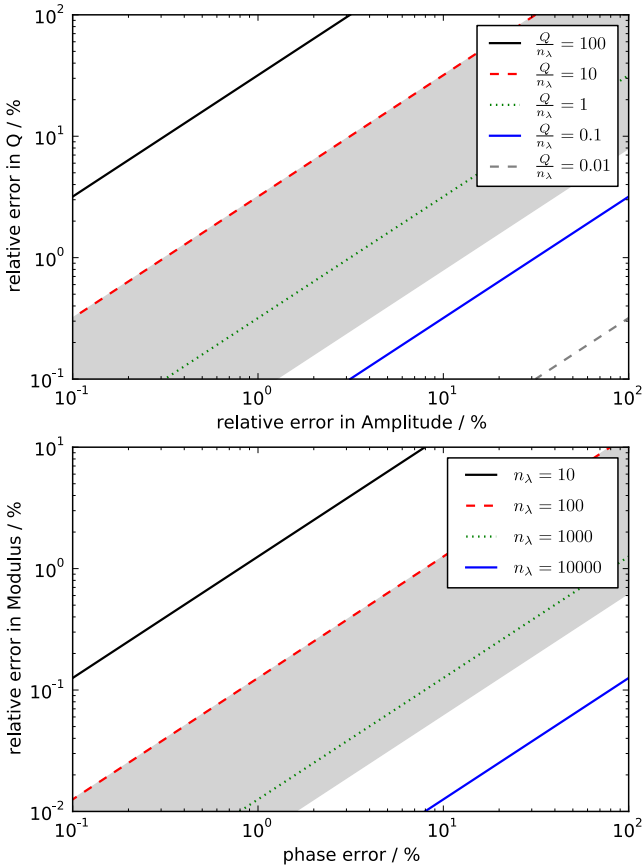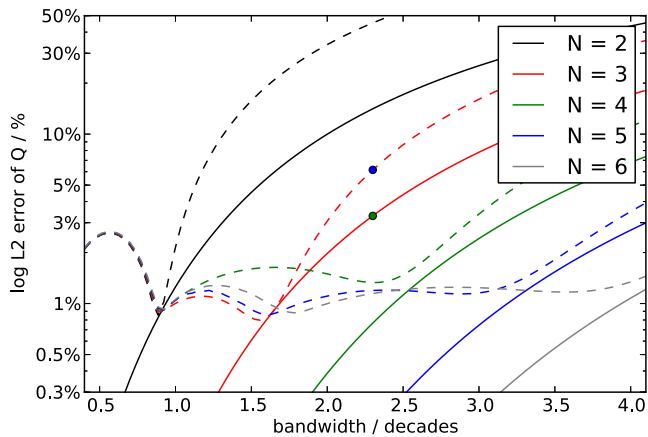

Figure 3. Given acceptable amplitude (top panel) or phase (bottom panel) error, what are the requirements for the parametrization of the anelastic medium in terms of acceptable error in quality factor Q and modulus M for a range of number of travelled wavelengths n_λ. Shaded areas indicate typical global body waves, that is, $t^* = 1$–$4\,s$, period 1–10 s, traveltime 1000–2000 s. Note the different scales on the y-axis.

Figure 4. Importance of the choice of relaxation frequencies for a good approximation of constant Q for various numbers of absorption bands N as a function of bandwidth: frequencies fixed log-spaced (dashed lines) versus inversion using simulated annealing (solid line). Note that for fixed relaxation frequencies Q does not converge towards the optimal value, neither for larger N, nor for smaller bandwidth, and the minimal error is about 1 per cent. The dots correspond to the first two examples in Fig. 5.

bands log-spaced. Inverting for the frequencies can reduce both the misfit in Q and consequently phase and amplitude errors of the seismograms in all frequency bands by factors of 2–4 (second column). Additional frequency weighting reduces the phase and envelope misfit by another factor of 2 in the highest frequency band at the cost of worse fit in the lower frequency range (third column), resulting in same order of magnitude misfits in all frequency bands.

2.2.2 The modulus

From the examples in Fig. 5, it can be observed that even when fitting $Q(\omega)$ with high accuracy for the high frequencies, the modulus has a maximum in the relative error close to the bounds of the frequency band. This can be understood from a relation between Q and the modulus that is valid in the approximation of large and almost constant Q (Dahlen & Tromp 1998, eq. 6.75):

$$\frac{\partial \ln M_1(\omega)}{\partial \ln \omega} \approx \frac{2}{\pi Q(\omega)}. \tag{19}$$

Thus, the modulus takes a maximum in its derivative at the peaks of $Q(\omega)$. Another disadvantage of optimizing Q only is that the choice of the reference frequency where the modulus is known has an important effect: if it coincides with one of the maxima of the approximation of the modulus, the average value of M_1 and hence the phase velocity are skewed. This can be avoided by relaxing the requirement that the approximate modulus matches the reference exactly at the reference frequency. Instead the ratio $M_R/M_1(\omega_r)$ can be added as a parameter to the inverse problem. $M(\omega)$ can be found via the Kramers–Kronig relations, either by numerical evaluation or analytical, for example, for constant Q (Kjartansson 1979)

$$M(\omega) = M(\omega_r) \left(\frac{\mathrm{i}\omega}{\omega_r} \right)^{\frac{2}{\pi} \tan^{-1} Q^{-1}}. \tag{20}$$

The fit of M_1 can then be added to the inverse problem with a weighting relative to the fit of Q. Fig. 6 shows an example, where the reference frequency was deliberately chosen inconvenient

($f_r = 2$ Hz) at the upper limit of the frequency range. If the modulus is determined by setting it to the reference value at the reference frequency, its average value is too low, leading to large phase errors. Adding the real part of the modulus to the optimization criterion circumvents this problem and makes the result independent of the choice of the reference frequency, while keeping essentially the same $Q(\omega)$ (green). Additionally, the parametrization can be tuned towards lower amplitude or phase errors by the weighting between Q and M_1 in the optimization criterion (red).

2.3 Determination of the optimal structural parameters in full-scale applications

In full seismic applications, the material parameters a_j and ω_j need to be determined for many values of Q and seismic velocities, typically for each spatial gridpoint, that is, $10^6 - 10^9$ times. The performance of an algorithm to find them is thus important and several approaches have been suggested.

For large Q values, eq. (7) can be linearized by realizing that $\delta M/M_R$ becomes small:

$$Q^{-1}(\omega) \approx \frac{\delta M}{M_R} \sum_j a_j \frac{\omega/\omega_j}{1 + (\omega/\omega_j)^2}. \tag{21}$$

This was proposed by Emmerich & Korn (1987) and later became more popular as the 'τ-method' by Blanch *et al.* (1995). The advantage of this method is that it results in only one inverse problem while finding the parameters for different Q is trivial due to the linearity of a_j in Q^{-1}. For low values of Q, this approximation introduces a logarithmic error to the resulting approximation of constant Q, see blue curve in Fig. 7(a). The largest error shows up at the high-frequency end of the frequency range of interest, that is, where the wavefield is most sensitive to such errors, see earlier. Furthermore, this problem becomes more dominant with increasing bandwidth.

Often, the relaxation frequencies ω_j are not inverted for, but fixed log-spaced, since this allows for a linear (hence faster) inversion for the parameters a_j as suggested by Emmerich & Korn (1987) by rearranging terms in eq. (7).

Similarly, Savage *et al.* (2010) use log-spaced frequencies, but a simplex approach for the non-linear inversion and suggest to use a look-up table to solve fewer inverse problems. However, these methods cannot handle the constraint $a_j > 0$.

Liu & Archuleta (2006) present empirical formulae to interpolate the parameters between the largest and smallest value of Q in the model and use a simulated annealing approach to solve the remaining two inverse problems. As they do not present a closed form to find the coefficients in their interpolation, their formulae can only be used for their specific setup of frequency range, number of absorption bands N and Q range.

Here, we suggest a correction to the linearization as an easy-to-implement and computationally light method, combining the advantages of the linearization ('τ-method') with increased accuracy for low Q. As this results in a single inverse problem, the relatively expensive simulated annealing method can be used allowing to invert for the relaxation frequencies as well as including the constraint $a_j > 0$.

This correction can be found by realizing that the discrete relaxation frequencies are widely spaced, hence

$$\frac{\omega_k/\omega_j}{1 + (\omega_k/\omega_j)^2} \approx \frac{1}{2} \delta_{jk} \tag{22}$$
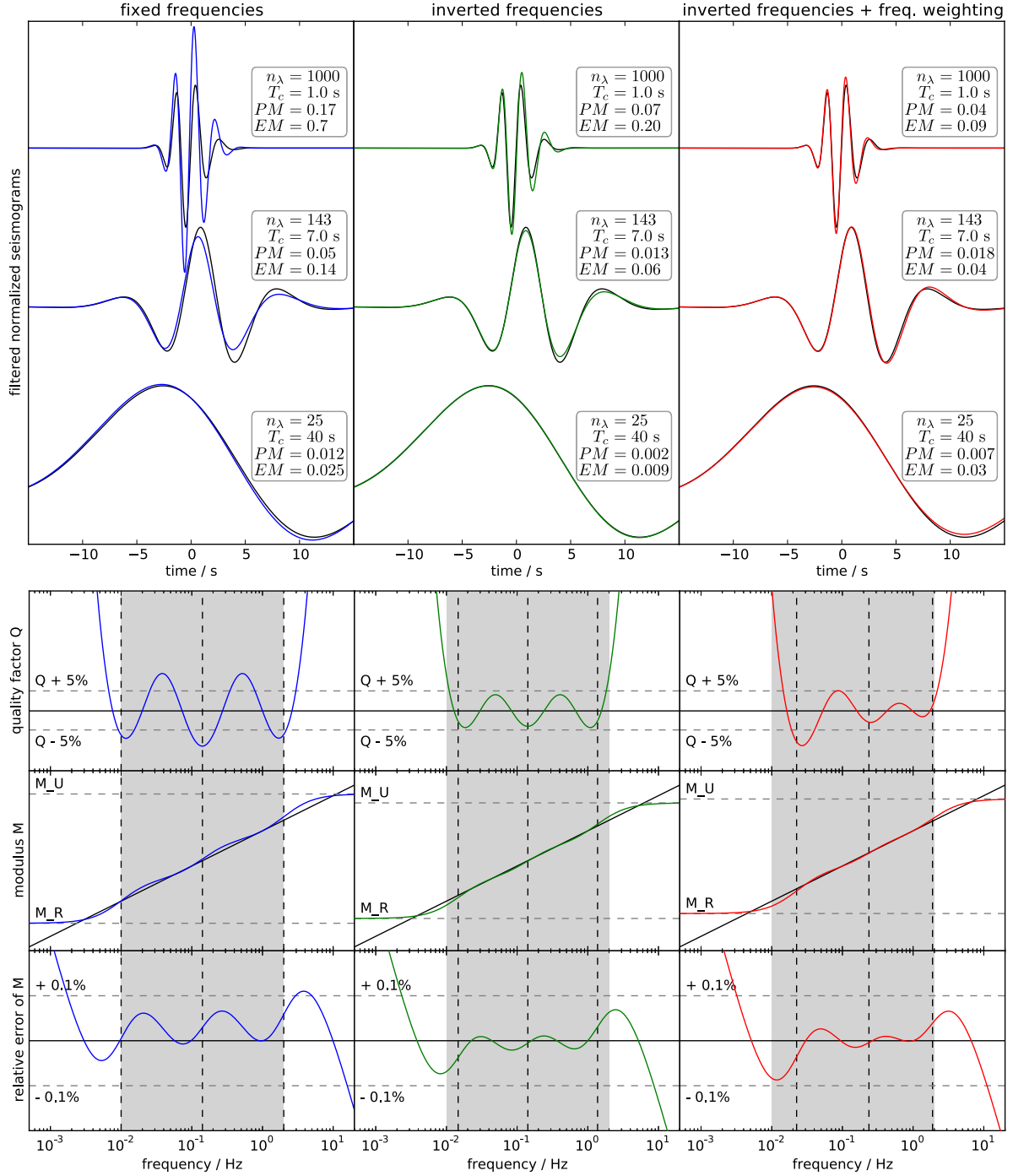
Figure 5. Influence of the parametrization of the medium on the accuracy of the seismograms simulated with dissipation operators for $t^* = 4$ s [typical for teleseismic *S* waves, see for example, Nolet (2008)] while keeping the complexity of the parametrization constant (three absorption bands). Left to right: relaxation frequencies fixed log-spaced in the frequency band of interest (2.3 decades), optimized frequencies using simulated annealing, additional linear weighting with the frequency. Top to bottom: seismograms filtered with Gabor-bandpass filters with centre period T_c, quality factor Q, modulus M and relative error of the modulus as approximated with standard linear solids. *EM* and *PM* denote the envelope and phase misfit (Kristeková *et al.* 2009). Shaded region indicates the frequency range used in the parameter optimization, vertical dashed lines the relaxation frequencies ω_j and solid black lines the reference medium. Optimization criterion was the log-l2 error of *Q*, see text for details. Note the large deviations of the optimal value in the modulus at the boundaries of the frequency range.

and

$$\frac{\left(\omega_k/\omega_j\right)^2}{1+\left(\omega_k/\omega_j\right)^2} \approx \begin{cases} 0, & (\omega_k < \omega_j) \\ \frac{1}{2}, & (\omega_k = \omega_j) \\ 1, & (\omega_k > \omega_j). \end{cases} \quad (23)$$

Then, we insert these approximations into the exact relation for Q^{-1} (eq. 7) evaluated at the relaxation frequencies ω_j. Comparison to the linearized version, eq. (21) leads to a correction factor for each a_j due to the term in the denominator neglected in the linearization. These factors can be defined recursively starting from the lowest
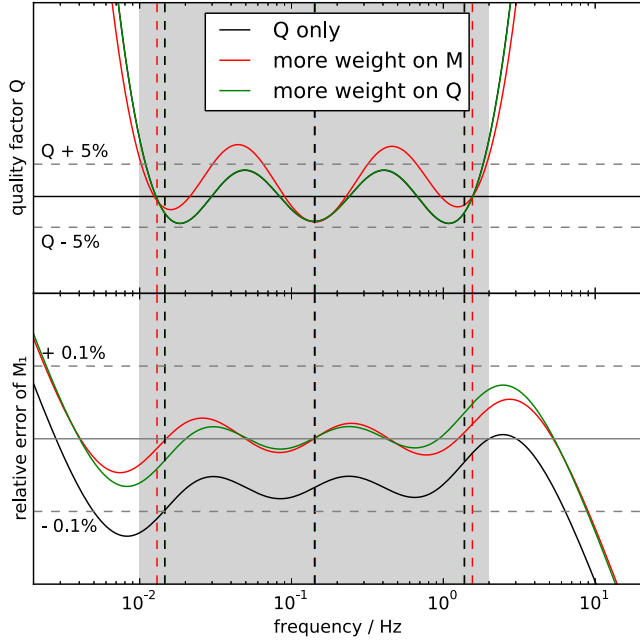
Figure 6. Optimized constant $Q \approx 250$ (black) versus simultaneous optimization of Q and real part of the modulus M_1 (red + green) using the relation by Kjartansson (1979) for a medium with three absorption bands: additional optimization of the modulus makes the parametrization independent of the choice of the reference frequency (here: $f_r = 2$ Hz) and allows for tuning the parametrization towards lower phase errors at the boundaries of the frequency range (red) or lower amplitude errors (green). Vertical dashed lines indicate the relaxation frequencies ω_j. Frequency weighting was not used for clarity.

relaxation frequency ω_0 with y_j as defined earlier:

$$\text{recursion}: \begin{cases} \delta_0 = 1 + \frac{1}{2} y_0 \\ \delta_{n+1} = \delta_n + \left(\delta_n - \frac{1}{2} \right) y_n + y_{n+1}, \end{cases}$$
$$y'_j = \delta_j \cdot y_j. \tag{24}$$

The performance of this correction is visualized in Fig. 7, where panel (a) compares Q computed using corrected parameters to the linearization and the exact result for $Q = 20$. While the linearization causes an error of almost 20 per cent in the high frequencies, the corrected version is still quite close to the optimal solution using the exact relation. Fig. 7(b) shows tests of the approximation as a function of Q for a variety of different bandwidths and numbers of absorption bands in terms of the frequency weighted log-l2 error as defined earlier. The more accurate the approximation of constant Q is desired, the higher the Q value where the linearization breaks down: for example, at 2.2 decades bandwidth and using three absorption bands (red), the linearization (dashed) doubles the error for Q values lower than 90, while the corrected ones hit this error bound only for $Q < 10$. On the global scale, where Q typically takes values larger than 50, this scheme allows to find the parameters a_j efficiently with very high accuracy.

This correction scheme has negligible computational imprint such that it may be used in the time loop, which means that the $2N$ coefficients a_j and w_j need to be stored only once and not for each gridpoint.
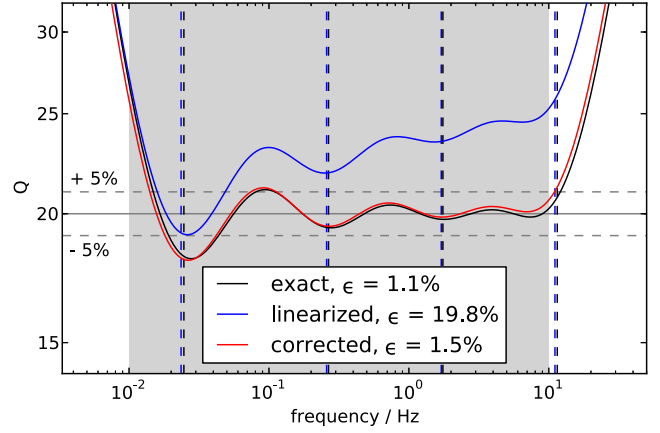
(a)
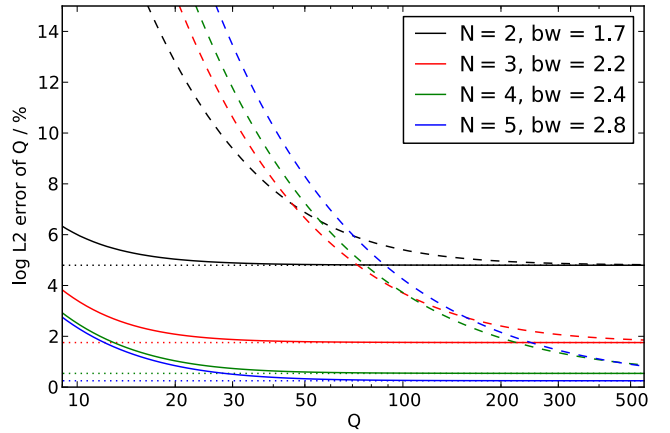
(b)

Figure 7. Recursive correction for the coefficients y_j as in eq. (24): (a) linearization of Q in y_j as suggested by Blanch *et al.* (1995; 'τ-method') leads to a logarithmic trend with the strongest effect in the high frequencies (blue), the correction (red) results in a behaviour very close to optimizing the exact relation (black) even for Q as low as 20 [on the global scale, the minimum value is $\approx$50 in the 3-D model QRLW8 by Gung & Romanowicz (2004)]. Errors are frequency-weighted, see text for details. (b) Frequency-weighted log-l2 error in Q as a function of Q for the linearization (dashed), the correction (solid) and the exact (dotted) relation.

2.4 Analytic time stepping

Many discrete schemes have been proposed to integrate the memory variable equation, eq. (5), over time to determine the values of the memory variables at the next time step: for example, finite differences (Day & Minster 1984), some unspecified second-order scheme (Emmerich & Korn 1987), second-order central difference (Kristek & Moczo 2003), ADER (Käser *et al.* 2007) or fourth-order Runge–Kutta (Komatitsch & Tromp 2002a; Savage *et al.* 2010). To our best knowledge, integrating the equation analytically has not been published in the seismological community.

The 'memory variable' equation is an ODE of the form

$$\dot{\zeta}^j(t) + \omega_j \zeta^j(t) = s^j(t), \tag{25}$$

with $s^j(t) = a_j \omega_j \delta M \epsilon(t)$. This can be solved by the standard method of multiplication with an integrating factor and the solution is

$$\zeta^j(t + \Delta t) = e^{-\omega_j \Delta t} \left[\zeta^j(t) + \int_t^{t+\Delta t} s^j(t') e^{\omega_j(t-t')} dt' \right]. \tag{26}$$

The integral can then be solved depending on the global time scheme and the corresponding time dependence of $s(t)$. For example, if the elastic equations are time integrated with a second-order scheme, where $s(t)$ is known at two times only [with linear interpolation: $s(t') = s(t) + \frac{t'-t}{\Delta t}(s(t + \Delta t) - s(t))$] the resulting discrete scheme is:

$$\zeta^j(t + \Delta t) = \zeta^j(t)e^{-\omega_j \Delta t}$$
$$+ \frac{s^j(t)}{\omega_j} \left[\frac{1}{\omega_j \Delta t}(1 - e^{-\omega_j \Delta t}) - e^{-\omega_j \Delta t} \right]$$
$$+ \frac{s^j(t + \Delta t)}{\omega_j} \left[1 - \frac{1}{\omega_j \Delta t}(1 - e^{-\omega_j \Delta t}) \right]. \quad (27)$$

Expanding the exponential functions in series, this agrees with the Runge–Kutta scheme by Savage *et al.* (2010) up to the fourth order in $\omega_j \Delta t$. In contrast to numerical integration schemes, the accuracy of this analytical scheme is only bounded by the accuracy of the source term (i.e. the elastic strain), which is determined by the time scheme used for the elastic equations. Thus, our formulation automatically ties the viscoelastic accuracy to that of the global time scheme.

2.5 Generalization to three dimensions

Generalization to three dimensions is rather straightforward and extensive literature is available, for example, Carcione (2007) and references therein. We restrict ourselves to slightly anisotropic media, where the effect of attenuation can still be treated as isotropic and the anisotropy of attenuation is neglected as a second-order effect. In this case, the eigenstrains and eigenstiffnesses are easy to find as dilation and shear stresses and the stress–strain relationship remains simple and computationally light, see for example, Carcione & Cavallini (1994). The validity of this approximation for global radial anisotropy as in PREM is verified by the benchmark in Section 5.

Many authors neglect the effect of bulk attenuation and take advantage by splitting the memory variable into the shear and bulk contribution such that only five instead of six memory variables for each absorption band are needed. However, this is no good approximation when there is significant contribution of the bulk quality factor Q_κ to the *P*-wave quality factor Q_p (e.g. Stein & Wysession 2003,3.7.6):

$$Q_p^{-1} = L Q_\mu^{-1} + (1 - L)Q_\kappa^{-1}, \quad (28)$$

$$L = \frac{4}{3}\left(\frac{v_s}{v_p}\right)^2. \quad (29)$$

In the PREM model (Dziewonski & Anderson 1981), this is the case in the inner core, where $Q_\kappa = 1327$ has a significant effect (around 30 per cent) on Q_p. Although there is no general agreement about the location of bulk attenuation in the Earth, it is needed to simultaneously fit high Q radial mode data and surface wave data (Romanowicz & Mitchell 2007). Hence we choose to keep the full 6 degrees of freedom and write the 3-D memory variable equation as:

$$\dot{\zeta}_l^j(t) + \omega_j \zeta_l^j(t)$$
$$= \begin{cases} \delta\kappa\, a_j^\kappa \omega_j \mathrm{tr}\epsilon + \frac{2}{3}\delta\mu\, a_j^\mu \omega_j (3\epsilon_l - \mathrm{tr}\epsilon), & (l = 1, 2, 3) \\ \delta\mu\, a_j^\mu \omega_j \epsilon_l, & (l = 4, 5, 6) \end{cases} \quad (30)$$

with index l denoting components in the Voigt notation (index mapping: $1 \to 11$, $2 \to 22$, $3 \to 33$, $4 \to 23$, $5 \to 31$, $6 \to 12$).

Importantly, using the scheme from above to find the medium parameters, ω_j are the same for bulk and shear Q, while the a_j are found depending on the actual value of Q_κ and Q_μ. Identifying the right-hand side with $s(t)$ from eq. (25), the analytical time stepping can readily be used in three dimensions.

3 *AxiSEM* DISCRETIZATION

In this section, we use *AxiSEM* (Nissen-Meyer *et al.* 2014) as an example and generalize the derivation of the equations of motions of the reduced 2-D equations in the weak form to include attenuation. The elastic anisotropic problem is treated by van Driel & Nissen-Meyer (2014) and the basic framework is derived in detail by Nissen-Meyer *et al.* (2007a,b, 2008). The reduced 2-D equations are found by projecting the wave equation onto test functions having the azimuthal dependence of monopole, dipole and quadrupole sources. Taking the dot product of the wave equation with a test function **w**, integrating over the domain Ω and using partial integration and the free surface boundary condition yields

$$\int_\Omega \left(\underbrace{\rho\mathbf{w} \cdot \ddot{\mathbf{u}}}_{\text{mass}} + \underbrace{\nabla\mathbf{w} : \sigma}_{\text{stiffness}} \right) d\Omega = \int_\Omega \underbrace{\mathbf{w} \cdot \mathbf{f}}_{\text{force}} d\Omega, \quad (31)$$

with the viscoelastic constitutional relation

$$\sigma = \sigma^{\text{el}} + \sigma^{\text{anel}} = c_U : \epsilon - \zeta. \quad (32)$$

Here $\zeta := \sum_{j=1}^N \zeta^j$ denotes the total stress contribution of all memory variables. As shown by Nissen-Meyer *et al.* (2007a) for spherically symmetric models and generalized to anisotropic axisymmetric models by van Driel & Nissen-Meyer (2014), the displacement in an axisymmetric model can be expanded in the series

$$\mathbf{u}(s, z, \varphi)$$
$$= \sum_{m=-\infty}^{\infty} \underbrace{\left[u_s^m(s, z)\,\hat{\mathbf{e}}_s(\varphi) + u_\varphi^m(s, z)\,\hat{\mathbf{e}}_\varphi(\varphi) + u_z^m(s, z)\,\hat{\mathbf{e}}_z \right]}_{=:\mathbf{u}_m(s,\varphi,z)} e^{im\varphi}, \quad (33)$$

where s, φ, z are the cylindrical coordinates as in Fig. 8. In the case of a moment tensor or single force point source on the symmetry axis, all contributions for $|m| > 2$ vanish. Equivalent expressions with $\sin\varphi$ and $\cos\varphi$ can be found by summing over pairs of $\pm m$ and using the fact that **u** is real. For monopole, dipole and quadrupole sources this results in

$$\mathbf{u}^m = \left[u_s^m\,\hat{\mathbf{e}}_s + u_z^m\,\hat{\mathbf{e}}_z \right] \cos(m\varphi + \varphi_0) - u_\varphi^m \sin(m\varphi + \varphi_0)\,\hat{\mathbf{e}}_\varphi, \quad (34)$$

where φ_0 depends on the orientation of the source. The reduced equations of motion can then be found by inserting this into eq. (31) and evaluating the integral in φ analytically.

3.1 Angular dependence and axial boundary conditions of the memory variables

As the initial condition for the memory variables is $\zeta = 0$ and the memory variable equation is linear, the angular dependence is determined by the source terms in eq. (30). The angular dependence
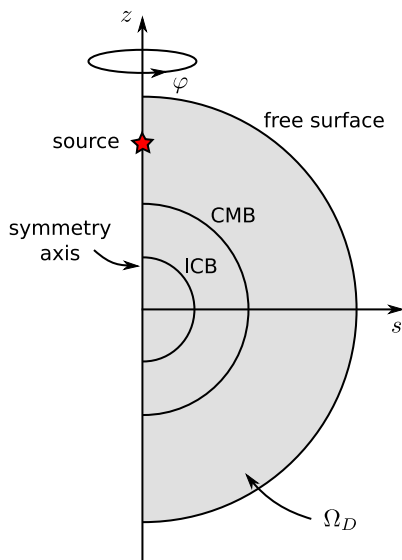
Figure 8. The cylindrical coordinate system (s, φ, z) and the reduced semicircular 2-D domain Ω_D for global wave propagation in axisymmetric media.

of the strain can be found from eq. (34) and we find

$$
\zeta_l(s, \varphi, z) = \begin{cases} \zeta_l(s, z) \cdot \cos{(m\varphi + \varphi_0)} & (l = 1, 2, 3, 5) \\ \zeta_l(s, z) \cdot \sin{(m\varphi + \varphi_0)} & (l = 4, 6), \end{cases} \quad (35)
$$

where $m = 0, 1, 2$ for monopole, dipole and quadrupole sources, respectively. With the same argumentation, the behaviour of the memory variables at the axis can be analysed based on the behaviour of the strain at the axis (Nissen-Meyer *et al.* 2007a). For the quadrupole case, ζ_4 and ζ_5 vanish at the axis with $\mathcal{O}(s)$, the others and in the case of monopole and dipole source all memory variables take finite values at the axis. These boundary conditions are needed to evaluate the Gauss–Jacobi quadrature at the axis in Section 3.3.

3.2 Anelastic stiffness terms in the weak form

The elastic stiffness terms σ^{el} remain the same as in purely elastic case, see van Driel & Nissen-Meyer (2014), where the elastic moduli are replaced with the unrelaxed moduli. Additionally, the contribution of the memory variables to the stiffness can be found for the monopole source as:

$$
\frac{1}{2\pi} \int_0^{2\pi} \nabla \mathbf{w} : \zeta \, \mathrm{d}\varphi
$$
$$
= \partial_s w_s \zeta_1 + \partial_z w_s \zeta_5 + \frac{w_s}{s} \zeta_2 + \partial_s w_z \zeta_5 + \partial_z w_z \zeta_3, \quad (36)
$$

Equivalent expressions for dipole and quadrupole sources can be found in Appendix A.

3.3 Spectral element discretization

The next step is to generalize the spatial discretization of the stiffness terms from the anisotropic elastic case presented in van Driel & Nissen-Meyer (2014) to the anelastic case. The approach is the same as in the isotropic elastic case and we refer the reader to section in Nissen-Meyer *et al.* (2007b) for details and restrict ourselves to a short summary of the method and important aspects of the notation

in the interest of brevity. A more general and rigorous approach can be found in Bernardi *et al.* (1999).

The collapsed 2-D domain Ω_D (Fig. 8) is divided into non-axial elements Ω_e and axial elements $\Omega_{\bar{e}}$. The mapping between reference coordinates $\xi, \eta \in [-1, 1]$ in each element and the physical coordinates s, z is provided by the Jacobian determinant

$$
\mathcal{J}(\xi, \eta) = \det \begin{pmatrix} s_\xi & s_\eta \\ z_\xi & z_\eta \end{pmatrix}, \quad (37)
$$

where the subscript denotes partial differentiation, $s_\xi = \partial_\xi s$, etc. Both the test function $\mathbf{w}$ and the field variables $\mathbf{u}$ and ζ are expanded in Lagrangian polynomials l_i of order N_{gll} (defined on the integration points, see below) within each element as

$$
u^\alpha(\xi, \eta, t) = \sum_{i,j=0}^{N_{\mathrm{gll}}} u_{ij}^\alpha(t) l_i(\xi) l_j(\eta),
$$
$$
\zeta_l(\xi, \eta, t) = \sum_{i,j=0}^{N_{\mathrm{gll}}} \zeta_{ij}^l(t) l_i(\xi) l_j(\eta), \quad (38)
$$

for each component $\alpha \in (s, \varphi, z)$ and $l \in \{1, \ldots, 6\}$ and equivalently to $\mathbf{u}$ for $\mathbf{w}$. For the axial elements $\xi = 0$ is the axis. The integral over the domain Ω_D is then split into a sum of integrals over elements and approximated using the Gauss–Lobatto integration rule

$$
\int_{\Omega_e} u(s, z) s \, \mathrm{d}s \, \mathrm{d}z \approx \sum_{pq} \sigma_p \sigma_q s(\xi_p, \eta_q) u_{pq} \mathcal{J}(\xi_p, \eta_q) \quad (39)
$$

with Gauss–Lobatto–Legendre (GLL) integration weights σ_p and integration points ξ_p and η_q. For the axial elements, Gauss–Lobatto–Jacobi (GLJ) quadrature is used for the ξ-direction with

$$
\int_{\Omega_{\bar{e}}} u(s, z) s \, \mathrm{d}s \, \mathrm{d}z \approx \sum_{pq} \bar{\sigma}_p (1 + \bar{\xi}_p)^{-1} \sigma_q s(\bar{\xi}_p, \eta_q) u_{pq} \mathcal{J}(\bar{\xi}_p, \eta_q), \quad (40)
$$

and GLJ integration weights $\bar{\sigma}_p$, integration points $\bar{\xi}_p$ and the Lagrangian interpolation polynomial on these points $\bar{l}(\xi)$. This allows to use l'Hospital's rule to calculate derivatives at the axis where needed.

Applying this discretization to eq. (31), choosing the set of test functions to be 1 in one component at a specific integration point and 0 at the others and summing over all elements we obtain the global set of ordinary differential equations in time

$$
\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) - \mathbf{z}(t) = \mathbf{f}(t), \quad (41)
$$

with the global mass matrix $\mathbf{M}$, stiffness matrix $\mathbf{K}$ and the additional anelastic stiffness vector $\mathbf{z}$. While the assembled mass matrix is diagonal in the GLL/GLJ basis (hence trivial to invert), it is unnecessary to compute $\mathbf{K}$ explicitly and we only evaluate its action on the displacement ($\mathbf{K}\mathbf{u}$). This term only appears on the right-hand side of the second-order system

$$
\ddot{\mathbf{u}}(t) = \mathbf{M}^{-1}[\mathbf{f}(t) - \mathbf{K}\mathbf{u}(t) + \mathbf{z}(t)], \quad (42)
$$

which is solved by explicit numerical time integration schemes, additional to the memory variable equation, eq. (30). The stiffness terms $\mathbf{K}\mathbf{u}$ and $\mathbf{z}$ are computed in each element first and the global stiffness is assembled subsequently (Nissen-Meyer *et al.* 2007b, section 4).

We split the original elemental stiffness integral into contributions from each component of the vectorial test function $\mathbf{w}$, denoted by the subscripts s and z. Furthermore, we revert to a tensorial notation instead of elemental sums and define the matrix–matrix

products

$$\mathbf{X} = \mathbf{A} \otimes \mathbf{B} : \qquad X_{ij} = \sum_k A_{ik} B_{kj}, \tag{43}$$

$$\mathbf{X} = \mathbf{A} \odot \mathbf{B} : \qquad X_{ij} = A_{ij} B_{ij}, \tag{44}$$

and vector–matrix and vector–vector products

$$\mathbf{X}^0 = \mathbf{A}^0 \otimes \mathbf{B} : \qquad X_{0j} = \sum_k A_{0k} B_{kj},$$

$$\mathbf{X}^0 = \mathbf{A}^0 \odot \mathbf{B}^0 : \qquad X_{0j} = A_{0j} B_{0j},$$

$$\mathbf{X} = \mathbf{A}^0 \mathbf{B}^0 : \qquad X_{ij} = A_{i0} B_{0j}. \tag{45}$$

The elemental anelastic stiffness can then be written in the monopole case as

$$\mathbf{z} = \mathbf{z}_s + \mathbf{z}_z \tag{46}$$

with the definitions from Table 1:

$$\mathbf{z}_s = \mathbf{D}_\xi \otimes \left(\mathbf{V}_{z_\eta} \odot \zeta_1 + \mathbf{V}_{s_\eta} \odot \zeta_5 \right)$$

$$+ \left(\mathbf{V}_{z_\xi} \odot \zeta_1 + \mathbf{V}_{s_\xi} \odot \zeta_5 \right) \otimes \mathbf{D}_\eta^\mathrm{T} + \mathbf{Y} \odot \zeta_2$$

$$+ \delta_{e\bar{e}} \mathbf{D}_\xi^0 \left[\mathbf{V}_{z_\eta}^0 \odot \zeta_1^0 + \mathbf{V}_{s_\eta}^0 \odot \zeta_5^0 \right],$$

$$\mathbf{z}_z = \mathbf{D}_\xi \otimes \left(\mathbf{V}_{z_\eta} \odot \zeta_5 + \mathbf{V}_{s_\eta} \odot \zeta_3 \right)$$

$$+ \left(\mathbf{V}_{z_\xi} \odot \zeta_5 + \mathbf{V}_{s_\xi} \odot \zeta_3 \right) \otimes \mathbf{D}_\eta^\mathrm{T}$$

$$+ \delta_{e\bar{e}} \Big\{ \mathbf{D}_\xi^0 \left[\mathbf{V}_{z_\eta}^0 \odot \zeta_5^0 + \mathbf{V}_{s_\eta}^0 \odot \zeta_3^0 \right]$$

$$+ \left[\mathbf{V}_{z_\eta}^0 \odot \zeta_5^0 + \mathbf{V}_{s_\eta}^0 \odot \zeta_3^0 \right] \otimes \mathbf{D}_\eta^\mathrm{T} \Big\}. \tag{47}$$

Here $\delta_{e\bar{e}}$ is 1 in axial elements and 0 in all others and ζ^0 is the vector of memory variables at the axis ($i = 0$). Equivalent terms for dipole and quadrupole case can be found in Appendix B.

4 COARSE-GRAINED MEMORY VARIABLES

In the following, we assume constant medium properties within each element. This assumption is used in the derivation of most spectral-element schemes, often without stating it explicitly. Empirically it is clear however, that correct results are obtained when this assumption is violated by small deviations only. We can then write the approximation of the theoretical or optimal modulus with standard linear solids as

$$M_\mathrm{opt}(\omega) \approx M_r + \delta M(\omega), \tag{48}$$

with the reference frequency ω_r that should be within the frequency band of interest and the modulus at this frequency M_r, that is, $\delta M(\omega_r) = 0$ (see Fig. 9 for a schematic sketch). This choice of reference is equivalent to the 'element-specific modulus' by Graves & Day (2003), that is, the unrelaxed modulus that is used for the elastic stiffness is not the reference and will hence be not constant over the element. The idea of coarse grained attenuation then is to find a medium where δM is a function of space with the constraints (1) that it behaves macroscopically as the homogeneous medium and (2) reduces the numerical burden. This calls for a method to compute the macroscopic behaviour of a medium with structure on
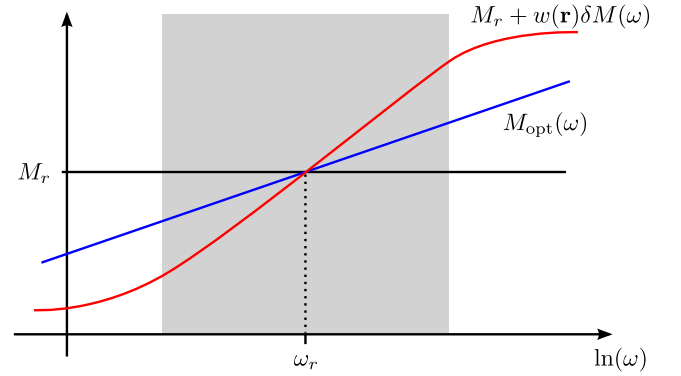

Figure 9. Coarse-grained attenuation schematically: constant Q corresponds to $M_\mathrm{opt}(\omega)$. The problem is to find $\delta M(\omega)$ using standard linear solids for the anelastic GLL points. The constraint is that averaged with the elastic GLL points with modulus M_0, the medium should behave as $M_\mathrm{opt}(\omega)$ in the frequency range of interest (shaded in grey).

the subwavelength scale where the deviations are relative small in amplitude.

In a seminal paper, Backus (1962) showed how to average a layered medium to find a homogeneous long-wavelength equivalent medium. A similar argument was introduced in the context of heterogeneous finite-difference methods by Boore (1972) and further developed by Zahradnik *et al.* (1993) and Moczo *et al.* (2002). In more recent developments of homogenization theory (e.g. Capdeville *et al.* 2010; Guillot *et al.* 2010), more rigorous analytical arguments are raised and in the 1-D case, it can be shown analytically that the correct way of averaging the modulus is the harmonic average (Cioranescu & Donato 1999). Graves & Day (2003) show that harmonic averaging also yields better results than arithmetic averaging in the context of the coarse-grained memory variables in the case of small values of $Q < 20$ (where δM is relatively large) and yields the same results as arithmetic averaging if $Q > 20$. Although Q is larger than 50 for most global models, it is not obvious to generalize this conclusion to global body waves, because the test cases presented by Graves & Day (2003) are dealing with an order of magnitude-less wavelengths propagation distance (compare Section 2.2). We thus consider arithmetic averaging at first because of its simplicity and benchmark it in the next section, finding that it does suffice for global high-frequency body waves.

In the case of arithmetic averaging, the modulus can be defined as

$$M(\mathbf{r}, \omega) = M_r + w(\mathbf{r}) \delta M(\omega), \tag{49}$$

with a weighting function $w(\mathbf{r})$ and subject to the normalization condition

$$V_e = \int_{V_e} \mathrm{d} V_e = \int_{V_e} w(\mathbf{r}) \, \mathrm{d} V_e, \tag{50}$$

where V_e denotes the volume of the element. The integrals are now replaced by the quadrature rules of the SEM scheme. For *AxiSEM*, we additionally assume an axisymmetric weighting function $w(\mathbf{r}) = w(s, z)$ and split the volume integral into the analytical integration in φ and the quadrature over the element. The constraint then reads

$$V_e = \sum_{pq} \gamma_w^{pq} = \sum_{pq} \gamma_w^{pq} w^{pq}, \tag{51}$$

Table 1. Definitions for pre-computable matrices (i.e. prior to the costly time extrapolation) of the anelastic stiffness terms, $\pm$ takes its value depending on the combination of x_ζ as in the right-hand table. Subscript reference coordinates denotes partial derivation, for example, $x_\xi = \partial_\xi x$. For consistency with the summation notation in Nissen-Meyer *et al.* (2007a), we use indices i, j and I, J, which all take the values in $\{0 \ldots N_{\text{gll}}\}$.

Matrix	Non-axial elements	Axial elements $(i > 0)$	$(i = 0)$	Axial vectors
$(\mathbf{V}_{x_\zeta})^{ij}$	$\pm \sigma_i \sigma_j x_\zeta^{ij} s^{ij}$	$\pm \bar{\sigma}_i \sigma_j (1 + \bar{\xi}_i)^{-1} x_\zeta^{ij} s^{ij}$	0	$(\mathbf{V}_{x_\zeta}^0)^j = \pm \bar{\sigma}_0 \sigma_j x_\zeta^{0j} s_\xi^{0j}$
$(\mathbf{Y})^{ij}$	$\sigma_i \sigma_j \mathcal{J}^{ij}$	$\bar{\sigma}_i \sigma_j (1 + \bar{\xi}_i)^{-1} \mathcal{J}^{ij}$	0	$(\mathbf{Y}^0)^j = \bar{\sigma}_0 \sigma_j \mathcal{J}^{0j}$
$(\mathbf{D}_\xi)^{Ii}$	$\partial_\xi l_I(\xi_i)$	$\partial_\xi \bar{l}_I(\bar{\xi}_i)$	$\partial_\xi \bar{l}_I(\bar{\xi}_0)$	$(\mathbf{D}_\xi^0)^I = \partial_\xi \bar{l}_I(\bar{\xi}_0)$
$(\mathbf{D}_\eta)^{Jj}$	$\partial_\eta l_J(\eta_j) = \partial_\xi l_J(\xi_j)$	$\partial_\eta l_J(\eta_j)$		

$\pm(x_\zeta)$	$\zeta = \xi$	$\zeta = \eta$
$x = s$	$+$	$-$
$x = z$	$-$	$+$

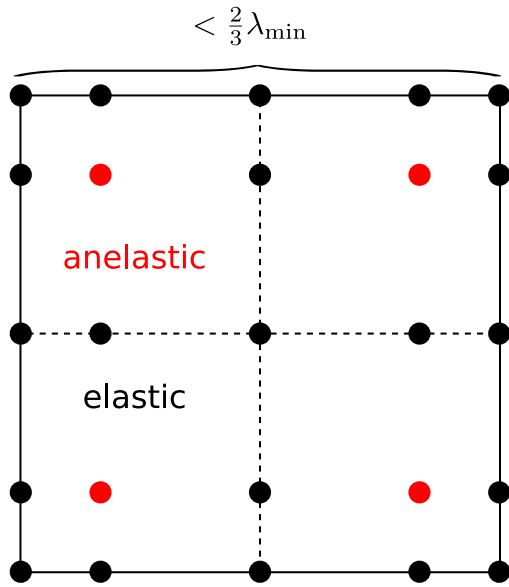$$< \tfrac{2}{3}\lambda_{\min}$$

Figure 10. Distribution of Gauss–Lobatto–Legendre (GLL) integration points in the reference element of the fourth-order scheme. For the coarse-grained attenuation, anelasticity is concentrated on the four red GLL points, while the black points are treated as elastic. During meshing, element sizes are chosen to be smaller than $\frac{2}{3}$ of the smallest wavelength, so that there are at least three anelastic points per wavelength.

with

$$\gamma_w^{pq} = \begin{cases} \sigma_p \sigma_q J^{pq} s^{pq}, & (\text{non} - \text{axial}) \\ \bar{\sigma}_p (1 + \xi_p)^{-1} \sigma_q J^{pq} s^{pq}, & (\text{axial}, \ p > 0) \\ \bar{\sigma}_0 \sigma_q J^{0q} s_\xi^{0q}, & (\text{axial}, \ p = 0) \end{cases} \tag{52}$$

and $w^{pq} = w(\xi_p, \eta_q)$. One possible solution is to assign the weight of neighbouring elastic GLL points to the selected anelastic GLL points as sketched in Fig. 10. The weighting function for the fourth-order spacial scheme could then, for example, be chosen as:

$$w^{11} = \frac{1}{\gamma_w^{11}} \left[\gamma_w^{00} + \gamma_w^{01} + \gamma_w^{10} + \gamma_w^{11} \right.$$

$$\left. + \frac{1}{2} \left(\gamma_w^{02} + \gamma_w^{12} + \gamma_w^{20} + \gamma_w^{21} \right) + \frac{1}{4} \gamma_w^{22} \right],$$

$$w^{13} = \frac{1}{\gamma_w^{13}} \left[\gamma_w^{03} + \gamma_w^{04} + \gamma_w^{13} + \gamma_w^{14} \right.$$

$$\left. + \frac{1}{2} \left(\gamma_w^{02} + \gamma_w^{12} + \gamma_w^{23} + \gamma_w^{24} \right) + \frac{1}{4} \gamma_w^{22} \right],$$

$$w^{31} = \frac{1}{\gamma_w^{31}} \left[\gamma_w^{30} + \gamma_w^{31} + \gamma_w^{40} + \gamma_w^{41} \right.$$

$$\left. + \frac{1}{2} \left(\gamma_w^{20} + \gamma_w^{21} + \gamma_w^{32} + \gamma_w^{42} \right) + \frac{1}{4} \gamma_w^{22} \right],$$

$$w^{33} = \frac{1}{\gamma_w^{33}} \left[\gamma_w^{33} + \gamma_w^{34} + \gamma_w^{43} + \gamma_w^{44} \right.$$

$$\left. + \frac{1}{2} \left(\gamma_w^{23} + \gamma_w^{24} + \gamma_w^{32} + \gamma_w^{42} \right) + \frac{1}{4} \gamma_w^{22} \right],$$

$$w^{ij} = 0, \qquad (\text{else}). \tag{53}$$

The choice of GLL points is guided by the criterion derived by Day (1998), stating that the periodicity of the medium should have smaller scale than half the shortest wavelength to be propagated, that is, the elastic equivalent of the Bragg condition for normal incidence. In the specific case of five GLL points per element in each dimension that are meshed with $\frac{2}{3}\lambda_s$ as element size, this leads to the choice of two anelastic points per element in each dimension. In principle, the dissipation mechanisms for the lower absorption bands could be placed even sparser, as they only affect longer wavelengths. While this would reduce memory usage and computational complexity in time stepping of the memory variables, we chose not to do so for simplicity in implementation and the fact that the stiffness term ζ would end up less sparse if another distribution was chosen. Also, additional source terms for the memory variables would cause extra cost, as the strain is no field variable in our scheme. This might be different depending on the numerical scheme used: if the strain is a field variable and does not require extra computation, the decision might be different.

4.1 Numerical efficiency

The performance can benefit from the coarse grained implementation in multiple ways: The number of memory variables per element is smaller, so these use less memory. Also, the number of memory variable equations that need to be integrated in time and hence the

Table 2. Memory usage relative to the elastic case in terms of a multiplicative factor (Newmark time integration).

N	Full memory variables	Coarse grained
3	1.7	1.1
4	1.8	1.1
5	1.9	1.1
6	2.0	1.1
7	2.2	1.2
8	2.3	1.2

number of points at which the source $s(t)$ of these equations needs to be computed is less. The pre-computed matrices for the anelastic stiffness need only be computed and stored for the elastic GLL points. On top of that, the sparsity of $\boldsymbol{\zeta}$ can be exploited by using sparse implementations of the matrix products $\otimes$ and $\odot$.

For the Hadamard product $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$, eq. (43), it is clear that $\mathbf{C}$ has the joint sparsity of $\mathbf{A}$ and $\mathbf{B}$. For the matrix product $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$, eq. (44), the implementation depends on which of the two matrices is sparse. In the specific case of chosen GLL points shown in Fig. 10 and sparse $\mathbf{A}$ we have

$$C^{ij} = \begin{cases} A^{i1} \cdot B^{1j} + A^{i3} \cdot B^{3j}, & (i \in \{1, 3\}, \text{ for all } j) \\ 0, & (\text{else}) \end{cases} \qquad (54)$$

and for sparse $\mathbf{B}$

$$C^{ij} = \begin{cases} A^{i1} \cdot B^{1j} + A^{i3} \cdot B^{3j}, & (j \in \{1, 3\}, \text{ for all } i) \\ 0, & (\text{else}). \end{cases} \qquad (55)$$

The observed performance gains in *AxiSEM* are summarized in Tables 2 and 3, showing roughly a factor of 2 in total memory reduction and a factor of 5 in computational time both for anelastic stiffness and memory variable time stepping, resulting in a total simulation speedup of around two. The additional cost of the anelastic terms compared to the elastic ones hence reduces from 150–200 per cent to 35–45 per cent for the Newmark time scheme. The ideal value for the speedup in anelastic stiffness and memory variable time stepping would be $\frac{25}{4} = 6.25$. In a full 3-D code, the dimensionality would increase this to $\frac{125}{8} \approx 16$. The anelastic addition to the elastic stiffness term in current implementations accounts for about 2/3 of the total runtime, and would then be diminished to the point where the elastic stiffness terms always dominate the computational cost.

5 BENCHMARK

Solving the full 3-D wave equation for arbitrary earthquake sources in axisymmetric anisotropic models, *AxiSEM* seems to be unique among available codes. For benchmarking we thus revert to spheri-

cally symmetric models. As a reference, we use *Yspec* by Al-Attar & Woodhouse (2008), which is a generalization of the direct radial integration method (Friederich & Dalkolmo 1995) including self-gravitation (switched off for the benchmark).

While Nissen-Meyer *et al.* (2008) could only perform benchmarks down to 20 s period due to limitations in the reference normal mode solution, this limit is overcome using *Yspec*. Also, *AxiSEM* since then has experienced some substantial development (Nissen-Meyer *et al.* 2014), specifically the improved parallelization allows us to perform production runs up to the highest frequencies observed for global body waves.

Fig. 11 shows a record section of seismograms computed for the anisotropic PREM model (Dziewonski & Anderson 1981) with continental crust computed with *Yspec* and *AxiSEM*. The source is a normal faulting event with a moment magnitude $M_w = 5.0$ in 117 km depth under the Tonga islands. The traces recorded at some selected GSN stations are filtered between 10 and 1 s period. Due to the high-frequency content, it is necessary to zoom in to see any differences at all: the agreement between the two methods is remarkable even though the highest frequencies have travelled more than 1000 wavelengths (given the low-pass filter at 1 s, the time axis is equivalent to the number of travelled wavelengths).

We use the phase and envelope misfit *PM* and *EM* as defined by Kristekova *et al.* (2009) for quantitative comparison within the zoom windows and find phase misfits below 1.6 per cent for all windows and envelope misfits below 3.1 per cent for all windows but the extremely small amplitude phase *ScS* at KNTN, where it reaches a maximum of 4.4 per cent. These differences between *AxiSEM* and *Yspec* are slightly larger than in the purely elastic case (van Driel & Nissen-Meyer 2014), which is not surprising for a number of reasons: While Q is approximated using standard linear solids in *AxiSEM*, *Yspec* uses exactly logarithmic dispersion. The symplectic time scheme was developed for conservative systems (Nissen-Meyer *et al.* 2008), while here the attenuation causes a slight energy loss. Also, the coarse-grained memory variable approach was implemented using arithmetic averaging instead of harmonic averaging and we neglected attenuation in the fluid outer core. Errors in amplitude and phase are therefore negligibly small compared to other errors when comparing these synthetics to data like, for example, noise or the assumption of a 1-D model.

The total cost of this extreme case, demanding run over 1800 travelled wavelengths at very high accuracy with *AxiSEM* was about 88K CPU hours on 11 048 cores [four simultaneous parallel jobs according to the decomposition described by Nissen-Meyer *et al.* (2007a)] using a fourth-order symplectic time scheme and five relaxation frequencies on a Cray XE6. The mesh was built for periods down to 0.8 s and the time step chosen 30 per cent below the CFL criterion, as this run was meant to prove convergence to the same result as *Yspec*. In applications where less accuracy is necessary one could either use the same traces at higher frequencies or reduce this cost substantially by choosing a larger time step and a coarser mesh (for instance, a dominant period of 1.6 s at otherwise fixed

Table 3. Computation time relative to elastic case in terms of a multiplicative factor (Newmark time integration).

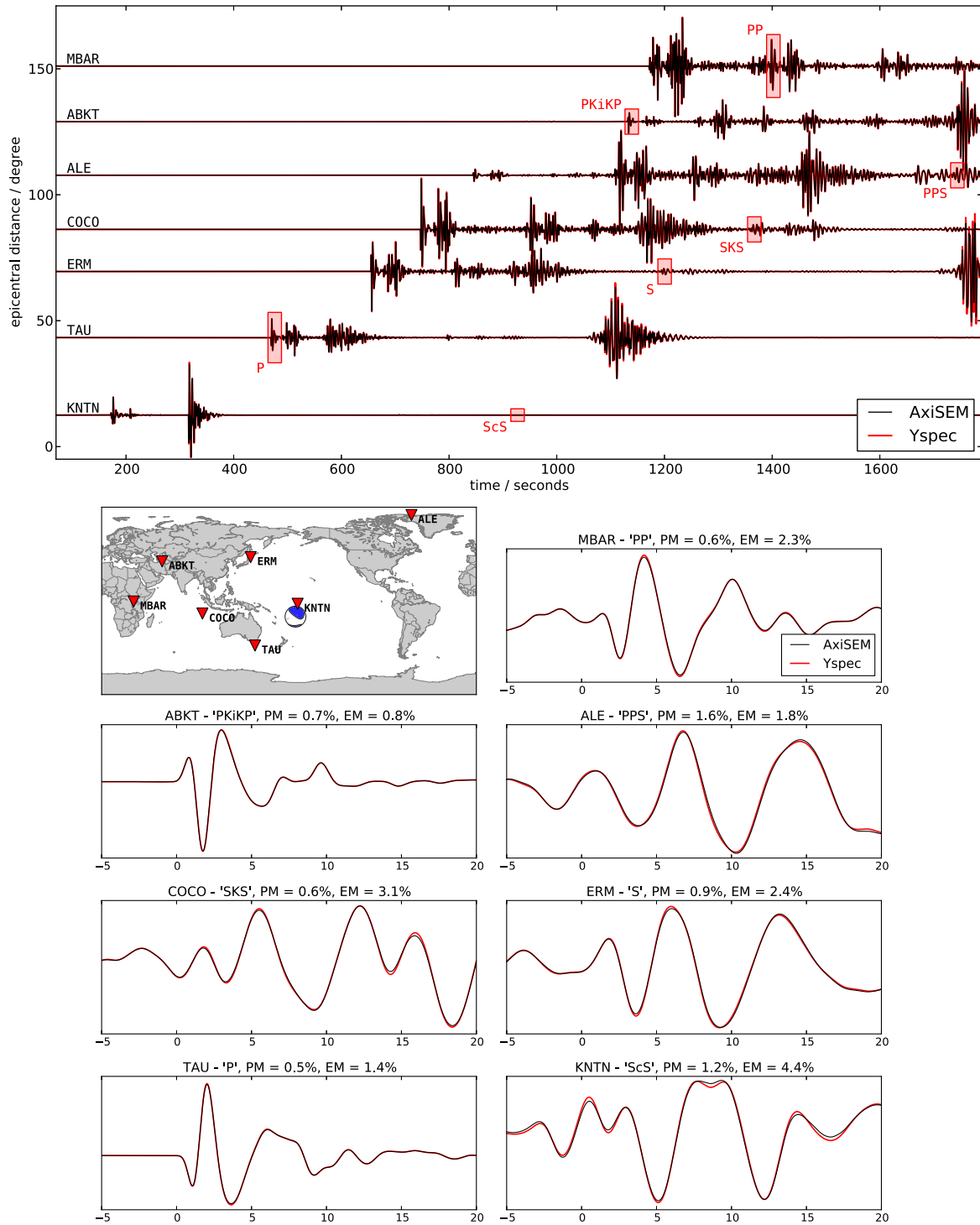| N	Elastic runtime	Full memory variables runtime	Coarse-grained runtime	Total speedup	Anelastic stiffness speedup	Anelastic time step speedup
3	1	2.55	1.35	1.9	5.3	4.3
4	1	2.75	1.38	2.0	5.2	4.5
5	1	3.0	1.41	2.1	5.3	4.7
6	1	3.16	1.44	2.2	5.5	4.7

Figure 11. Comparison of vertical displacement seismograms (bandpass filtered from 10 to 1 s period) for a moment magnitude $M_w = 5.0$ event in 126 km depth under the Tonga islands, computed with *AxiSEM* and *Yspec* in the anisotropic PREM model without ocean but including attenuation. The traces are recorded at the GSN stations indicated in the map. The zoom windows are indicated with red rectangles in the record section and the timescale is relative to the ray-theoretical arrival. *EM* and *PM* denote the envelope and phase misfit in the time window plotted (Kristekova *et al.* 2009). The largest envelope error is 4.4 per cent for the small amplitude phase *ScS* at KNTN.

parameters is already an order of magnitude cheaper to compute due to the scaling with the third power of the period).

The additional cost for attenuation using the fourth-order symplectic time scheme and coarse-grained memory variables compared to a purely elastic simulation was only 26 per cent.

6 CONCLUSION AND OUTLOOK

In this paper, we presented three improvements for more efficient implementation of attenuation in time-domain wave-propagation methods. First, we showed how to find the optimal set of medium parameters to minimize errors in the wavefield at a fixed

numerical complexity. To do so, we analysed first-order effects of errors in this parametrization onto the wavefield. Also, we derived an approximate method to find these parameters in a full-scale application, that is, for 10^6 or more different values of Q and the modulu. Furthermore, we suggested an analytical time-stepping scheme, that ties the error of the anelastic time integration to that of the global scheme. These findings are completely independent of the spatial scheme.

Finally, we generalized the coarse-grained memory variable approach to the spectral element scheme, noting that the same method could be used in other high-order finite-element schemes. This is a physical approximation reducing the cost of attenuation by a factor of 5 in the 2-D case of *AxiSEM*. Future work includes implementation of the coarse-grained scheme into high-order 3-D methods including homogeneous averaging for lower Q values.

REFERENCES

Al-Attar, D. & Woodhouse, J. H., 2008. Calculation of seismic displacement fields in self-gravitating earth models: applications of minors vectors and symplectic structure, *Geophys. J. Int.,* **175**(3), 1176–1208.

Alterman, Z. & Karal, F., 1968. Propagation of elastic waves in layered media by finite difference methods, *Bull. seism. Soc. Am.,* **58**(1), 367–398.

Backus, G., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.,* **67**(11), 4427–4440.

Bernardi, C., Dauge, M. & Maday, Y., 1999. *Spectral Methods for Axisymmetric Domains,* Gauthier-Villars.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. ObsPy: a python toolbox for seismology, *Seismol. Res. Lett.,* **81**(3), 530–533.

Blanch, J.O., Robertsson, J.O.A. & Symes, W.W., 1995. Modeling of a constant Q: methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique, *Geophysics,* **60**(1), 176–184.

Boore, D.M., 1972. Finite-difference methods for seismic wave propagation in heterogeneous materials, in *Methods in Computational Physics,* Vol. **11**, pp. 1–37, ed. Bolt, B.A., Academic Press.

Capdeville, Y., Guillot, L. & Marigo, J.-J., 2010. 1-D non-periodic homogenization for the seismic wave equation, *Geophys. J. Int.,* **181**, 897–910.

Carcione, J.M., 2007. *Wave Fields in Real Media,* 2nd edn, Elsevier.

Carcione, J.M. & Cavallini, F., 1994. A rheological model for anelastic anisotropic media with applications to seismic wave propagation, *Geophys. J. Int.,* **119**(1), 338–348.

Carcione, J., Kosloff, D. & Kosloff, R., 1988. Wave propagation simulation in a linear viscoelastic medium, *Geophys. J.,* **95**, 597–611.

Carrington, L., Komatitsch, D., Laurenzano, M., Tikir, M., Michéa, D., Le Goff, N., Snavely, A. & Tromp, J., 2008. High-frequency simulations of global seismic wave propagation using SPECFEM3D_GLOBE, in *Proceedings of the ACM/IEEE Supercomputing SC'2008 Conference,* Austin, TX, pp. 1–11.

Chaljub, E. & Tarantola, A., 1997. Sensitivity of SS precursors to topography on the uppermantle 660 km discontinuity, *Geophys. Res. Lett.,* **24**(21), 2613–2616.

Christensen, R., 1982. *Theory of Viscoelasticity: An Introduction,* 2nd edn, Academic Press.

Cioranescu, D. & Donato, P., 1999. *An Introduction to Homogenization,* Oxford Univ. Press.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology,* pbk. edn, Princeton Univ. Press.

Day, S.M., 1998. Efficient simulation of constant Q using coarse-grained memory variables, *Bull. seism. Soc. Am.,* **88**(4), 1051–1062.

Day, S. & Bradley, C., 2001. Memory-efficient simulation of anelastic wave propagation, *Bull. seism. Soc. Am.,* **91**, 520–531.

Day, S. & Minster, J., 1984. Numerical simulation of attenuated wavefields using a Padé approximant method, *Geophys. J. R. astr. Soc.,* **78**, 104–118.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25**(4), 297–356.

Emmerich, H. & Korn, M., 1987. Incorporation of attenuation into time-domain computations of seismic wave fields, *Geophysics,* **52**(9), 1252–1264.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.,* **179**(3), 1703–1725.

Friederich, W. & Dalkolmo, J., 1995. Complete synthetic seismograms for a spherically symmetric earth by a numerical computation of the Green's function in the frequency domain, *Geophys. J. Int.,* **122**(2), 537–550.

Furumura, T., Kennett, B.L.N. & Furumura, M., 1998. Seismic wavefield calculation for laterally heterogeneous whole earth models using the pseudospectral method, *Geophys. J. Int.,* **135**, 845–860.

Graves, R.W. & Day, S.M., 2003. Stability and accuracy analysis of coarse-grain viscoelastic simulations, *Bull. seism. Soc. Am.,* **93**(1), 283–300.

Guillot, L., Capdeville, Y. & Marigo, J.-J., 2010. 2-D non-periodic homogenization of the elastic wave equation: SH case, *Geophys. J. Int.,* **182**(3), 1438–1454.

Gung, Y. & Romanowicz, B., 2004. Q tomography of the upper mantle using three-component long-period waveforms, *Geophys. J. Int.,* **157**(2), 813–830.

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. & Weber, M., 1996. P-SV wave propagation in the Earth's mantle using finite differences: application to heterogeneous lowermost mantle structure, *Geophys. Res. Lett.,* **23**(5), 415–418.

Igel, H., Mora, P. & Riollet, B., 1995. Anisotropic wave propagation through finite-difference grids, *Geophysics,* **60**(4), 1203–1216.

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.

Käser, M., Dumbser, M., de la Puente, J. & Igel, H., 2007. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes: III. Viscoelastic attenuation, *Geophys. J. Int.,* **168**(1), 224–242.

Kjartansson, E., 1979. Constant Q-wave propagation and attenuation, *J. geophys. Res.,* **84**, 4737–4748.

Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation-II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophys. J. Int.,* **150**(1), 303–318.

Komatitsch, D. & Tromp, J., 2002b. Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.,* **149**(2), 390–412.

Kristek, J. & Moczo, P., 2003. Seismic-wave propagation in viscoelastic media with material discontinuities: a 3D fourth-order staggered-grid finite-difference modeling, *Bull. seism. Soc. Am.,* **93**(5), 2273–2280.

Kristekova, 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.

Liu, P. & Archuleta, R.J., 2006. Efficient modeling of Q for 3D numerical simulation of wave propagation, *Bull. seism. Soc. Am.,* **96**(4A), 1352–1358.

Liu, H.-P., Anderson, D.L. & Kanamori, H., 1976. Velocity dispersion due to anelasticity: implications for seismology and mantle composition, *Geophys. J. R. astr. Soc.,* **47**(1), 41–58.

Ma, S. & Liu, P., 2006. Modeling of the perfectly matched layer absorbing boundaries and intrinsic attenuation in explicit finite-element methods, *Bull. seism. Soc. Am.,* **96**(5), 1779–1794.

Megies, T., Beyreuther, M., Barsch, R., Krischer, L. & Wassermann, J., 2011. ObsPy: what can it do for data centers and observatories? *Ann. Geophys.,* **54**(1), 47–58.

Moczo, P. & Kristek, J., 2005. On the rheological models used for time-domain methods of seismic wave propagation, *Geophys. Res. Lett.,* **32**(1), 1–5.

Moczo, P., Kristek, J., Vavrarycuk, 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.

Moczo, P., Kristek, J. & Galis, M., 2014. *The Finite-Difference Modelling of Earthquake Motions: Waves and Ruptures,* Cambridge Univ. Press.

Müller, G., 1983. Rheological properties and velocity dispersion of a medium with power-law dependence of Q on frequency, *J. Geophys.,* **54**(1), 20–29.

Nissen-Meyer, T., Dahlen, F.A. & Fournier, A., 2007a. Spherical-earth Fréchet sensitivity kernels, *Geophys. J. Int.,* **168**(3), 1051–1066.

Nissen-Meyer, T., Fournier, A. & Dahlen, F.A., 2007b. A two-dimensional spectral-element method for computing spherical-earth seismograms: I. Moment-tensor source, *Geophys. J. Int.,* **168**(3), 1067–1092.

Nissen-Meyer, T., Fournier, A. & Dahlen, F.A., 2008. A 2-D spectral-element method for computing spherical-earth seismograms: II. Waves in solid-fluid media, *Geophys. J. Int.,* **174**(3), 873–888.

Nissen-Meyer, T., van Driel, M., Stähler, S.C., Hosseini, K., Hempel, S., Auer, L., Colombi, A. & Fournier, A., 2014. AxiSEM: broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth,* **5**(1), 425–445.

Nolet, G., 2008. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun,* Cambridge Univ. Press.

Romanowicz, B. & Mitchell, B., 2007. Deep Earth structure: Q of the Earth from crust to core, in *Treatise on Geophysics,* pp. 731–774, ed. Schubert, G., Elsevier.

Savage, B., Komatitsch, D. & Tromp, J., 2010. Effects of 3D attenuation on seismic wave amplitude and phase measurements, *Bull. seism. Soc. Am.,* **100**(3), 1241–1251.

Stein, S. & Wysession, M., 2003. *An Introduction to Seismology, Earthquakes and Earth Structure,* Vol. **9,** Blackwell Publishing.

Takenaka, H., Tanaka, H., Okamoto, T. & Kennett, B.L.N., 2003. Quasi-cylindrical 2.5D wave modeling for large-scale seismic surveys, *Geophys. Res. Lett.,* **30**(21), doi:10.1029/2003GL018068.

Thomas, C., Igel, H., Weber, M. & Scherbaum, F., 2000. Acoustic simulation of P-wave propagation in a heterogeneous spherical earth: numerical method and application to precursor waves to PKPdf, *Geophys. J. Int.,* **141**, 307–320.

Toyokuni, G. & Takenaka, H., 2006. FDM computation of seismic wavefield for an axisymmetric earth with a moment tensor point source, *Earth Planets Space,* **58**, e29–e32.

Toyokuni, G. & Takenaka, H., 2012. Accurate and efficient modeling of global seismic wave propagation for an attenuative Earth model including the center, *Phys. Earth planet. Inter.,* **200–201**, 45–55.

Toyokuni, G., Takenaka, H., Wang, Y. & Kennett, B.L.N., 2005. Quasi-spherical approach for seismic wave modeling in a 2-D slice of a global Earth model with lateral heterogeneity, *Geophys. Res. Lett.,* **32**, doi:10.1029/2004GL022180.

van Driel, M. & Nissen-Meyer, T., 2014. Seismic wave propagation in fully anisotropic axisymmetric media, *Geophys. J. Int.,* doi:10.1093/gji/ggu269.

Zahradnik, J., Moczo, P. & Hron, F., 1993. Testing four elastic finite-difference schemes for behavior at discontinuities, *Bull. seism. Soc. Am.,* **83**(1), 107–129.

APPENDIX A: ANELASTIC STIFFNESS TERMS IN THE WEAK FORM

The anelastic stiffness in the weak form for the dipole source ($m = 1$) reads

$$\frac{1}{\pi}\int_0^{2\pi} \nabla \mathbf{w} : \zeta \, d\varphi = \partial_s w_+(\zeta_1 - \zeta_6) + \partial_z w_+(\zeta_5 - \zeta_4)$$

$$+ \partial_s w_-(\zeta_1 + \zeta_6) + \partial_z w_-(\zeta_5 + \zeta_4) + 2\frac{w_-}{s}(\zeta_2 - \zeta_6)$$

$$+ \partial_s w_z \zeta_5 + \partial_z w_z \zeta_3 - \frac{w_z}{s}\zeta_4, \tag{A1}$$

and for the quadrupole source ($m = 2$):

$$\frac{1}{2\pi}\int_0^{2\pi} \nabla \mathbf{w} : \zeta \, d\varphi = \partial_s w_s \zeta_1 + \partial_z w_s \zeta_5 + \frac{w_s}{s}(\zeta_2 - 2\zeta_6)$$

$$- \partial_s w_\varphi \zeta_6 - \partial_z w_\varphi \zeta_4 + \frac{w_\varphi}{s}(\zeta_6 - 2\zeta_2)$$

$$+ \partial_s w_z \zeta_5 + \partial_z w_z \zeta_3 - 2\frac{w_z}{s}\zeta_4. \tag{A2}$$

APPENDIX B: SPECTRAL ELEMENT DISCRETIZATION

In the spectral element discretization, the anelastic stiffness terms for the dipole source are

$$\mathbf{z} = \mathbf{z}_+ + \mathbf{z}_- + \mathbf{z}_z \tag{B1}$$

with

$$\mathbf{z}_+ = \mathbf{D}_\xi \otimes \left[\mathbf{V}_{z_\eta} \odot (\zeta_1 - \zeta_6) + \mathbf{V}_{s_\eta} \odot (\zeta_5 - \zeta_4) \right]$$

$$+ \left[\mathbf{V}_{z_\xi} \odot (\zeta_1 - \zeta_6) + \mathbf{V}_{s_\xi} \odot (\zeta_5 - \zeta_4) \right] \otimes \mathbf{D}_\eta^T$$

$$+ \delta_{e\bar{e}} \Big\{ \mathbf{D}_\xi^0 \left[\mathbf{V}_{z_\eta}^0 \odot (\zeta_1^0 - \zeta_6^0) + \mathbf{V}_{s_\eta}^0 \odot (\zeta_5^0 - \zeta_4^0) \right]$$

$$+ \left[\mathbf{V}_{z_\eta}^0 \odot (\zeta_1^0 - \zeta_6^0) + \mathbf{V}_{s_\eta}^0 \odot (\zeta_5^0 - \zeta_4^0) \right] \otimes \mathbf{D}_\eta^T \Big\}, \tag{B2}$$

$$\mathbf{z}_- = \mathbf{D}_\xi \otimes \left[\mathbf{V}_{z_\eta} \odot (\zeta_1 + \zeta_6) + \mathbf{V}_{s_\eta} \odot (\zeta_5 + \zeta_4) \right]$$

$$+ \left[\mathbf{V}_{z_\xi} \odot (\zeta_1 + \zeta_6) + \mathbf{V}_{s_\xi} \odot (\zeta_5 + \zeta_4) \right] \otimes \mathbf{D}_\eta^T$$

$$+ 2\mathbf{Y} \odot (\zeta_2 - \zeta_6)$$

$$+ \delta_{e\bar{e}} \mathbf{D}_\xi^0 \left[\mathbf{V}_{z_\eta}^0 \odot (\zeta_1^0 + \zeta_6^0) + \mathbf{V}_{s_\eta}^0 \odot (\zeta_5^0 + \zeta_4^0) \right]$$

$$+ 2\mathbf{Y}^0 \odot (\zeta_2^0 - \zeta_6^0) \Big], \tag{B3}$$

$$\mathbf{z}_z = \mathbf{D}_\xi \otimes \left(\mathbf{V}_{z_\eta} \odot \zeta_5 + \mathbf{V}_{s_\eta} \odot \zeta_3 \right)$$

$$+ \left(\mathbf{V}_{z_\xi} \odot \zeta_5 + \mathbf{V}_{s_\xi} \odot \zeta_3 \right) \otimes \mathbf{D}_\eta^T - \mathbf{Y} \odot \zeta_4$$

$$+ \delta_{e\bar{e}} \mathbf{D}_\xi^0 \left[\mathbf{V}_{z_\eta}^0 \odot \zeta_5^0 + \mathbf{V}_{s_\eta}^0 \odot \zeta_3^0 - \mathbf{Y}^0 \odot \zeta_4^0 \right] \tag{B4}$$

and for the quadrupole source

$$\mathbf{z} = \mathbf{z}_s + \mathbf{z}_\varphi + \mathbf{z}_z \tag{B5}$$

with

$$
\begin{aligned}
\mathbf{z}_s &= \mathbf{D}_\xi \otimes \left(\mathbf{V}_{z_\eta} \odot \zeta_1 + \mathbf{V}_{s_\eta} \odot \zeta_5 \right) \\
&\quad + \left(\mathbf{V}_{z_\xi} \odot \zeta_1 + \mathbf{V}_{s_\xi} \odot \zeta_5 \right) \otimes \mathbf{D}_\eta^{\mathrm{T}} + \mathbf{Y} \odot (\zeta_2 - 2\,\zeta_6) \\
&\quad + \delta_{e\bar{e}} \mathbf{D}_\xi^0 \left[\mathbf{V}_{z_\eta}^0 \odot \zeta_1^0 - \mathbf{Y}^0 \odot \left(\zeta_2^0 - 2\,\zeta_6^0 \right) \right],
\end{aligned} \tag{B6}
$$

$$
\begin{aligned}
\mathbf{z}_\varphi &= -\mathbf{D}_\xi \otimes \left(\mathbf{V}_{z_\eta} \odot \zeta_6 + \mathbf{V}_{s_\eta} \odot \zeta_4 \right) \\
&\quad - \left(\mathbf{V}_{z_\xi} \odot \zeta_6 + \mathbf{V}_{s_\xi} \odot \zeta_4 \right) \otimes \mathbf{D}_\eta^{\mathrm{T}} + \mathbf{Y} \odot (\zeta_6 - 2\,\zeta_2) \\
&\quad + \delta_{e\bar{e}} \mathbf{D}_\xi^0 \left[-\mathbf{V}_{z_\eta}^0 \odot \zeta_6^0 + \mathbf{Y}^0 \odot \left(\zeta_6^0 - 2\,\zeta_2^0 \right) \right],
\end{aligned} \tag{B7}
$$

$$
\begin{aligned}
\mathbf{z}_z &= \mathbf{D}_\xi \otimes \left(\mathbf{V}_{z_\eta} \odot \zeta_5 + \mathbf{V}_{s_\eta} \odot \zeta_3 \right) + \left(\mathbf{V}_{z_\xi} \odot \zeta_5 + \mathbf{V}_{s_\xi} \odot \zeta_3 \right) \otimes \mathbf{D}_\eta^{\mathrm{T}} \\
&\quad - 2\mathbf{Y} \odot \zeta_4 + \delta_{e\bar{e}} \mathbf{D}_\xi^0 \left[\mathbf{V}_{z_\eta}^0 \odot \zeta_3^0 \right].
\end{aligned} \tag{B8}
$$