

# Generalized interferometry – I: theory for interstation correlations

Andreas Fichtner,<sup>1</sup> Laurent Stehly,<sup>2</sup> Laura Ermert<sup>1</sup> and Christian Boehm<sup>1</sup>

<sup>1</sup>Department of Earth Sciences, ETH Zurich, Switzerland. E-mail: [andreas.fichtner@erdw.ethz.ch](mailto:andreas.fichtner@erdw.ethz.ch)

<sup>2</sup>ISTerre, Université Joseph Fourier, Grenoble, France

Accepted 2016 November 7. Received 2016 November 3; in original form 2016 August 15

## SUMMARY

We develop a general theory for interferometry by correlation that (i) properly accounts for heterogeneously distributed sources of continuous or transient nature, (ii) fully incorporates any type of linear and nonlinear processing, such as one-bit normalization, spectral whitening and phase-weighted stacking, (iii) operates for any type of medium, including 3-D elastic, heterogeneous and attenuating media, (iv) enables the exploitation of complete correlation waveforms, including seemingly unphysical arrivals, and (v) unifies the earthquake-based two-station method and ambient noise correlations. Our central theme is not to equate interferometry with Green function retrieval, and to extract information directly from processed interstation correlations, regardless of their relation to the Green function. We demonstrate that processing transforms the actual wavefield sources and actual wave propagation physics into effective sources and effective wave propagation. This transformation is uniquely determined by the processing applied to the observed data, and can be easily computed. The effective forward model, that links effective sources and propagation to synthetic interstation correlations, may not be perfect. A forward modelling error, induced by processing, describes the extent to which processed correlations can actually be interpreted as proper correlations, that is, as resulting from some effective source and some effective wave propagation. The magnitude of the forward modelling error is controlled by the processing scheme and the temporal variability of the sources. Applying adjoint techniques to the effective forward model, we derive finite-frequency Fréchet kernels for the sources of the wavefield and Earth structure, that should be inverted jointly. The structure kernels depend on the sources of the wavefield and the processing scheme applied to the raw data. Therefore, both must be taken into account correctly in order to make accurate inferences on Earth structure. Not making any restrictive assumptions on the nature of the wavefield sources, our theory can be applied to earthquake and ambient noise data, either separately or combined. This allows us (i) to locate earthquakes using interstation correlations and without knowledge of the origin time, (ii) to unify the earthquake-based two-station method and noise correlations without the need to exclude either of the two data types, and (iii) to eliminate the requirement to remove earthquake signals from noise recordings prior to the computation of correlation functions. In addition to the basic theory for acoustic wavefields, we present numerical examples for 2-D media, an extension to the most general viscoelastic case, and a method for the design of optimal processing schemes that eliminate the forward modelling error completely. This work is intended to provide a comprehensive theoretical foundation of full-waveform interferometry by correlation, and to suggest improvements to current passive monitoring methods.

**Key words:** Inverse theory; Interferometry; Seismic tomography; Computational seismology; Theoretical seismology.

## 1 INTRODUCTION

Seismic waves in the Earth are continuously excited. Their sources range from widely distributed, long-lasting ocean waves to spatially localized explosions and earthquakes of short duration. While classical seismology developed largely around the study of transient signals, it has long been known that valuable information may be contained in the space-time coherence of the low-amplitude, quasi-random ambient field; also referred to as ambient noise or microtremors (Aki 1957). Early attempts to extract this information were not always successful.

Inspired by the autocorrelation method of Claerbout (1968), Cole (1995) computed noise correlations for an array of 4056 geophones deployed on the Stanford campus. However, no coherent signals could be found, possibly because the recording length was too short and the instrumental noise too strong.

With improvements in station quality and quantity, interferometry based on interstation correlations of ambient noise has become a standard tool, complementing techniques based on earthquakes or explosions. In seismic exploration, interferometry is a low-cost alternative to active-source imaging and time-dependent monitoring (e.g. Bussat & Kugler 2011; Mordret *et al.* 2013, 2014; de Ridder *et al.* 2014; de Ridder & Biondi 2015). In regional crustal studies (e.g. Sabra *et al.* 2005; Shapiro *et al.* 2005; Stehly *et al.* 2009), and continental-scale tomography (e.g. Lin *et al.* 2008; Zheng *et al.* 2011; Verbeke *et al.* 2012; Saygin & Kennett 2012), interferometry provides data sets with short propagation distances that allow tomographers to fit higher-frequency recordings, and thereby achieve higher resolution. Imaging of the global upper mantle (e.g. Nishida & Montagner 2009; Haned *et al.* 2016), internal discontinuities (e.g. Poli *et al.* 2012; Boué *et al.* 2013; Poli *et al.* 2015), and the inner core (e.g. Lin *et al.* 2013; Huang *et al.* 2015) equally benefits from noise interferometry, as it increases coverage in regions that are difficult to illuminate with earthquake data.

### 1.1 Theoretical foundations of Green function recovery

Interferometry by correlation rests on the approximation of an empirical Green function by the cross-correlation of a diffuse wavefield recorded at a pair of receivers. Having been demonstrated in lab experiments (Malcolm *et al.* 2004), theoretical justifications for Green function recovery have been proposed for models with variable degrees of approximation to the true physics of wave propagation in the complex Earth.

For closed systems, Lobkis & Weaver (2001) demonstrated that the interstation correlation of a wavefield with equipartitioned, that is, random and uncorrelated modes, is proportional to the interstation Green function. While theoretically appealing, the normal-mode argument has limited applicability to the Earth because heterogeneously distributed sources – in the form of noise sources, earthquakes, or scatterers – prevent the ambient field from being equipartitioned across the complete seismically observable frequency band (e.g. Nawa *et al.* 1998; Hillers *et al.* 2012; Liu & Ben-Zion 2013; Nishida 2013).

For open, that is, spatially unbounded, systems, representation theorems can be used to justify Green function recovery (Wapenaar 2004; Wapenaar & Fokkema 2006). In addition to the absence of attenuation, a homogeneous distribution of uncorrelated sources on a closed surface is required in the representation theorem approach. Furthermore, sufficiently smooth Earth structure is needed to extract material properties from under the representation theorem integrals and to obtain expressions useful for noise correlation.

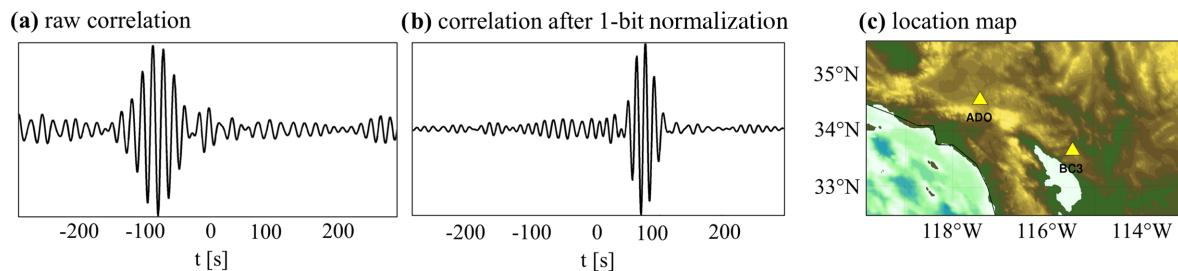
The most simplified but also most physically insightful approach uses a model of plane wave propagation in a 2-D, isotropic, acoustic and unbounded medium without attenuation (e.g. Tsai 2009; Boschi *et al.* 2013). The Green function is recovered when the incident plane waves arrive with equal strength from all azimuths with randomly distributed and statistically independent amplitudes, that is, when the wavefield is perfectly equipartitioned. While being over-simplified, the plane wave model can be used to understand the effects on amplitudes and traveltimes of the empirical Green function when the noise source distribution is not homogeneous (e.g. Yao & van der Hilst 2009; Harmon *et al.* 2010).

Realizing that the stringent requirements for Green function recovery are difficult to meet, various authors derived approximate relations between correlations and Green functions that hold under more relaxed conditions. Using a linearized version of the ray and stationary phase approximations, Godin (2009) suggested that the Green function may also be estimated when sources are correlated. Tsai (2010) derived relations between correlations and Green functions for closed systems without attenuation when the modes are not equipartitioned. That sources, for example, in the form of scatterers, are only needed in the stationary phase region was shown by Snieder (2004) assuming wave propagation in homogeneous media without attenuation. While providing valuable insight into the physics of Green function recovery, these approaches themselves introduce new assumptions and approximations, the quality of which may be difficult to assess in real-data applications.

The effects of attenuation and heterogeneous, non-stationary sources on Green function retrieval have been studied extensively. Depending on the specific situation, they include amplitude and traveltime biases (Tsai 2009; Yao & van der Hilst 2009; Froment *et al.* 2010; Tsai 2011; Zhan *et al.* 2013; Fichtner 2014; Yanovskaya *et al.* 2016), the appearance of spurious arrivals (Halliday & Curtis 2008, 2009; Kimman & Trampert 2010), and the absence of seismic phases such as body- or higher-mode surface waves. Approaches to correct for systematic traveltime errors have been developed, but also rely on strong simplifying assumptions, for example, that seismic waves are 2-D plane waves (Yao & van der Hilst 2009) or that the phase of noise correlations is not affected by sources that vary in space and time (Daskalakis *et al.* 2016).

### 1.2 Data processing

Noticing that the seismic noise field is ‘polluted’ by earthquake signals and excited by spatially heterogeneous, non-stationary sources, numerous processing and stacking schemes have been proposed in order to improve Green function recovery. These include the averaging of causal and acausal correlation branches, spectral whitening, time-domain running averages, and frequency-domain normalization (e.g. Bensen *et al.* 2007; Groos *et al.* 2012), as well as one-bit normalization (e.g. Larose *et al.* 2004; Shapiro & Campillo 2004; Cupillard *et al.* 2011; Hanasoge & Branicki 2013), phase-weighted stacks based on the Hilbert transform (Schimmel & Paulssen 1997; Schimmel *et al.* 2011) or the



**Figure 1.** Modification of the correlation function between the Californian stations ADO and BC3. (a) Raw correlation function computed without pre-processing of the individual recordings. (b) Correlation function computed after one-bit normalization of the recordings. The change in symmetry is likely due to the continuous recording of low-amplitude waves originating in the Atlantic that contribute to the one-bit normalized correlations but not much to the raw correlations. Energetic sources of short duration located nearby the Pacific coastline contribute to raw correlations and not to one-bit correlations. (c) Location map of stations ADO and BC3. [CR] (see Section 1.4 for the definition of the reproducibility indicators ‘easily reproducible’ [ER], ‘conditionally reproducible’ [CR] and ‘not reproducible’ [NR]).

S transform (Baig *et al.* 2009), directional balancing (Curtis & Halliday 2010), Welch’s method of overlapping time windows (Welch 1967; Seats *et al.* 2012), the application of curvelet denoising filters (Stehly *et al.* 2011), or a sequence of selection and noise suppression filters (Nakata *et al.* 2015).

During the past decade, experience has shown that the optimal processing is dependent on both the nature of a specific data set and on the wave type that one wishes to extract. In most applications, a quantitative measure for the goodness of the processing sequence is difficult to obtain because the reference, that is, the actual Green functions between *all* receiver pairs, is generally unknown. As a consequence, processing is to some extent subjective, and often guided by the need to obtain stable measurements, and to maximize some measure of signal-to-noise ratio or cross-correlation symmetry. Differences in processing can lead to significant differences in the correlation functions (Bensen *et al.* 2007). These differences leave an unavoidable imprint on the sensitivity to Earth structure and noise sources (Fichtner 2014; Stehly & Cupillard 2016). An example of how processing can modify the correlation function is presented in Fig. 1.

### 1.3 Problem statement and motivation

The failure to fully meet all requirements for Green function retrieval, combined with semi-subjective processing has a series of negative effects on our ability to infer Earth structure using seismic interferometry:

Systematic traveltimes and amplitude errors in particular, may lead to incorrect inferences of seismic velocities, anisotropy and attenuation (Harmon *et al.* 2010; Stehly & Cupillard 2016). While these effects may be small in the case of large velocity heterogeneities inferred from traveltimes, they can become significant or even dominant in monitoring applications where small velocity variations must be resolved over short time scales (e.g. Zhan *et al.* 2013; Delaney *et al.* 2016).

Since the details of correlation waveforms depend strongly on both the source characteristics and the processing, modern tomographic techniques that exploit these details for the benefit of improved resolution cannot be applied reliably. These techniques include finite-frequency tomography for surface waves (e.g. Yomogida 1992; Friederich 2003; Yoshizawa & Kennett 2004; Zhou *et al.* 2004), for body waves (e.g. Dahlen *et al.* 2000; Dahlen & Baig 2002; Sigloch *et al.* 2008), and full-waveform inversion (e.g. Chen *et al.* 2007; Fichtner *et al.* 2009; Tape *et al.* 2010; Fichtner *et al.* 2013b).

As a consequence of heterogeneously distributed noise sources, correlation functions may not resemble plausible Green functions. Properly converged correlation functions that carry valuable information on Earth structure may thus need to be excluded from tomographic inversions that assume nearly perfect Green function recovery. This may affect up to 70 per cent of a correlation data set (e.g. Stehly *et al.* 2009).

Finally, reproducibility suffers from the unavoidable subjectivity of the processing schemes that are mostly based on experience and intuition. This aspect becomes increasingly relevant as our Earth models become more and more detailed, and as we extract more and more information from them.

### 1.4 Objectives and outline

This work attempts to address these shortcomings and is guided by the following objectives: (i) Provide a unified theory for interferometry by correlation that accounts for heterogeneous noise sources, linear and non-linear processing, and the presence of transient sources such as earthquakes. (ii) Enable the exploitation of seemingly unphysical correlation waveforms that result from heterogeneous noise sources and that are not present in the actual interstation Green function. (iii) Bridge the gap between interstation correlations of ambient noise (noise interferometry) and earthquake recordings (two-station method). (iv) Function for any type of medium, including a realistic elastic Earth with attenuation and anisotropy. (v) Establish the theoretical foundation of interferometric full-waveform inversion for sources and Earth structure.

The central theme of this work is not to equate interferometry with Green function retrieval, not even attempting to approximate a Green function through the correlation of noise. Instead, we will extract information on wavefield sources and Earth structure directly from the interstation correlation; regardless of its proximity to or distance from the Green function.

Ideas going in a similar direction have been presented before. Schuster *et al.* (2004) propose to image with correlations, not explicitly mentioning their relation to a Green function, but also without analysis of potentially important cross terms. Later, van Driel *et al.* (2015) modelled correlation functions by simulating random wavefields excited by arbitrary distributions of sources without invoking Green function recovery. Cupillard (2008) constructed a surface of effective sources that mimick a heterogeneous distribution of noise sources. In his pioneering work, Woodard (1997) developed a deterministic representation of the interstation correlation that only requires the power-spectral density distribution of the noise sources. This concept was used extensively in helioseismology (e.g. Gizon & Birch 2002; Duvall *et al.* 2006; Hanasoge *et al.* 2011) and adopted to terrestrial seismology by Tromp *et al.* (2010). We will most closely follow Woodard's approach, and extend it to account for data processing typically applied in seismic applications.

This paper is organized as follows: In Section 2, we develop a forward modelling theory for raw and processed interstation correlations. We demonstrate that processing leads to an effective wave propagation and to effective sources that generally differ from the true physical wave propagation and the true sources. The effective theory for processed correlations may suffer from forward modelling errors that are controlled by the type of processing applied to the raw data. These new concepts of processing-induced effectiveness and modelling errors are illustrated with numerical examples in Section 3. Section 4 is dedicated to the formulation of inverse problems for Earth structure and wavefield sources. We derive sensitivity kernels for raw and processed correlations, and we demonstrate that processing must be taken into account in order to ensure accurate inferences on Earth structure and noise sources. Building on the developments for scalar, acoustic waves, Section 5 offers a generalization to vectorial, elastic waves. In Section 6, we present inverse theoretical interpretations of our method, and we make suggestions for optimal processing schemes that eliminate any forward modelling errors in the effective theory.

We complement all our developments with numerical examples for 2-D, acoustic and unbounded media. This simplification does not arise from a limitation of the theory, but from our decision to provide physically insightful and fully reproducible illustrations of basic concepts. Further generalizations and real-data applications will appear in later publications.

To indicate the degree of reproducibility, we adopt the codex introduced by the Stanford Exploration Project. Figures labelled [ER] are easily reproducible using codes and input files provided in the Supporting Information. Conditional reproducibility [CR] indicates that larger computational resources or specific data are needed. Figures that are not reproducible, for example, schematic illustrations, are labelled [NR].

## 2 EFFECTIVE THEORY FOR MODELLING INTERSTATION CORRELATIONS - THE ACOUSTIC CASE

### 2.1 Fundamental theories

In the interest of simplicity, we start our development with the acoustic case where wave propagation is governed by the scalar wave equation

$$\mathcal{L}_x u = -\omega^2 \frac{1}{\kappa} u - \nabla \cdot \left( \frac{1}{\rho} \nabla u \right) = s, \quad (1)$$

with the acoustic wave operator  $\mathcal{L}_x$  that differentiates with respect to the spatial variable  $\mathbf{x}$ . Eq. (1) links the bulk modulus  $\kappa(\mathbf{x}, \omega)$ , the mass density  $\rho(\mathbf{x})$ , and the external source  $s(\mathbf{x}, \omega)$  to the acoustic pressure field  $u(\mathbf{x}, \omega)$ . We depart from the more conventional symbol for pressure,  $p(\mathbf{x}, \omega)$ , to facilitate the transition to the elastic case in Section 5. The bulk modulus is allowed to be frequency-dependent and complex-valued to account for viscoacoustic dissipation.

In our developments, the acoustic wave equation does not play the role of an approximation to more complex elastic wave propagation, that may suffer from unmodelled shear waves, instability or unphysical anisotropy (e.g. Alkhalifah 2000; Barnes & Charara 2008; Fletcher *et al.* 2008; Operto *et al.* 2009; Marelli *et al.* 2012; Cance & Capdeville 2015). Instead, we assume that the medium is truly acoustic, meaning that observational evidence for elastic effects is absent. This implies that we can find  $\rho$ ,  $\kappa$ , and  $s$  such that *any* observed wavefield  ${}^o u$  can be matched by a synthetic wavefield  $u$  computed with eq. (1) to within the observational uncertainties, that is,

$${}^o u(\mathbf{x}, \omega) = u(\mathbf{x}, \omega) + \varepsilon(\mathbf{x}, \omega), \quad (2)$$

with a hopefully small error  $\varepsilon$ . In the following, we will classify a forward modelling theory for which there is no observational evidence of its inaccuracy as a *fundamental theory*; fundamental in the sense of providing the basis for further developments.

### 2.2 Raw correlations

Interferograms for raw, that is, unprocessed, data can be computed by correlating frequency-domain wavefields  ${}^o u(\mathbf{x}_i, \omega)$ , observed at positions  $\mathbf{x}_i$ . The superscript  $o$  denotes observations, and the subscript  $n$  refers to a time interval  $[t_n, t_{n+1}]$  for which the Fourier transform has been computed. To avoid contamination of the Fourier spectrum by the finiteness of the time window, we assume that  $(t_{n+1} - t_n)^{-1}$  is much smaller

than the minimum frequency of interest. From the correlation of recordings at positions  $\mathbf{x}_i$  and  $\mathbf{x}_k$  we compute the observed interferogram, or correlation function, for time interval  $n$

$${}^oI(\mathbf{x}_i, \mathbf{x}_k) = {}_n\mu(\mathbf{x}_i) {}_n\mu^*(\mathbf{x}_k), \quad (3)$$

where we omitted  $\omega$  for a condensed notation. We use the terms ‘interferogram’ and ‘correlation’ interchangeably, noticing that the former is more general, comprising more than correlation. Taking the average over all intervals  $n = 1, \dots, N$ , we obtain the ensemble interferogram,

$${}^oI(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N {}_n\mu(\mathbf{x}_i) {}_n\mu^*(\mathbf{x}_k). \quad (4)$$

The omission of the subscript  $_n$  symbolizes averaging. To model observed interferograms, we start with the frequency-domain representation theorem (e.g. Kennett 2001; Aki & Richards 2002)

$${}_n\mu(\mathbf{x}_i) = \int_{\oplus} G(\mathbf{m}; \mathbf{x}_i, \mathbf{x}) {}_nF(\mathbf{x}) d\mathbf{x}. \quad (5)$$

It translates a source  ${}_nF(\mathbf{x})$  into a synthetic wavefield  ${}_n\mu(\mathbf{x}_i)$  via the Green function  $G(\mathbf{m}; \mathbf{x}_i, \mathbf{x})$ , which depends on the structural model  $\mathbf{m} = (\rho, \kappa)$ . The integration in eq. (5) is over the entire model volume  $\oplus$ . The source  ${}_nF(\mathbf{x})$  may comprise any combination of quasi-random, deterministic, point-localized and distributed sources. This includes sources of ambient noise, earthquakes and explosions. Using eq. (5), we compute the ensemble synthetic interferogram by cross-correlation and subsequent summation:

$$I(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N {}_nI(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N {}_n\mu(\mathbf{x}_i) {}_n\mu^*(\mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') {}_nF(\mathbf{x}) {}_nF^*(\mathbf{x}') d\mathbf{x} d\mathbf{x}'. \quad (6)$$

The omission of the superscript  $o$  in  $I(\mathbf{x}_i, \mathbf{x}_k)$  indicates synthetics, in contrast to observations  ${}^oI(\mathbf{x}_i, \mathbf{x}_k)$ . Substituting the power-spectral density (psd) of the sources, that is, their average correlation between positions  $\mathbf{x}$  and  $\mathbf{x}'$ ,

$$S(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_nS(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_nF(\mathbf{x}) {}_nF^*(\mathbf{x}'), \quad (7)$$

yields the forward modelling equation for unprocessed synthetic interstation correlations,

$$\underbrace{I(\mathbf{x}_i, \mathbf{x}_k)}_{\text{interf. wavefield}} = \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) \underbrace{\left[ \int_{\oplus} G^*(\mathbf{x}_k, \mathbf{x}') S(\mathbf{x}, \mathbf{x}') d\mathbf{x}' \right]}_{\text{source of interf. wavefield}} d\mathbf{x}. \quad (8)$$

Interpreting  $\mathbf{x}_i$  as a free space variable, eq. (8) can be regarded as a representation theorem where a source  $\int G^*(\mathbf{x}_k, \mathbf{x}') S(\mathbf{x}, \mathbf{x}') d\mathbf{x}'$  drives the propagation of an *interferometric wavefield*  $I(\mathbf{x}_i, \mathbf{x}_k)$  via the Green function  $G(\mathbf{x}_i, \mathbf{x})$ . Should the sources  ${}_nF$  at neighbouring positions  $\mathbf{x}$  and  $\mathbf{x}'$  be on average temporally uncorrelated in the sense

$$\frac{1}{N} \sum_{n=1}^N {}_nF(\mathbf{x}) {}_nF^*(\mathbf{x}') = S(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}'), \quad (9)$$

the forward modelling eq. (8) further reduces to

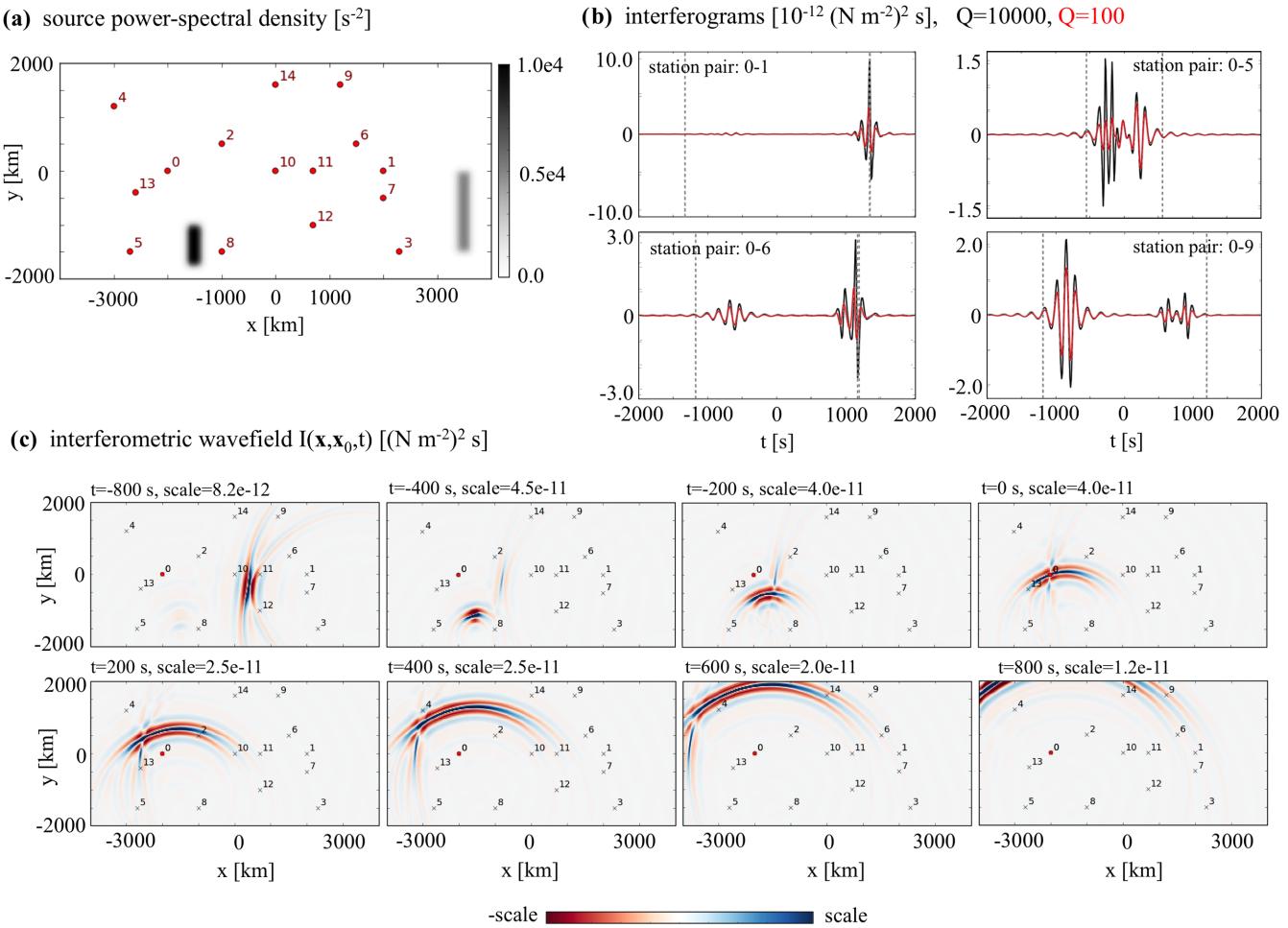
$$\underbrace{I(\mathbf{x}_i, \mathbf{x}_k)}_{\text{interf. wavefield}} = \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) \underbrace{[G^*(\mathbf{x}_k, \mathbf{x}) S(\mathbf{x})]}_{\text{source of interf. wavefield}} d\mathbf{x}. \quad (10)$$

The approximation (9) is frequently made to reduce computational costs (e.g. Woodard 1997; Tromp *et al.* 2010; Hanasoge *et al.* 2011; Basini *et al.* 2013; Hanasoge 2013; Fichtner 2014; Nishida 2014). Its quality needs to be assessed on a case-by-case basis.

The derivation of the forward modelling eq. (8) did not require any additional assumptions. Thus, when the acoustic wave eq. (1) constitutes a fundamental theory for the modelling of the wavefield itself, eq. (8) must be a fundamental theory for the forward modelling of interstation correlations.

### 2.2.1 Computability

A key element of eqs (8) and (10) is computability: Only the psd of the sources, that is, a deterministic quantity, is required to compute the synthetic correlation function. The forward problem can thus be solved without simulating long-duration realizations of random wavefields. While this may be possible in normal-mode approaches for spherically symmetric Earth models (e.g. Cupillard & Capdeville 2010; Gualtieri *et al.* 2013; Nishida 2014), it would be prohibitively expensive for modern 3-D time-domain wave equation solvers that fully account for lateral heterogeneities (e.g. Moczo *et al.* 2002; de la Puente *et al.* 2007; Peter *et al.* 2011; Gokhberg & Fichtner 2016). Stehly *et al.* (2011) estimated that the seismogram length needed to retrieve high-quality Green functions over 720 km distance in the period range 20–40 s is on



**Figure 2.** Illustration of the synthetic interferometric wavefield  $I(\mathbf{x}, \mathbf{x}_0)$  with the reference receiver at position  $\mathbf{x}_0$ . The medium is 2-D homogeneous with velocity  $v = 3000 \text{ m s}^{-1}$ , and density  $\rho = 3000 \text{ kg m}^{-3}$ . (a) Distribution of the frequency-domain source psd  $S(\mathbf{x})$ . The positions of receivers, used also in later examples, are marked by red dots. This distribution is the same for all frequencies. (b) Time-domain interferograms for receiver pairs 0–1, 0–5, 0–6 and 0–9 for  $Q = 10\,000$  (black) and  $Q = 100$  (red). Grey dashed lines indicate the arrival times of ballistic waves travelling between the receivers. The heterogeneous source distribution causes some wave packets to arrive before the ballistic wave. (c) Snapshots of the interferometric wavefield at different times for  $Q = 10\,000$  and the reference receiver 0 at position  $\mathbf{x}_0$ . At  $t = -800 \text{ s}$ , a low-amplitude part of the wavefield has just passed receiver 9, generating the complicated wave packet on the acausal side seen in panel (b) for pair 0–9. At  $t = 0$ , the wavefield focuses near receiver 0. Around  $t = 800 \text{ s}$ , the wavefield again passes by receiver 9, generating the wave packet on the causal side of the interferogram 0–9 shown in panel (b). [ER]

the order of 100 days. Roughly estimating that explicit time integration schemes used in numerical wave propagation require  $\sim 10$  time steps per minimum period, this translates to around 5 million time steps. This number challenges currently available computational resources and even advanced time stepping schemes (e.g. Bogey & Bailly 2004; Dumbser *et al.* 2007; Wang & Xu 2015).

### 2.2.2 Illustration

We illustrate the computation of the synthetic interferometric wavefield  $I(\mathbf{x}_i, \mathbf{x}_k)$  for the case of an unbounded, homogeneous medium. In the interest of a physically insightful and easily reproducible illustration, we assume line sources extending infinitely in  $z$ -direction, thereby making the problem effectively 2-D. The corresponding Green function is given by

$$G(\mathbf{x}_i, \mathbf{x}) = -i \frac{1}{4\rho v^2} \sqrt{\frac{2v}{\pi\omega r}} e^{-i\frac{\omega}{v}r} e^{-\frac{\omega r}{2vQ}} e^{i\frac{\pi}{4}}, \quad (11)$$

with the phase velocity  $v = \sqrt{\text{Re } \kappa/\rho}$ , the source-receiver distance  $r = |\mathbf{x} - \mathbf{x}_i|$ , and the quality factor  $Q$  (e.g. Båth 1968; Igel 2016). Fig. 2 shows interferometric wavefields computed using eq. (10) with  $Q = 10\,000$  and  $Q = 100$ . The source psd is heterogeneous, leading to asymmetric correlations with wave packets arriving earlier than the ballistic wave travelling between the receivers. Note that the amplitudes of different parts of the interferograms react differently to changes in  $Q$ .

### 2.3 Effective processing and the transfer coefficient

In practice, the computation of interferograms by cross-correlation is less straightforward than suggested in Section 2.2. Processing is generally applied to make correlations more suitable for inversion, that is, facilitate the extraction of information on Earth structure and wavefield sources. It transforms the observed *raw interferogram* for time interval  $n$ ,  ${}^nI(\mathbf{x}_i, \mathbf{x}_k)$ , into the observed *processed interferogram*,  ${}^n\tilde{I}(\mathbf{x}_i, \mathbf{x}_k)$ . The two are related by the frequency-dependent *transfer coefficient*  ${}_nT_{ik}$  that is implicitly defined by,

$${}^n\tilde{I}(\mathbf{x}_i, \mathbf{x}_k) = {}_nT_{ik} {}^nI(\mathbf{x}_i, \mathbf{x}_k). \quad (12)$$

As can be seen from eq. (12), the transfer coefficient can always be computed in the frequency domain by dividing the processed interferogram  ${}^n\tilde{I}(\mathbf{x}_i, \mathbf{x}_k)$  by the raw interferogram  ${}^nI(\mathbf{x}_i, \mathbf{x}_k)$ . This operation may require some regularization to prevent division by zero. Cases where  ${}^nI(\mathbf{x}_i, \mathbf{x}_k)$  is practically zero are further discussed in Section 6.4. The transfer coefficient constitutes an *effective processing*. It encapsulates the complete processing chain, including filtering, normalizing, weighting, and any other operation that is performed on the raw data per time interval. As will be discussed in more detail in Section 6.2, the transfer coefficient constitutes a specific evaluation of the generally nonlinear transfer function for a specific piece of raw data.

To gain intuitive understanding of the transfer coefficient, prior to a rigorous analysis that will follow later, we factorize  ${}_nT_{ik}$  as

$${}_nT_{ik} = {}_nf g_{ik} + {}_ne_{ik}. \quad (13)$$

The factorization (13) is always mathematically possible, and it can be found such that the factorization residual  ${}_ne_{ik}$  is optimally small (see Appendix A1). For a fixed station pair,  $(\mathbf{x}_i, \mathbf{x}_k)$ , the factor  $g_{ik}$  encapsulates effective processing applied to any time interval  $n$ . It is, by design, independent of the time window index  $n$  and entirely wave path-specific. Thus, intuitively, it comprises processing done to account or correct for propagation effects along the wave path  $\mathbf{x}_i \leftrightarrow \mathbf{x}_k$ .

The factor  ${}_nf$  contains that part of the effective processing that is applied to all receiver pairs in time interval  $n$ . Thus,  ${}_nf$  does not depend on a specific wave path, but only on the wavefield sources. It therefore accounts or corrects for source effects.

Finally, the residual  ${}_ne_{ik}$  in the factorization (13) is specific for each time interval  $n$  and receiver pair  $(\mathbf{x}_i, \mathbf{x}_k)$ . It is that part of the effective processing that cannot be decomposed into a propagation effect (covered by  $g_{ik}$ ) and a source effect (covered by  ${}_nf$ ).

While the factorization of the transfer coefficient helps at this stage to understand its physical meaning, its true value will become apparent in the formulation of an effective forward modelling theory in Section 2.5. In Appendix A1, we show how the factorization (13) can be computed efficiently such that the time-averaged factorization residual equals 0,

$$\frac{1}{N} \sum_{n=1}^N {}_ne_{ik} = 0, \quad \text{for all } (i, k), \quad (14)$$

and the time-averaged source factor is equal to 1,

$$\frac{1}{N} \sum_{n=1}^N {}_nf = 1. \quad (15)$$

In Sections 2.5.2 and 2.5.4, we will provide physics-based justifications for these specific choices.

### 2.4 Admissible effective theories

The value of a fundamental theory defined in Section 2.1 is that a structure-source model can be found such that any observed wavefield can be matched by a synthetically computed wavefield to within the observational uncertainties. In Section 2.2, we concluded that eq. (8) constitutes a fundamental theory for the modelling of unprocessed correlations, because no additional assumptions were needed for its derivation that went beyond the acoustic wave eq. (1). Thus, we can find triples of bulk modulus, mass density, and psd,  $\mathbf{p}_{\text{fit}} = (\kappa_{\text{fit}}, \rho_{\text{fit}}, S_{\text{fit}})$  such that the raw synthetic interferogram  ${}_nI(\mathbf{x}_i, \mathbf{x}_k; \mathbf{p}_{\text{fit}})$  matches the raw observed interferogram  ${}^nI(\mathbf{x}_i, \mathbf{x}_k)$  to within the observational uncertainties,

$${}^nI(\mathbf{x}_i, \mathbf{x}_k) = {}_nI(\mathbf{x}_i, \mathbf{x}_k; \mathbf{p}_{\text{fit}}) + {}_n\epsilon_{ik}. \quad (16)$$

Due to non-uniqueness, there will generally be infinitely many  $\mathbf{p}_{\text{fit}}$  that satisfy eq. (16). We denote their ensemble, the space of data-fitting models, by  $\mathbb{P}_{\text{fit}}$ . To ensure meaningful inverse problem solutions, we must request that an *admissible* effective theory for processed interferograms must be a fundamental theory as well. This means, in particular, that the triples  $\mathbf{p} = (\kappa, \rho, S)$  that satisfy (16) must also satisfy the corresponding equation for processed interferograms,

$${}^n\tilde{I}(\mathbf{x}_i, \mathbf{x}_k) = {}^n\tilde{I}(\mathbf{x}_i, \mathbf{x}_k; \mathbf{p}) + {}^n\tilde{\epsilon}_{ik}, \quad \text{for all } \mathbf{p} \in \mathbb{P}_{\text{fit}}. \quad (17)$$

If the admissibility condition is not met, the matching of processed and raw interferograms unavoidably leads to conflicting inferences on the structure of the Earth and the sources of the wavefield. From the comparison of eqs (16) and (17), we can derive an admissible effective theory for processed interferograms, that is, a definition of  ${}^n\tilde{I}(\mathbf{x}_i, \mathbf{x}_k)$ . Multiplying (16) by the transfer coefficient  ${}_nT_{ik}$  we obtain

$${}^n\tilde{I}(\mathbf{x}_i, \mathbf{x}_k) = {}_nT_{ik} {}^nI(\mathbf{x}_i, \mathbf{x}_k) = {}_nT_{ik} {}_nI(\mathbf{x}_i, \mathbf{x}_k; \mathbf{p}_{\text{fit}}) + {}_nT_{ik} {}_n\epsilon_{ik}. \quad (18)$$

It follows that the effective forward problem

$$\tilde{I}(\mathbf{x}_i, \mathbf{x}_k) = {}_n T_{ik} {}_n I(\mathbf{x}_i, \mathbf{x}_k), \quad \text{for any } \mathbf{p}, \quad (19)$$

is able to match the processed observed interferogram  $\tilde{I}(\mathbf{x}_i, \mathbf{x}_k)$  to within the processed observational errors  $\tilde{\epsilon}_{ik} = {}_n T_{ik} {}_n \epsilon_{ik}$ , meaning that it is admissible. In the following we will use eq. (19) to develop a computable forward modelling theory for processed interferograms. In Section 6.2, we provide an extended discussion on the transfer coefficient and its role in the definition of an effective theory.

## 2.5 Effective theory for the modelling of processed interferograms

In deriving an effective forward modelling theory for processed interferograms, we exploit the factorization of the transfer coefficient, introduced in eq. (13). Combining eq. (6) for the modelling of unprocessed interferograms with eq. (19) and the factorization  ${}_n T_{ik} = {}_n f g_{ik} + {}_n e_{ik}$ , yields

$$\tilde{I}(\mathbf{x}_i, \mathbf{x}_k) = ({}_n f g_{ik} + {}_n e_{ik}) \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') {}_n F(\mathbf{x}) {}_n F^*(\mathbf{x}') d\mathbf{x} d\mathbf{x}'. \quad (20)$$

Taking the average over time intervals  $n$ , yields the processed synthetic ensemble interferogram:

$$\begin{aligned} \tilde{I}(\mathbf{x}_i, \mathbf{x}_k) &= \int_{\oplus} \int_{\oplus} g_{ik} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') \left[ \frac{1}{N} \sum_{n=1}^N {}_n f {}_n F(\mathbf{x}) {}_n F^*(\mathbf{x}') \right] d\mathbf{x} d\mathbf{x}' \\ &+ \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') \left[ \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} {}_n F(\mathbf{x}) {}_n F^*(\mathbf{x}') \right] d\mathbf{x} d\mathbf{x}'. \end{aligned} \quad (21)$$

The first term in eq. (21) resembles eq. (6) for the unprocessed synthetic interferogram. The only differences are the appearances of  ${}_n f$  and  $g_{ik}$ . In the second part of eq. (21), the factorization residual  ${}_n e_{ik}$  depends on both the time interval  $n$  and the receiver pair indices  $(i, k)$ . Defining, in analogy to eq. (7), the *effective power-spectral density* of the sources

$$\widehat{S}(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_n \widehat{S}(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_n f {}_n S(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_n f {}_n F(\mathbf{x}) {}_n F^*(\mathbf{x}'), \quad (22)$$

and the *forward modelling error*

$$E(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') \left[ \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} {}_n S(\mathbf{x}, \mathbf{x}') \right] d\mathbf{x} d\mathbf{x}', \quad (23)$$

we obtain a simplified version of eq. (21):

$$\begin{aligned} \underbrace{\tilde{I}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{proc. interf. wavefield}} &= \int_{\oplus} \underbrace{g_{ik} G(\mathbf{x}_i, \mathbf{x})}_{\text{eff. Green function}} \underbrace{\left[ \int_{\oplus} G^*(\mathbf{x}_k, \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x}' \right]}_{\text{effective source}} d\mathbf{x} \\ &+ \underbrace{E(\mathbf{x}_i, \mathbf{x}_k)}_{\text{forward modelling error}} = \underbrace{\widehat{I}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{effective interf. wavefield}} + E(\mathbf{x}_i, \mathbf{x}_k). \end{aligned} \quad (24)$$

The effective interferometric wavefield  $\widehat{I}(\mathbf{x}_i, \mathbf{x}_k)$  constitutes, by construction, an admissible and computable effective theory for the modelling of processed correlations. Similar to eq. (8), eq. (24) can be regarded as a representation theorem for an interferometric wavefield where the receiver position  $\mathbf{x}_i$  is a free variable. The effective source  $\int G^*(\mathbf{x}_k, \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x}'$  drives the propagation of the *processed interferometric wavefield*  $\tilde{I}(\mathbf{x}_i, \mathbf{x}_k)$  via the *effective Green function*  $g_{ik} G(\mathbf{x}_i, \mathbf{x})$ . In the specific case where sources are uncorrelated in the sense of  $\widehat{S}(\mathbf{x}, \mathbf{x}') = \widehat{S}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}')$ , (e.g. Woodard 1997; Tromp *et al.* 2010; Hanasoge *et al.* 2011; Hanasoge 2013; Fichtner 2014; Nishida 2014) we can eliminate one of the volume integrals in the effective interferogram, which yields

$$\underbrace{\tilde{I}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{proc. interf. wavefield}} = \int_{\oplus} \underbrace{g_{ik} G(\mathbf{x}_i, \mathbf{x})}_{\text{eff. Green function}} \underbrace{[G^*(\mathbf{x}_k, \mathbf{x}) \widehat{S}(\mathbf{x})]}_{\text{effective source}} d\mathbf{x} + \underbrace{E(\mathbf{x}_i, \mathbf{x}_k)}_{\text{forward modelling error}} = \underbrace{\widehat{I}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{effective interf. wavefield}} + E(\mathbf{x}_i, \mathbf{x}_k). \quad (25)$$

Since eqs (22) to (24) are central to this work, we provide a detailed interpretation of their meaning in terms of both wave propagation physics and inverse problem solutions in the following.

### 2.5.1 Interpretation of the effective theory

The factor  $g_{ik}$  plays the role of a *propagation correction term* that modifies the propagation of the unprocessed interferometric wavefield through a multiplication with the Green function. It may thus be interpreted as describing wave propagation through an effective medium ‘seen’ by the processed interferogram. The *source correction factor*  ${}_n f$  is a correction term to the source for time interval  $n$ . The presence of  ${}_n f$  transforms the actual psd  $S$  into the effective psd  $\widehat{S}$  that excites the propagation of the processed interferometric wavefield.

The first part of eqs (24) and (25) represents a computable theory that is formally equivalent to the one from eq. (8) for unprocessed correlation functions. The effect of processing is reflected in an effective source and an effective propagation. The forward modelling error  $E(\mathbf{x}_i, \mathbf{x}_k)$  describes the extent to which processed correlations cannot be interpreted as correlations with effective source and effective propagation. It plays the role of a systematic error in the forward theory that must be taken into account in the solution of an inverse problem (see Section 4).

Forward modelling errors are the result of processing. While they could in theory be avoided by not applying any processing, this is hardly possible in practice. Processing is a means to emphasize specific aspects of the data, thereby facilitating the solution of inverse problems. For instance, processing schemes used to reduce the impact of large earthquake signals emphasize waves travelling between receivers. Several limiting cases of forward modelling errors deserve a more detailed discussion:

(i) *The perfect effective theory.* When a factorization  ${}_n T_{ik} = {}_n f g_{ik} + {}_n e_{ik}$  can be found such that the forward modelling error  $E(\mathbf{x}_i, \mathbf{x}_k)$  vanishes, the effective theory is perfect. This means that we can solve an effective forward problem (24) that has exactly the same functional shape as the original forward problem (8) without suffering from any forward modelling errors. In particular, this implies that the processed interferogram  $\widetilde{I}(\mathbf{x}_i, \mathbf{x}_k)$  can indeed be interpreted as a proper correlation function.

(ii) *The perfect effective theory with correction factors equal to 1.* When  $E(\mathbf{x}_i, \mathbf{x}_k)$  vanishes, and the correction factors equal 1 ( ${}_n f = 1$ ,  $g_{ik} = 1$ ) the effective source and the effective medium are equal to the original source and the original medium. It is only under these circumstances that the processed correlations actually see the correct physical source and the correct propagation medium.

(iii) *Large error term.* When the forward modelling error  $E(\mathbf{x}_i, \mathbf{x}_k)$  is significantly larger than the observational uncertainties, an effective theory that takes the same functional form as the original one (eq. 8) does not exist. Large errors can mean either of two things: (1) The effective source seen by the processed interferogram is a different one for each receiver pair. (2) The effective medium seen by the processed correlations is a different one for each time interval.

Different applications come with different requirements concerning the correction factors and the forward modelling error. In full-waveform inversion for sources and/or structure, the processing must be chosen such that the complete forward modelling error is negligible. For traditional noise tomography based on the Green function approximation, the phase of the propagation correctors should vanish to ensure that the original and the effective medium have identical velocity structure. Furthermore, the forward modelling error is allowed to introduce amplitude but no phase errors. For time-lapse monitoring, the effective source should be stable in time.

When  $E(\mathbf{x}_i, \mathbf{x}_k)$  is too large, the processed interferograms  $\widetilde{I}(\mathbf{x}_i, \mathbf{x}_k)$  can generally not be interpreted as proper correlation functions. It follows that a *necessary condition* for the meaningful solution of the inverse problem, namely consistency of the data with the forward model, is not satisfied (Parker 1977). Any attempt to invert processed interferograms under the assumption that these are correlations, would fail because the misfit between observations and synthetics will be dominated by the forward modelling error.

### 2.5.2 Analysis and minimization of the forward modelling error

Similar to the actual interferogram (24), the forward modelling error  $E(\mathbf{x}_i, \mathbf{x}_k)$ , defined in eq. (23), can be interpreted as a propagating wavefield, excited by the source

$$f_{ik}^{\text{error}}(\mathbf{x}) = \int_{\oplus} G^*(\mathbf{x}_k, \mathbf{x}') \left[ \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} {}_n S(\mathbf{x}, \mathbf{x}') \right] d\mathbf{x}' . \quad (26)$$

In contrast to the source of the interferometric wavefield, the error source  $f_{ik}^{\text{error}}(\mathbf{x})$  depends on the receiver pair  $(i, k)$ , meaning that the error wavefield for any receiver pair is excited by a different source. To understand the nature of the forward modelling error, we rewrite the term in square brackets in eqs (23) and (26):

$$\frac{1}{N} \sum_{n=1}^N {}_n e_{ik} {}_n S(\mathbf{x}, \mathbf{x}') = \frac{1}{N} S(\mathbf{x}, \mathbf{x}') \sum_{n=1}^N {}_n e_{ik} + \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} [ {}_n S(\mathbf{x}, \mathbf{x}') - S(\mathbf{x}, \mathbf{x}') ] , \quad (27)$$

where  $S(\mathbf{x}, \mathbf{x}')$  is the time average of the original psd, defined in eq. (7). We can force the first term on the right-hand side to zero by constructing the factorization of the transfer coefficient from eq. (13) such that the time-averaged factorization residual is equal to 0, that is,

$$e_{ik} = \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} = 0 , \quad (28)$$

as earlier requested in eq. (14). In Appendix A1, we demonstrate that a factorization that satisfies eq. (28) can always be found. It follows that the remaining forward modelling error is controlled by the difference  ${}_n S(\mathbf{x}, \mathbf{x}') - S(\mathbf{x}, \mathbf{x}')$ , that is, the temporal variations of the source psd

from its mean. In the special case of stationary (strictly time-invariant) noise sources, we have  ${}_nS(\mathbf{x}, \mathbf{x}') = S(\mathbf{x}, \mathbf{x}')$ , and the forward modelling error vanishes.

### 2.5.3 Estimation of the forward modelling error

As all forward modelling errors, caused by pragmatic simplifications of the underlying physics,  $E(\mathbf{x}_i, \mathbf{x}_k)$  is inherently difficult to quantify, as it requires information that the simplified problem is by construction unable to provide. When estimates of the time-dependent psd  ${}_nS(\mathbf{x}, \mathbf{x}')$  are available, for instance from physics-based models (e.g. Longuet-Higgins 1950; Hasselmann 1963; Ardhuin *et al.* 2011; Stutzmann *et al.* 2012; Traer & Gerstoft 2014; Ardhuin *et al.* 2015), the second term in eq. (27) may be constrained to some extent, thus providing an estimate of the forward modelling error. Seismic noise source inversion techniques (e.g. Schulte-Pelkum *et al.* 2004; Kurrale & Widmer-Schnidrig 2006; Rhee & Romanowicz 2006; Reading *et al.* 2014; Ermert *et al.* 2016) may be less suitable in this context as they require long-time averaging that prevents the resolution of temporal details in  ${}_nS(\mathbf{x}, \mathbf{x}')$ .

The observed data themselves offer a pragmatic estimate of the forward modelling error. Defining the effective, observed interferogram as

$$\widehat{I}(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N {}_n f g_{ik} {}_n I(\mathbf{x}_i, \mathbf{x}_k) = {}^o\widetilde{I}(\mathbf{x}_i, \mathbf{x}_k) - \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} {}_n I(\mathbf{x}_i, \mathbf{x}_k), \quad (29)$$

we obtain the *empirical forward modelling error*

$${}^o E(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} {}_n I(\mathbf{x}_i, \mathbf{x}_k) = {}^o\widetilde{I}(\mathbf{x}_i, \mathbf{x}_k) - \widehat{I}(\mathbf{x}_i, \mathbf{x}_k), \quad (30)$$

as the difference between the processed and the effective observations. The empirical forward modelling error can be computed entirely from the observations, without the need to run wavefield simulations. In the vicinity of an optimal source and structure model where  $\widehat{I}(\mathbf{x}_i, \mathbf{x}_k) \approx \widetilde{I}(\mathbf{x}_i, \mathbf{x}_k)$  and  ${}^o\widetilde{I}(\mathbf{x}_i, \mathbf{x}_k) \approx \widetilde{I}(\mathbf{x}_i, \mathbf{x}_k)$ , the empirical forward modelling error  ${}^o E(\mathbf{x}_i, \mathbf{x}_k)$  may serve as a proxy for the true forward modelling error  $E(\mathbf{x}_i, \mathbf{x}_k)$ .

### 2.5.4 Analysis of the effective power-spectral density

Similar to the forward modelling error, we can write the effective source psd, introduced in eq. (22), in terms of the temporal variability of the original source psd:

$$\widehat{S}(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_n f {}_n S(\mathbf{x}, \mathbf{x}') = \frac{1}{N} S(\mathbf{x}, \mathbf{x}') \sum_{n=1}^N {}_n f + \frac{1}{N} \sum_{n=1}^N {}_n f [ {}_n S(\mathbf{x}, \mathbf{x}') - S(\mathbf{x}, \mathbf{x}') ] \quad (31)$$

The first term on the right-hand side is equal to the original psd  $S(\mathbf{x}, \mathbf{x}')$  when we request that the average source correction is equal to 1 (see eq. 15):

$$\frac{1}{N} \sum_{n=1}^N {}_n f = 1, \quad (32)$$

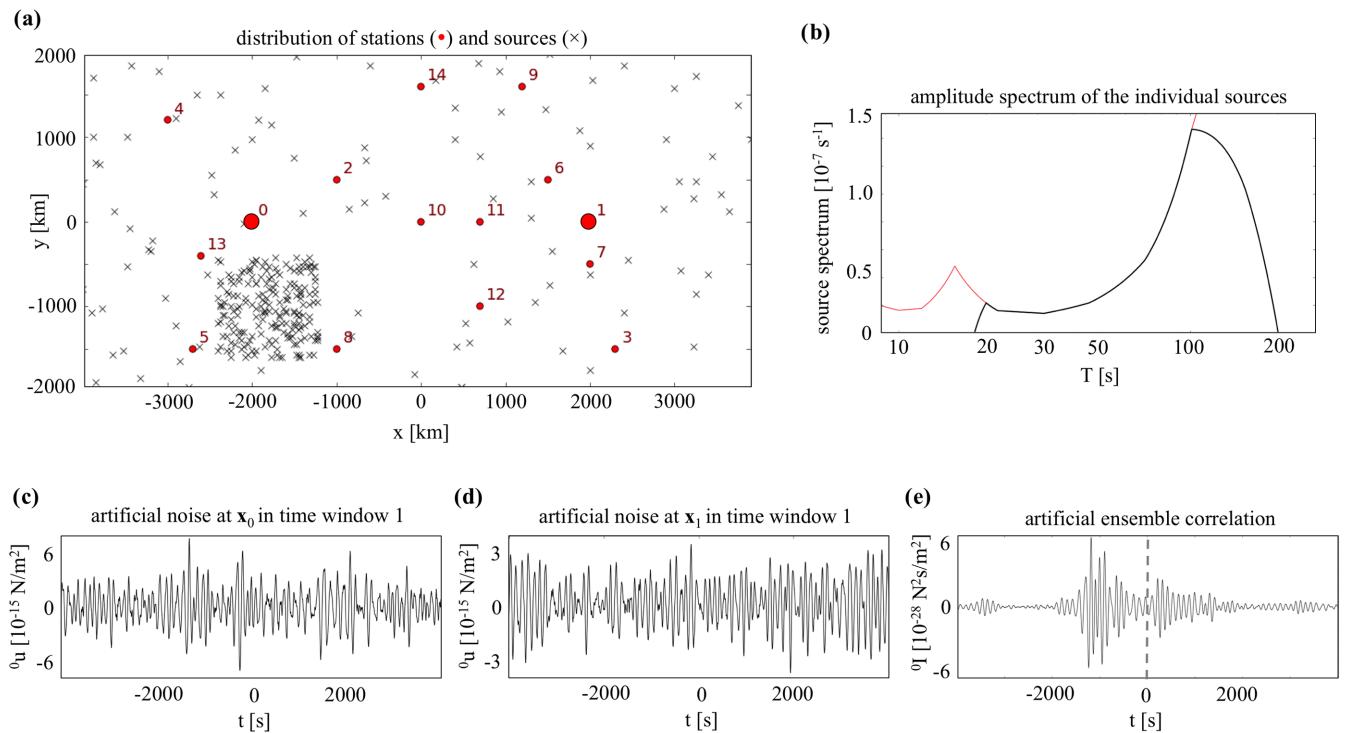
which simplifies the effective psd to

$$\widehat{S}(\mathbf{x}, \mathbf{x}') = S(\mathbf{x}, \mathbf{x}') + \frac{1}{N} \sum_{n=1}^N {}_n f [ {}_n S(\mathbf{x}, \mathbf{x}') - S(\mathbf{x}, \mathbf{x}') ]. \quad (33)$$

Eq. (33) illustrates that the difference between the effective psd  $\widehat{S}(\mathbf{x}, \mathbf{x}')$  and the original psd  $S(\mathbf{x}, \mathbf{x}')$  is controlled by the temporal variations of the sources  ${}_n S(\mathbf{x}, \mathbf{x}') - S(\mathbf{x}, \mathbf{x}')$ . In this sense, effectiveness is induced by non-stationarity, that is, the time dependence of the sources. For time-independent sources, the original and effective psd's are identical, that is,  $\widehat{S}(\mathbf{x}, \mathbf{x}') = S(\mathbf{x}, \mathbf{x}')$ .

In the context of forward modelling, the effective psd can be computed from eq. (33) provided that a model for the temporal details of the source is available. As mentioned in Section 2.5.3, this information may come from physics-based models or data-based source inversion. Conceptually, the computation of the effective psd is similar to the computation of a long-wavelength equivalent effective medium using structural homogenization techniques (e.g. Backus 1962; Fichtner & Igel 2008; Capdeville *et al.* 2010, 2015).

In the context of inverse problems, it is clear from the forward problem eqs (21) and (24) that ensemble interferograms are inherently unable to constrain the temporal variability of the sources. An inverse problem solution for either structure or sources must therefore start with an initial guess for the *effective* sources, which may then be improved to yield an optimal effective source. The true optimal source cannot be determined unambiguously from the effective optimal source because the individual  ${}_n S$  are unknown. Again invoking the analogy with structural homogenization, we note that any tomographic inversion can only constrain a long-wavelength equivalent Earth model. The underlying fine-scale structure can also not be constrained uniquely, unless independent data are added (e.g. Capdeville *et al.* 2013; Fichtner *et al.* 2013a; Bodin *et al.* 2015). We will discuss the inverse problem aspect in more detail in Section 4.



**Figure 3.** Setup for the modelling of artificial noise and noise correlations. (a) Distribution of noise point sources (black crosses) and stations (red dots). Receiver pair 0–1, used in most of the following examples, is emphasized. (b) Amplitude spectrum of the individual point-localized noise sources consisting of a bandpass with cut-offs at 20 and 100 s, times the *New Low Noise Model* of Peterson (1993). The latter is plotted in red. The periods used in our experiments are slightly above the microseismic peak and close to the hum at long periods. (c,d) Artificial noise time series at receivers  $x_0$  and  $x_1$  in the first of 5000 time windows. (e) Ensemble correlation between  $x_0$  and  $x_1$ , obtained by averaging correlations over all time windows. [ER]

### 3 NUMERICAL EXAMPLES FOR EFFECTIVE MODELLING IN THE ACOUSTIC CASE

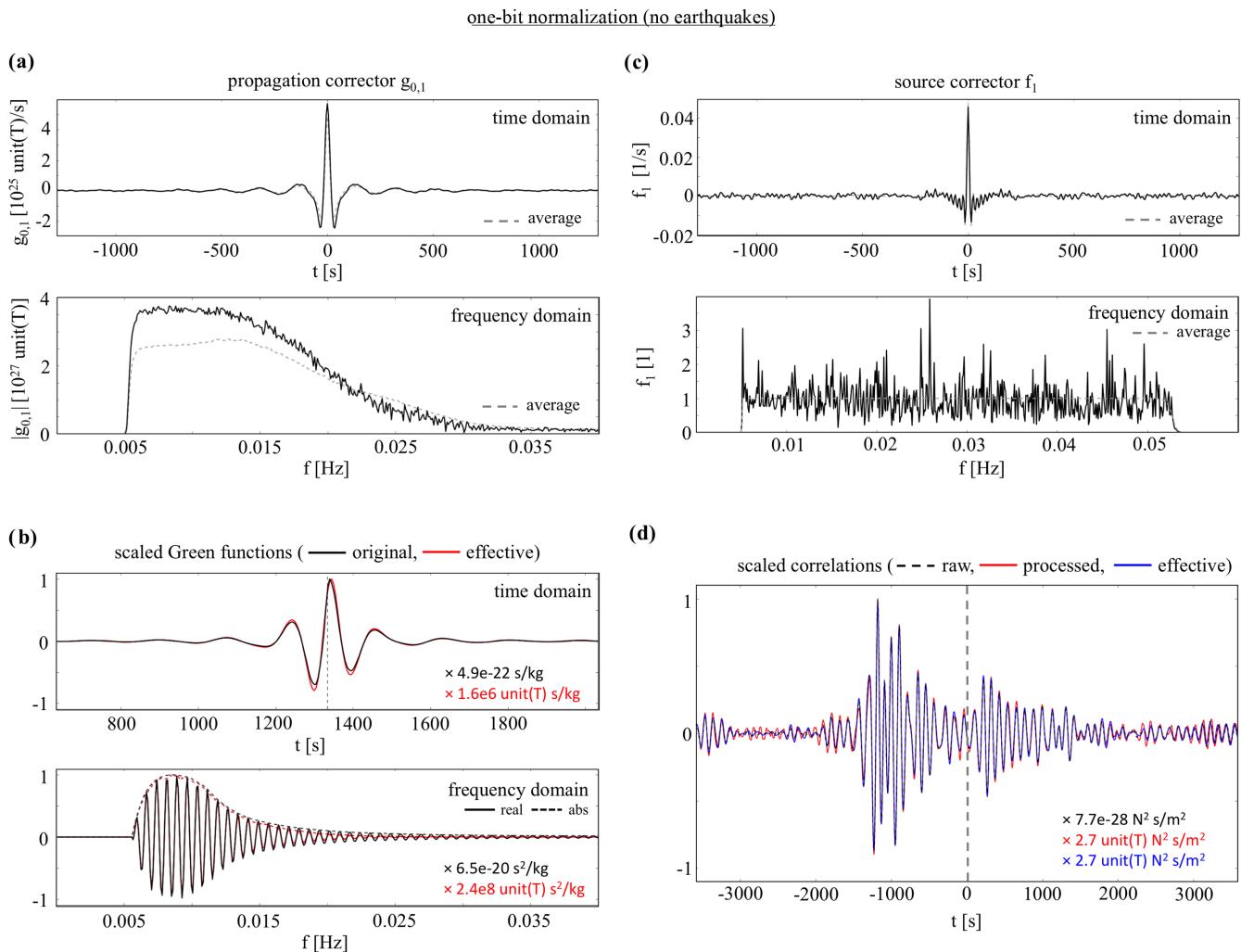
To illustrate our developments from Section 2, we consider a series of simple and easily reproducible numerical examples. Starting with scenarios where no earthquakes are present, we will investigate the influence of various widely used processing schemes on the effective wave propagation and the forward modelling error. Subsequently, we will transition to examples where earthquakes play an increasingly important role in the source distribution.

For all examples, we use the same 2-D computational domain as in Fig. 2, and we assume that the medium is homogeneous, with velocity  $v = 3000 \text{ m s}^{-1}$ , density  $\rho = 3000 \text{ kg m}^{-3}$ , and  $Q = 100$ . Since the Green function for the homogeneous 2-D medium is given explicitly by eq. (11), we can afford to compute an artificial observed noise field, excited by a distribution of point-localized sources with a different random phase for each time window. In our basic setup, shown in Fig. 3(a), 100 point sources are randomly distributed throughout the domain. Additional 200 point sources are placed inside a smaller subdomain, thus leading to an overall heterogeneous source density. To obtain realistic artificial noise, we use the *New Low Noise Model* of Peterson (1993) as temporal source power spectrum, and we apply a bandpass between 0.01 and 0.05 Hz to limit computational requirements. The complete spectrum is shown in Fig. 3(b). From 5000 artificial noise traces of 8000 s duration at each receiver, we compute raw and processed ensemble correlations that we treat as data. The total duration of  $5000 \times 8000 \text{ s}$  ( $\sim 463$  days) ensures that the correlations, shown in Fig. 3(e) for receiver pair  $(x_0, x_1)$ , have converged to the extent that the effect of any processing is larger than the difference to the expected correlation for infinitely many time windows. The heterogeneous source distribution causes the wave packet on the causal branch to arrive earlier ( $\sim 500$  s) than the wave packet on the acausal branch ( $\sim 1000$  s).

Using the artificial data, we compute source correctors  ${}_n f$  and propagation correctors  $g_{ik}$  as described in Appendix A. They determine the effective source psd and the effective Green function. To emphasize the influence of processing on the effectiveness, we then compute effective correlations using the correct medium properties and the correct time-dependent psd's  ${}_n S$ . The difference between processed correlations  $\tilde{\mathcal{I}}(x_i, x_k)$  and effective correlations  $\hat{\mathcal{I}}(x_i, x_k)$  therefore equals the forward modelling error. In real-world applications where the true medium and the true time-dependent sources are not known, we will generally not enjoy the luxury of being able to quantify the forward modelling error as easily.

#### 3.1 Random noise sources in the absence of earthquakes

We start our tour of numerical examples with the simplest scenario where no earthquakes are present. Realizing that the term ‘earthquake’ is not perfectly appropriate in an acoustic medium, we nevertheless use it to designate point-localized transient sources, and to emphasize the analogy with solid-Earth seismology.



**Figure 4.** Effective forward modelling for station pair  $(\mathbf{x}_0, \mathbf{x}_1)$ . The processing is limited to one-bit normalization. (a) Time-domain propagation corrector and absolute value of the frequency-domain propagation corrector  $g_{0,1}$ . The grey dashed lines mark the average of the propagation correctors over all receiver pairs. (b) Original Green function  $G(\mathbf{x}_0, \mathbf{x}_1)$  (black) and effective Green function  $g_{0,1} G(\mathbf{x}_0, \mathbf{x}_1)$  (red) in the time- and frequency domains. For better visibility, the Green functions are scaled to a maximum value of 1, with scaling factors marked in the figures. Solid lines in the frequency-domain plot denote the real part, and dashed lines denote the absolute value. (c) Time- and frequency-domain source corrector for the first time window,  $f_1$ . The average of all source correctors over all time windows is shown as grey dashed line. In the frequency domain, it is 1 by construction. (d) Scaled time-domain correlation functions. Raw (unprocessed) correlations are shown as black dashed curve, processed correlations in red, and effective correlations in blue. The forward modelling error is the difference between processed and effective correlations. [ER]

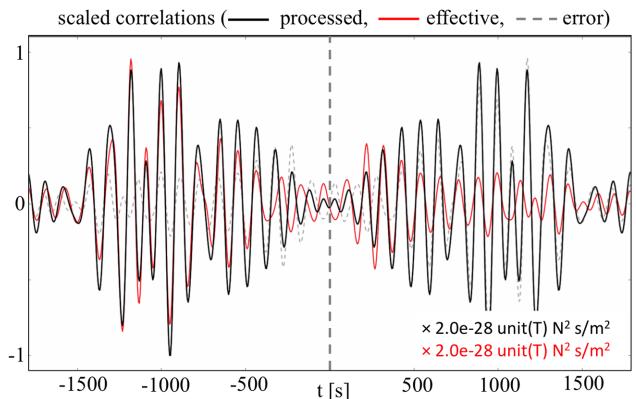
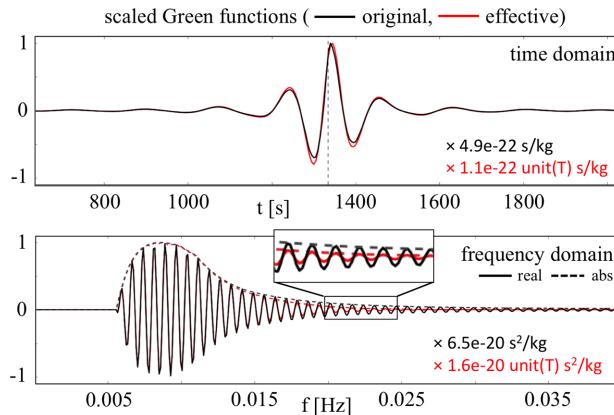
### 3.1.1 One-bit normalization

As the first processing scheme, we consider one-bit normalization, where the samples of the raw time series are set to 1 for positive and to  $-1$  for negative amplitudes. The effective forward modelling for receiver pair  $(\mathbf{x}_0, \mathbf{x}_1)$  is summarized in Fig. 4. The time-domain propagation corrector  $g_{0,1}$  (Fig. 4a) resembles a band-limited version of a  $\delta$ -pulse, centred around  $t = 0$ . Its frequency-domain version is nearly flat from 0.005–0.015 Hz where the signal, filtered by the bandpass and the natural source spectrum, has its largest amplitude. For higher frequencies, the amplitude of the propagation corrector drops, indicating that these are propagated less efficiently in the effective medium.

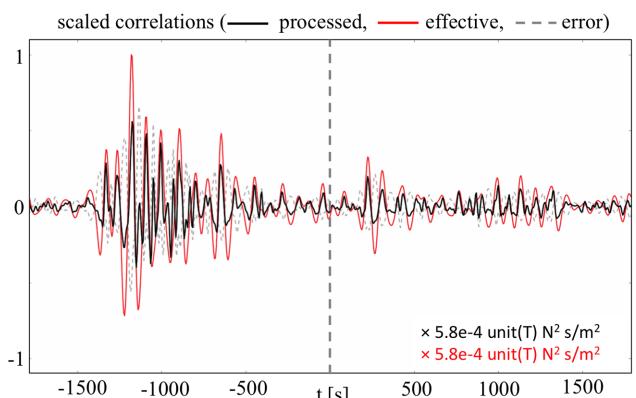
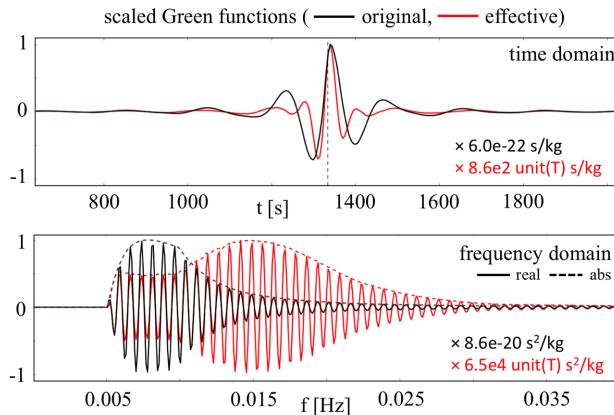
The original Green function  $G(\mathbf{x}_0, \mathbf{x}_1)$  and its effective counterpart  $g_{0,1} G(\mathbf{x}_0, \mathbf{x}_1)$ , displayed in Fig. 4(b), are nearly perfectly in phase for all frequencies. The amplitude differences at higher frequencies result from the amplitude characteristic of the propagation corrector. The source corrector for the first time window,  $f_1$  (Fig. 4c) is also peaked around  $t = 0$  in the time domain, but shows strong oscillations in the frequency domain. By construction, the average of all frequency-domain source correctors, shown as grey dashed line, is equal to 1, at least within the pass band. Fig. 4(d) shows the raw, processed and effective correlation functions. They are almost perfectly in phase, and the amplitudes of the processed and the effective correlation functions are nearly identical. It follows that the forward modelling error, that is, the difference between processed and effective correlations, is small compared to the amplitude of the correlation functions themselves. These results are similar for all other receiver pairs.

Our numerical experiment is in good agreement with theoretical studies on correlations of one-bit normalized noise that predict vanishing phase shifts between raw and processed correlations in the case of stationary sources (van Vleck & Middleton 1966; Cupillard &

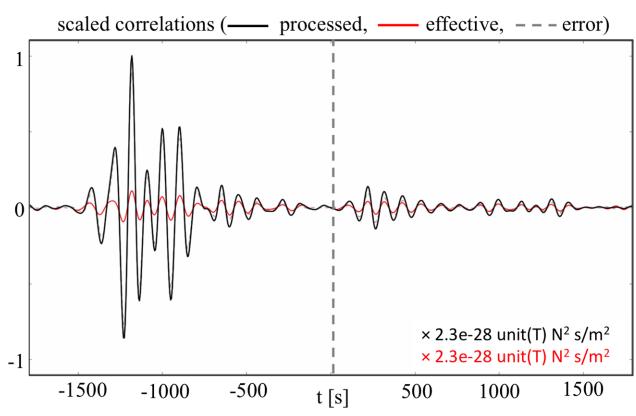
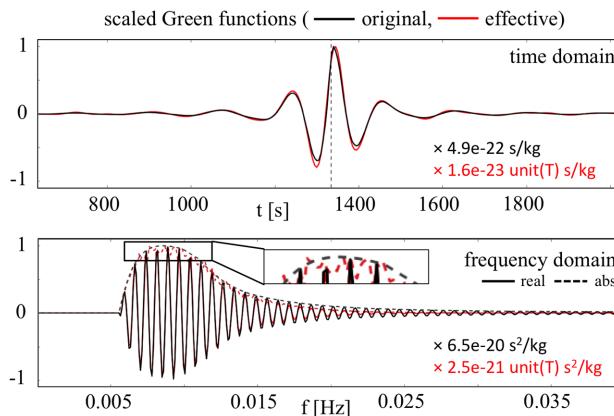
## (a) causal/acausal average



## (b) spectral whitening



## (c) phase-weighted stack

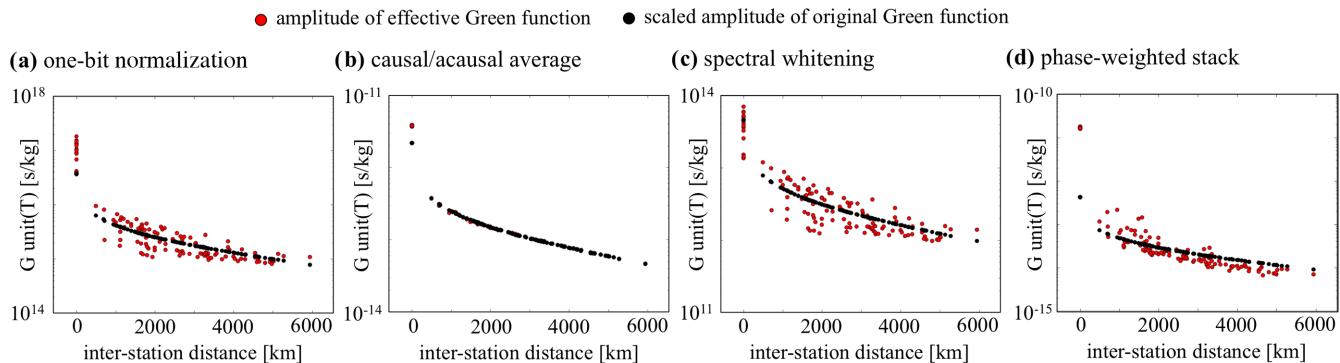


**Figure 5.** Green functions (original and effective) and correlation functions (processed and effective) for station pair ( $\mathbf{x}_0, \mathbf{x}_1$ ) and a selection of processing schemes. (a) Averaging of the causal and acausal branches of the correlation function. The original (black) and effective (red) Green functions between the two stations are shown to the left; processed (black) and effective (red) correlation functions to the right. The forward modelling error, that is, the difference between processed and effective correlations, is plotted as faint dashed line. (b) Spectral whitening of the individual recordings. (c) Phase-weighted stacking. The dashed curve of the forward modelling error is mostly hidden by the curve for the processed correlations. [ER]

Capdeville 2010; Cupillard *et al.* 2011; Hanasoge & Branicki 2013). Therefore, this special case constitutes a successful test of the effective forward modelling theory developed in the previous sections.

### 3.1.2 Other processing schemes

Acknowledging that the range of existing processing schemes is too large to be presented in a comprehensive fashion, we limit ourselves to a few widely used and instructive scenarios, summarized in Fig. 5.



**Figure 6.** Amplitude versus interstation distance of the effective Green functions  $g_{ik}G(\mathbf{x}_i, \mathbf{x}_k)$  at frequency 0.15 Hz for various processing schemes and all possible receiver pairs (red dots). Appropriately scaled amplitudes of the original interstation Green functions are shown for reference as black dots. The applied processings are (a) one-bit normalization, (b) causal/causal averaging, (c) spectral whitening, and (d) phase-weighted stacking. [ER]

Averaging the causal and acausal branches of the correlation function (Fig. 5a) has nearly no effect on the phase of the Green function. Waves in the original and effective media are almost in phase for all frequencies. However, the effective medium propagates higher frequencies ( $\gtrsim 0.2$  Hz) less efficiently than lower frequencies ( $\lesssim 0.2$  Hz), which is the result of preferentially cancelling higher-frequency oscillations in the causal/causal averaging. The processed and effective correlation functions in the right panel of Fig. 5(a) differ significantly, thus implying that the forward modelling error is large. While the processed correlation is by design symmetric, the effective correlation is not. It follows that there is no combination of effective medium and effective source such that the processed correlation can be represented as a proper correlation function. This result can be understood intuitively: Wave packets on the causal and acausal branches of the raw correlation arrive at different times ( $\sim 500$  s versus  $\sim 1000$  s). The averaged correlation therefore has two wave packets arriving at different times on both branches, which would require an impossible acoustic medium that propagates two waves at different velocities instead of one.

Spectral whitening, shown in Fig. 5(b), does not affect the phase of the Green function but removes the  $1/\sqrt{\omega}$  amplitude decay (equation 11). The effective medium propagates high frequencies more efficiently. This is also reflected in the effective correlation function, which resembles the processed correlation function, though without capturing all its details.

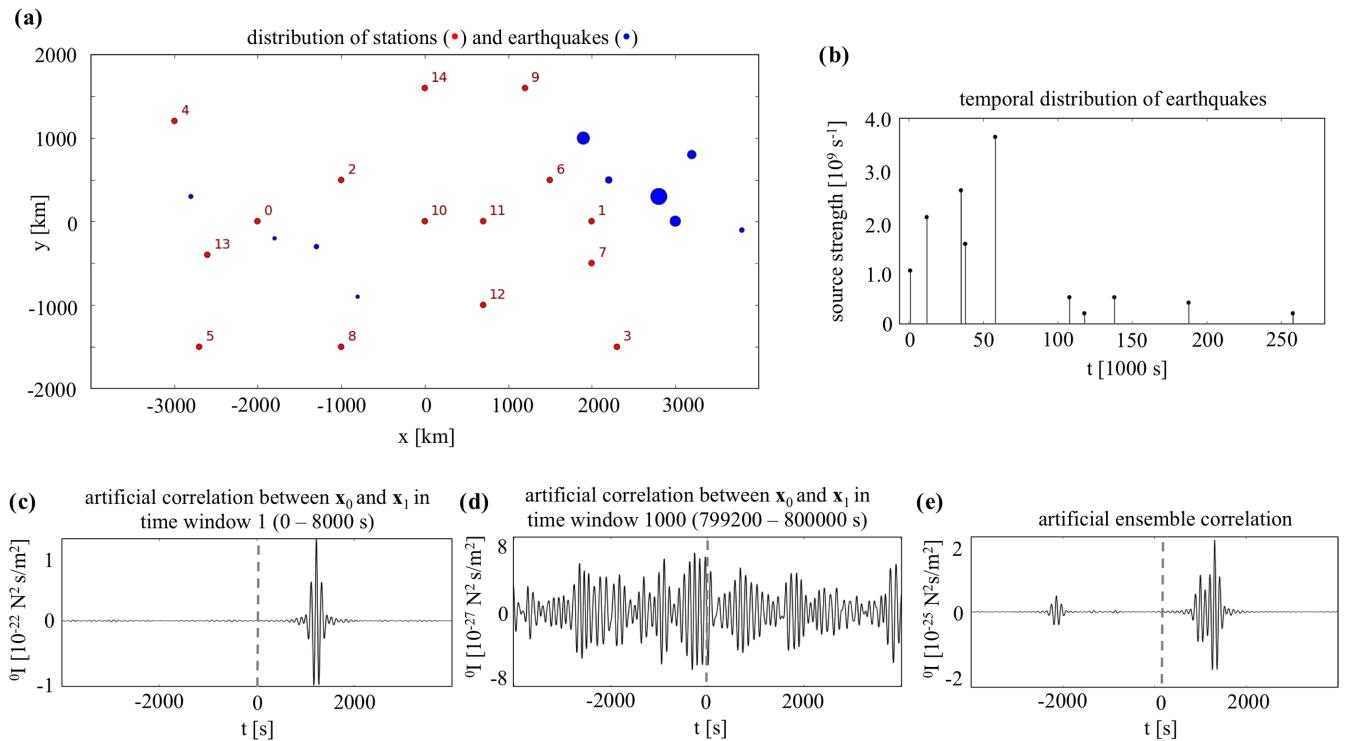
Phase-weighted instead of linear stacking (averaging) emphasizes coherent signals by using the phase stack as time-dependent weighting function (Schimmel & Paulsen 1997). As shown in Fig. 5(c), phase-weighted stacking causes the effective Green function to have an amplitude spectrum that differs slightly from the original Green function, while having no effect on its phase. As expected, the processed and effective correlation functions are less complex than the original, containing only few compact wave trains. While processed and effective correlations have nearly identical waveforms, they differ significantly in amplitude, indicating that the processed correlations cannot be interpreted as an effective correlation function. However, measurements of surface wave dispersion could be useful.

As the previous examples have shown, different types of processing lead to different properties of the effective propagation. One of these properties, the amplitude of the effective Green function  $g_{ik}G(\mathbf{x}_i, \mathbf{x}_k)$  versus interstation distance, is displayed in Fig. 6. While the amplitude decay of the effective Green function for causal/causal averaging is the same as for the original Green function, all other previously considered processing schemes lead to a more complex amplitude-versus-distance behaviour. Indeed, effective Green functions for identical interstation distances may have different amplitudes, thus demonstrating that the effective medium is not necessarily in accord with our intuitive understanding of wave propagation physics in a homogeneous medium. The effective medium is effectively heterogeneous.

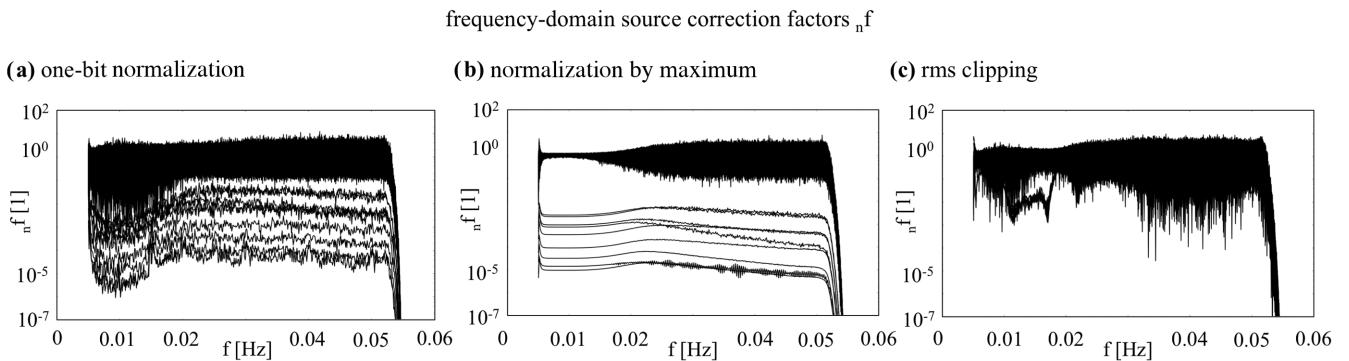
Among the four processing schemes considered in the previous paragraphs, only one-bit normalization leads to a forward modelling error that may be considered negligible for practical purposes. The nature of the forward modelling error for the other processing schemes places constraints on the misfit functionals that can be used to extract meaningful information from the difference between observed processed and synthetic effective correlations. While time-like measures of misfit such as cross-correlation time shifts (Luo & Schuster 1991; Dahlen *et al.* 2000; van Leeuwen & Mulder 2010), time-frequency phase misfits (Kristekova *et al.* 2006; Fichtner *et al.* 2008) or the temporal Generalized Seismological Data Functionals (Gee & Jordan 1992) may be used with caution for all processing schemes shown in Fig. 5, reliable amplitude information can certainly not be extracted. Before transitioning to examples where earthquakes are included, we remark that a sequence of different processings may of course be used instead of the isolated ones that we have analysed here in the interest of a simple interpretation.

### 3.2 Random noise sources with earthquakes

Numerous processing schemes have been developed to eliminate large-amplitude transient signals such as earthquakes in an automatic fashion. Not having made restrictive assumptions on the nature of the sources  ${}_nF(\mathbf{x})$ , we can extend our analysis to scenarios where earthquakes are present. For this, we add 10 earthquakes of variable size that cluster in the northeastern corner of the domain, as illustrated in Figs 7(a) and (b). Earthquakes are modelled as  $\delta$ -functions in space and time, with variable amplitude. Correlations for time windows with and without



**Figure 7.** Setup for the modelling of artificial correlations in the presence of earthquakes. (a) Distribution of stations (red dots) and earthquakes (blue dots) with size proportional to source strength. Earthquakes are superimposed on the noise sources shown in Fig. 3. (b) Temporal distribution of the earthquakes. (c,d) Artificial correlations between  $x_0$  and  $x_1$  in time window 1 (0–8000 s) and time window 1000 (799200–800000 s), respectively. (e) Ensemble correlation between  $x_0$  and  $x_1$ , obtained by averaging correlations over all time windows. [ER]

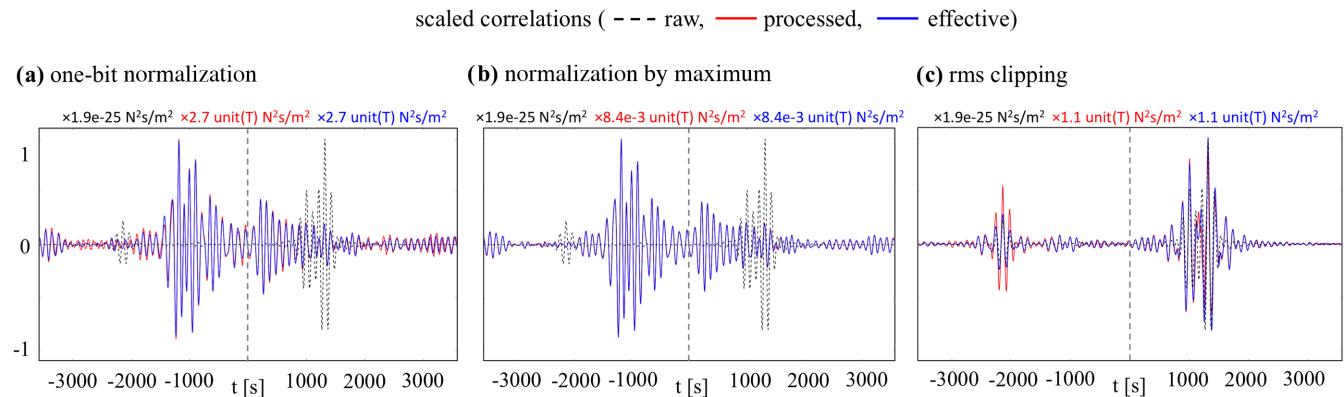


**Figure 8.** Frequency-domain source correctors  $n_f$  for all 5000 time windows and different processing schemes: (a) one-bit normalization, (b) normalization of correlations for individual time windows by their maximum amplitude and (c) clipping of recordings for individual time windows above their rms value. [ER]

earthquakes, and the ensemble correlation are displayed in Figs 7(c)–(e) for station pair ( $x_0, x_1$ ). The ensemble correlation is dominated by the earthquakes and therefore differs significantly from the ensemble correlation in Fig. 3, where earthquakes are absent.

Again realizing that not all possible processing schemes can be presented, we limit our attention to three special cases that are useful to convey basic concepts: In order to connect to Section 3.1.1, we continue to analyse one-bit normalization. Furthermore, we consider the normalization of correlations for individual time windows by their respective maximum amplitudes, and the clipping of individual recordings at their rms value. The collection of source correctors  $n_f$  for all 5000 time windows is displayed in Fig. 8.

The source correctors for one-bit normalization, shown in Fig. 8(a), are close to 1 for all frequencies, with the exception of those nine correctors that correspond to the nine time windows containing earthquakes. The source psd's for these nine windows,  $n_S$ , are down-weighted by their source correctors by up to 6 orders of magnitude. As desired, the effective source  $\widehat{S}$  is practically free of earthquakes. The down-weighting by the source correctors is frequency dependent, with lower frequencies around 0.01 Hz being suppressed more efficiently than higher frequencies above 0.02 Hz.



**Figure 9.** Scaled time-domain correlation functions. Raw (unprocessed) correlations are shown as black dashed curve, processed correlations in red and effective correlations in blue. The forward modelling error is the difference between processed and effective correlations. Processing schemes are the same as in Fig. 8: (a) one-bit normalization, (b) normalization of correlations for individual time windows by their maximum amplitude and (c) clipping of recordings for individual time windows above their rms value. [ER]

Normalization by maximum in Fig. 8(b) has a similar effect, the most notable difference being a less pronounced frequency dependence. All source correctors for rms clipping in Fig. 8(c) are comparatively close to 1 because the large rms of time windows with earthquakes prevents a sufficient amplitude reduction to eliminate the transient signal.

The properties of the source correctors translate into the properties of processed and effective correlations, shown in Fig. 9. One-bit normalization and normalization by maximum effectively eliminate the few time windows with earthquakes. The result is an ensemble correlation built from slightly fewer time windows, with little effect on the expected value. Consequently, the ensemble correlations for both processing schemes are nearly indistinguishable, and the difference between processed and effective correlations, that is, the forward modelling error, is very small. For rms clipping, the ensemble correlation remains dominated by the earthquakes. Notable differences between processed and effective correlations suggest that the forward modelling error may not be negligible.

#### 4 SOLVING INVERSE PROBLEMS

The effective interferogram

$$\widehat{I}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} g_{ik} G(\mathbf{x}_i, \mathbf{x}) \left[ \int_{\oplus} G^*(\mathbf{x}_k, \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x}' \right] d\mathbf{x}, \quad (34)$$

first introduced in eq. (24), constitutes, by construction, a deterministic and computable forward model for processed correlations. It can, as such, be used to solve inverse problems for the effective psd  $\widehat{S}$  and the Earth model parameters  $\mathbf{m}$  through a comparison with observed processed correlations  $\widetilde{I}(\mathbf{x}_i, \mathbf{x}_k)$ . In the case of small model spaces and computationally inexpensive forward simulations, the inverse problem may be solved probabilistically (e.g. Mosegaard & Tarantola 1995; Sambridge & Mosegaard 2002; Sambridge *et al.* 2013). Otherwise, deterministic methods driven by gradient-based iterative descent schemes will most likely be used to find an optimal psd and Earth model that minimize a misfit or measurement functional  $\chi = \chi[\widehat{I}(\mathbf{x}_i, \mathbf{x}_k)]$  to within the uncertainties. The latter are composed of (i) the observational uncertainties, (ii) the processing-induced forward modelling errors estimated, for instance, via the empirical forward modelling error defined in Section 2.5.3, and (iii) possible convergence errors in the ensemble interferograms, especially in the case where asymptotically uncorrelated sources in the sense of  $\widehat{S}(\mathbf{x}, \mathbf{x}') = \widehat{S}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}')$  are assumed. All of these must be taken into account through the application of appropriate data covariance matrices (e.g. Tarantola 2005).

In the following paragraphs, we will derive Fréchet kernels for the effective psd  $\widehat{S}$  and the Earth model parameters  $\mathbf{m}$ , that can be used in gradient-based optimization. For this, we note that the variation or Fréchet derivative of the misfit functional  $\chi$  due to variations in the effective interferogram  $\widehat{I}$  can generally be written as (Fichtner 2014)

$$\delta\chi = 2 \operatorname{Re} \int_{\omega=0}^{\infty} \delta\widehat{I}(\mathbf{x}_i, \mathbf{x}_k) f d\omega \quad (35)$$

with a frequency-dependent integral kernel  $f(\omega)$  that we will later identify as the adjoint source. Eq. (35) is a consequence of Riesz' theorem, which states that any linear operator—including the Fréchet derivative  $\delta\chi$  that is by definition linear in the perturbation  $\delta\widehat{I}$ —can be written as a scalar product over  $\delta\widehat{I}$  and some other function, that is,  $f$  in our case (e.g. Rudin 1966). The specific form of  $f$  is determined by the definition of the misfit  $\chi$ . Examples for traveltimes and waveform energy measurements are provided in Appendix B2.

#### 4.1 Fréchet kernels for the effective source power-spectral density

##### 4.1.1 Derivation

In the specific case where the variations of  $\widehat{I}$  are due to a variation in the effective source psd  $\widehat{S}$ , we have

$$\delta\widehat{I}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} g_{ik} G(\mathbf{x}_i, \mathbf{x}) \left[ \int_{\oplus} G^*(\mathbf{x}_k, \mathbf{x}') \delta\widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x}' \right] d\mathbf{x}. \quad (36)$$

Inserting this expression into eq. (35) for the variation of  $\chi$ , we obtain

$$\delta\chi = 2 \operatorname{Re} \int_{\omega=0}^{\infty} \int_{\oplus} \int_{\oplus} g_{ik} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') f \delta\widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x}' d\mathbf{x} d\omega = \int_{\omega=0}^{\infty} \int_{\oplus} \int_{\oplus} K(\mathbf{x}, \mathbf{x}') \delta\widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x}' d\mathbf{x} d\omega, \quad (37)$$

with the space- and frequency-dependent Fréchet kernel for the effective source psd

$$K(\mathbf{x}, \mathbf{x}') = 2 \operatorname{Re} g_{ik} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') f. \quad (38)$$

The kernel  $K$  describes the relation between variations in the source psd and the resulting variations of the measurement or misfit. In the special case of uncorrelated sources, in the sense of  $\widehat{S}(\mathbf{x}, \mathbf{x}') = \widehat{S}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}')$ , eqs (37) and (38) simplify to

$$\delta\chi = \int_{\omega=0}^{\infty} \int_{\oplus} K(\mathbf{x}) \delta\widehat{S}(\mathbf{x}) d\mathbf{x} d\omega, \quad K(\mathbf{x}) = 2 \operatorname{Re} g_{ik} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}) f. \quad (39)$$

A further simplification is possible when  $\widehat{S}$  is independent of frequency, that is, a constant in the integration over frequency in eqs (37) and (39):

$$\delta\chi = \int_{\oplus} K(\mathbf{x}) \delta\widehat{S}(\mathbf{x}) d\mathbf{x}, \quad K(\mathbf{x}) = 2 \operatorname{Re} \int_{\omega=0}^{\infty} g_{ik} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}) f d\omega. \quad (40)$$

Eq. (40) forms the basis of the numerical examples that we present in the following sections for 2-D, acoustic, unbounded media.

##### 4.1.2 Example I: source kernels for raw and processed correlations

To illustrate source kernels and their dependence on processing, we continue with the earthquake-free scenario introduced in Section 3.1 and Fig. 3. We again consider station pair  $(\mathbf{x}_0, \mathbf{x}_1)$ , and for simplicity we assume that the (effective) psd is frequency-independent so that eq. (40) becomes applicable.

Fig. 10(a) shows the raw correlation function with grey-shaded time windows within which we measure the waveform energy. The source kernels corresponding to the different windows are displayed in Fig. 10(c). The waveform in window 1, roughly corresponding to the ballistic wave travelling from  $\mathbf{x}_0$  to  $\mathbf{x}_1$ , is most sensitive to source perturbations in a hyperbolically shaped region west of receiver  $\mathbf{x}_0$ . Sensitivity within the first Fresnel zone is positive, indicating that increasing source psd leads to increasing waveform energy. For time window 2, sensitivity in the first Fresnel zone is flipped from positive to negative, and higher Fresnel zones gain importance. Time window 3, located near zero lag time, has dominant sensitivity in higher Fresnel zones. Time window 4 is centred around the ballistic wave travelling from  $\mathbf{x}_1$  to  $\mathbf{x}_0$ , and is therefore most sensitive to source perturbations in the first Fresnel zone east of receiver  $\mathbf{x}_1$ .

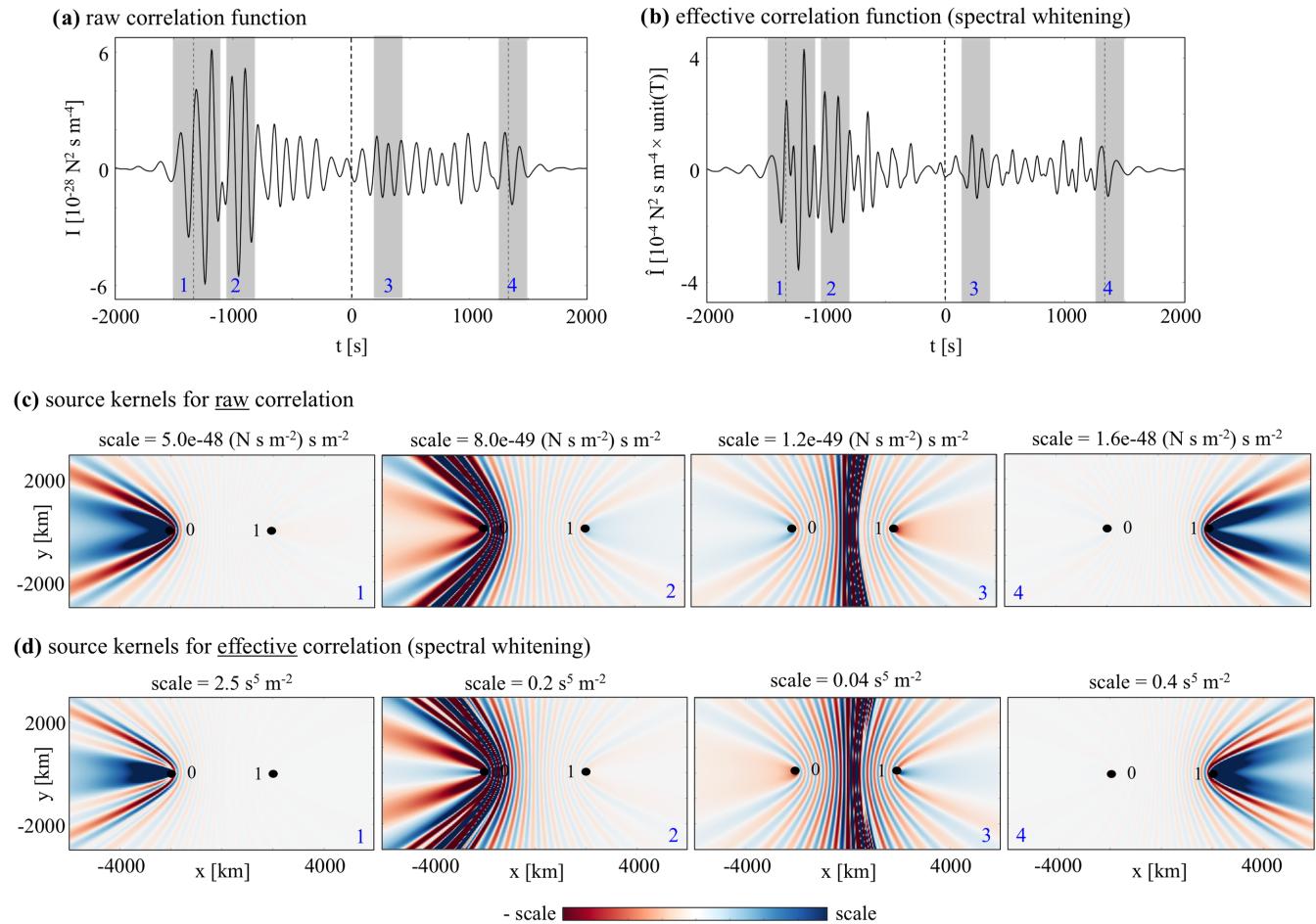
Spectral whitening emphasizes higher frequencies in the correlation function shown in Fig. 10(b). The source kernels corresponding to the same time windows used before are presented in Fig. 10(d). In addition to having different physical units and amplitudes, the kernels show different sensitivity patterns with a more rapidly oscillating succession of Fresnel zones. These differences result from (i) the higher-frequency signal within the time windows that results in a higher-frequency adjoint source, and (ii) the effective propagation of the adjoint field that further emphasizes higher frequencies, as illustrated in Fig. 5(b).

##### 4.1.3 Example II: locating earthquakes

When correlations are dominated by large-amplitude earthquakes, source kernels can be used to constrain their location. A measurement window centred around an earthquake-dominated correlation waveform at time  $t = T$ , produces an adjoint source proportional to  $e^{i\omega T}$  (Appendix B2). It follows from eq. (40) that the source kernel is then proportional to a frequency integral with oscillating integrand:

$$K(\mathbf{x}) \propto 2 \operatorname{Re} \int_{\omega=0}^{\infty} e^{-i\frac{\omega}{v}|\mathbf{x}-\mathbf{x}_i|} e^{i\frac{\omega}{v}|\mathbf{x}-\mathbf{x}_k|} e^{i\omega T} d\omega. \quad (41)$$

The integral is significantly different from zero near the stationary phase region where the exponent vanishes, that is, for  $\mathbf{x}$  on the hyperbola defined by  $|\mathbf{x} - \mathbf{x}_i| = |\mathbf{x} - \mathbf{x}_k| + vT$ . All points along the hyperbola affect the measurement and may thus be considered an infinite set of



**Figure 10.** (a,b) Raw and processed (spectrally whitened) correlation functions, and (c,d) the corresponding source kernels for different time windows indicated by grey shading and blue numbers from 1 to 4. Kernels for raw and processed correlations have different units, amplitudes and sensitivity patterns. [ER]

candidate points for the earthquake location. The ambiguity can be reduced by adding receiver pairs. In 3-D, four receivers are required to uniquely determine the location, assuming that knowledge of the medium and measurements are perfect.

In the case of infinite bandwidth, earthquake location using correlations is equivalent to ray-theory-based multilateration where a source location is estimated from time difference measurements. This equivalence results from the correspondence of an absolute time measurement on a correlation and a time difference measurement on the original signals. Commonly used in navigation and air traffic control, various analytical and numerical approaches to multilateration are available (e.g. Fang 1990; Bucher & Misra 2002).

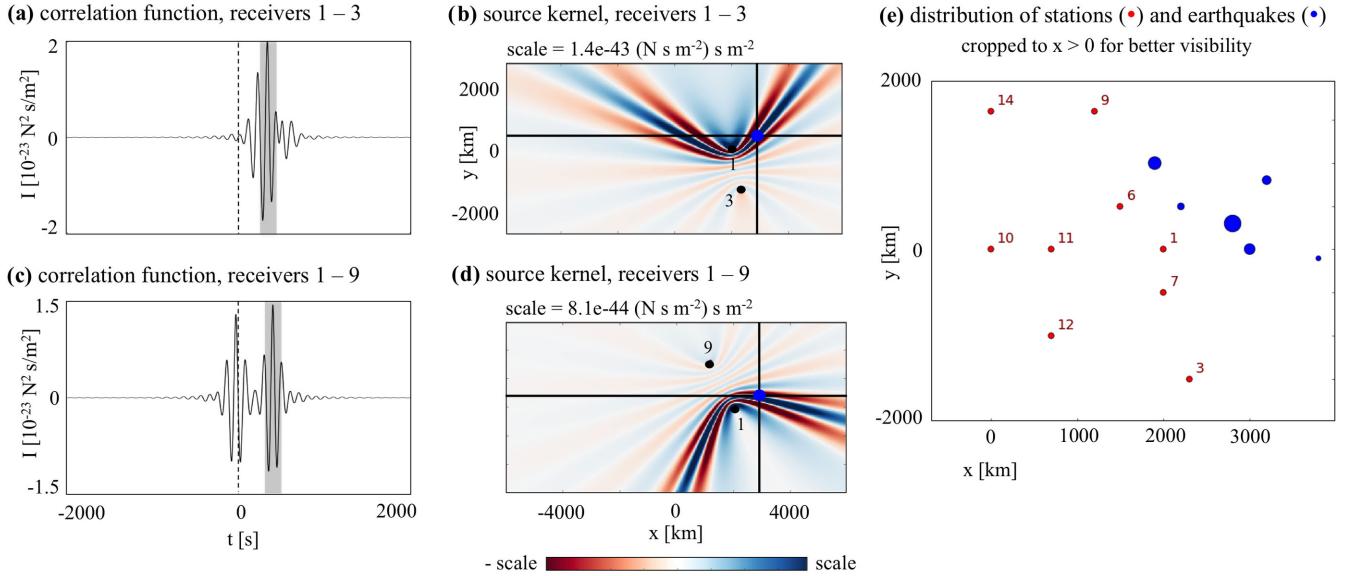
With bandlimited signals, rays expand into finite-frequency kernels, thereby increasing location uncertainty that may be compensated by incorporating additional data. However, the basic principle of multilateration remains intact, as illustrated in Fig. 11 using the example with earthquake-dominated correlations from Section 3.2. Measuring the energy of the largest-amplitude correlation waveforms for station pairs  $(\mathbf{x}_1, \mathbf{x}_3)$  and  $(\mathbf{x}_1, \mathbf{x}_9)$ , produces two hyperbolically shaped source kernels. The kernels overlap near the position of the largest earthquake, indicating that it can indeed be localized.

While this example illustrates the concept, the location of individual earthquakes from ensemble correlations will be more difficult in practice unless one earthquake is clearly dominant. More often, an effective superposition of numerous earthquake and ambient noise sources will be imaged.

#### 4.2 Fréchet kernels for structure

The derivation of sensitivity kernels for Earth structure  $\mathbf{m}$  is more involved than for the effective psd because  $\mathbf{m}$  does not appear explicitly in the forward modelling eqs (21) and (25), but implicitly in the Green function  $G$ . Our strategy is to first derive an expression for the variation of the interferometric wavefield  $\delta \hat{I}$  with respect to variations  $\delta \mathbf{m}$ , and then to insert this expression into the generic form of the misfit variation  $\delta \chi$ , given in eq. (35). In the interest of obtaining simpler expressions, and also in anticipation of the more complex elastic case treated in Section 5, we introduce a new interferometric wavefield

$$\hat{I}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') \hat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' = g_{ik}^{-1} \hat{I}(\mathbf{x}_i, \mathbf{x}_k), \quad (42)$$



**Figure 11.** Illustration of finite-frequency multilateration using correlation source kernels. (a) Correlation between receivers at  $\mathbf{x}_1$  and  $\mathbf{x}_3$  in the earthquake-dominated setup introduced in Section 3.2 and Fig. 7. The energy of the largest amplitudes within the grey-shaded window is measured, which leads to the source kernel displayed in (b). Panels (c) and (d) show the same for the correlation between receivers at  $\mathbf{x}_1$  and  $\mathbf{x}_9$ . The hyperbolically shaped source kernels for both receiver pairs overlap near the location of the largest earthquake, marked by a blue dot. (e) Distribution of receivers (red dots) and earthquakes (blue dots, with size proportional to source strength). [ER]

which is excited by the effective psd  $\widehat{S}$  but propagates through the original medium without propagation correction. For variations in the Earth model parameters  $\delta\mathbf{m}$  the resulting variation of  $\bar{I}$  follows from the product rule:

$$\delta \bar{I}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} \int_{\oplus} \delta G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' + \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) \delta G^*(\mathbf{x}_k, \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}'. \quad (43)$$

The appearance of the Green function variation  $\delta G$  makes eq. (43) difficult to tackle numerically because it cannot be computed efficiently for all possible  $\delta\mathbf{m}$ , unless  $G$  takes a simple analytical form. To eliminate the variation of the Green function from eq. (43), we use the relation

$$\delta G(\mathbf{x}, \xi) = - \int_{\oplus} G^\dagger(\mathbf{x}'', \mathbf{x}) [\delta \mathcal{L}_{\mathbf{x}''} G(\mathbf{x}'', \xi)] d\mathbf{x}'', \quad (44)$$

derived in Appendix B1. In eq. (44),  $\mathcal{L}_{\mathbf{x}''}$  symbolizes the frequency-domain, acoustic wave equation operator defined in (1), or any other operator governing the propagation of a scalar wavefield. Its variation with respect to the model parameters  $\mathbf{m} = (\kappa, \rho)$  is denoted by  $\delta \mathcal{L}_{\mathbf{x}''}$ . The adjoint Green function is denoted by  $G^\dagger$ . Substituting eq. (44) into eq. (43), we obtain

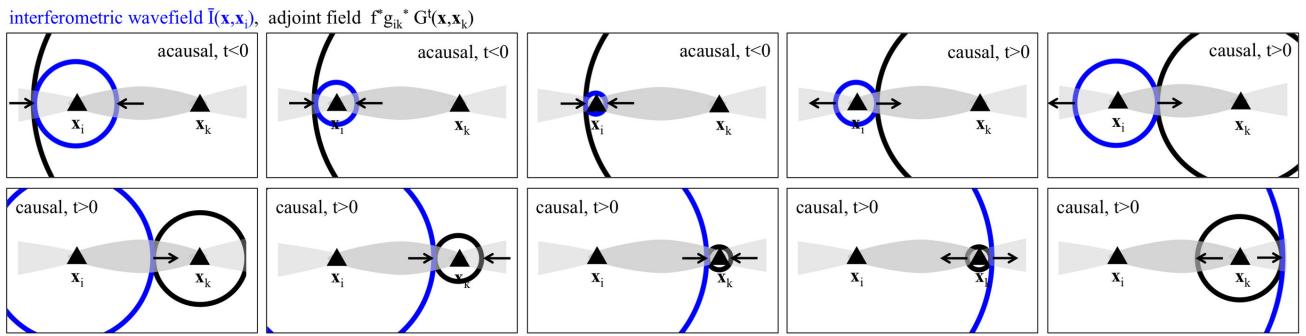
$$\begin{aligned} \delta \bar{I}(\mathbf{x}_i, \mathbf{x}_k) &= - \int_{\oplus} \int_{\oplus} \int_{\oplus} G^\dagger(\mathbf{x}'', \mathbf{x}_i) [\delta \mathcal{L}_{\mathbf{x}''} G(\mathbf{x}'', \mathbf{x})] G^*(\mathbf{x}_k, \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' d\mathbf{x}'' \\ &\quad - \int_{\oplus} \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}'', \mathbf{x}_k) [\delta \mathcal{L}_{\mathbf{x}''}^* G^*(\mathbf{x}'', \mathbf{x}')] \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' d\mathbf{x}''. \end{aligned} \quad (45)$$

Realizing that all functions that are independent of  $\mathbf{x}''$  are constants for the differential operator  $\delta \mathcal{L}_{\mathbf{x}''}$ , we can rearrange eq. (45) as follows:

$$\begin{aligned} \delta \bar{I}(\mathbf{x}_i, \mathbf{x}_k) &= - \int_{\oplus} G^\dagger(\mathbf{x}'', \mathbf{x}_i) \delta \mathcal{L}_{\mathbf{x}''} \left[ \int_{\oplus} \int_{\oplus} G(\mathbf{x}'', \mathbf{x}) G^*(\mathbf{x}_k, \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' \right] d\mathbf{x}'' \\ &\quad - \int_{\oplus} G^*(\mathbf{x}'', \mathbf{x}_k) \delta \mathcal{L}_{\mathbf{x}''}^* \left[ \int_{\oplus} \int_{\oplus} G(\mathbf{x}_i, \mathbf{x}) G^*(\mathbf{x}'', \mathbf{x}') \widehat{S}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' \right] d\mathbf{x}''. \end{aligned} \quad (46)$$

A comparison with eq. (42) shows that the square brackets in eq. (46) contain copies of the interferometric wavefield  $\bar{I}$ , which condenses the expression for  $\delta \bar{I}(\mathbf{x}_i, \mathbf{x}_k)$  to

$$\delta \bar{I}(\mathbf{x}_i, \mathbf{x}_k) = - \int_{\oplus} G^\dagger(\mathbf{x}, \mathbf{x}_i) \delta \mathcal{L}_{\mathbf{x}} \bar{I}(\mathbf{x}, \mathbf{x}_k) d\mathbf{x} - \left[ \int_{\oplus} G^\dagger(\mathbf{x}, \mathbf{x}_k) \delta \mathcal{L}_{\mathbf{x}} \bar{I}(\mathbf{x}, \mathbf{x}_i) d\mathbf{x} \right]^*. \quad (47)$$



**Figure 12.** Schematic illustration of the construction of a partial Fréchet kernel for structural parameters  $\mathbf{m}$  through the interaction of the interferometric field with focus  $\mathbf{x}_i$  and the adjoint field with source  $\mathbf{x}_k$ , that is, the second term in eqs (48), (51) and (52). Starting at negative times, the interferometric field  $\bar{I}(\mathbf{x}, \mathbf{x}_k)$  collapses into the receiver position  $\mathbf{x}_i$  at time  $t = 0$  and then propagates outwards for increasing times  $t > 0$ . The adjoint field  $f^* g_{ik}^* \nabla G^\dagger(\mathbf{x}, \mathbf{x}_k)$  originates from the other receiver position,  $\mathbf{x}_k$ , and may also extend into negative times, depending on the measurement. The partial kernel, shown in grey, is non-zero where the interferometric field and the adjoint field overlap. A second partial kernel, corresponding to the first term in eqs (48), (51) and (52), results from the interaction of the interferometric field  $\bar{I}(\mathbf{x}, \mathbf{x}_k)$  with focus at  $\mathbf{x}_k$  and the adjoint field  $f^* g_{ik}^* \nabla G^\dagger(\mathbf{x}, \mathbf{x}_i)$  with source at  $\mathbf{x}_i$ . The complete kernel is the sum of the partial kernels. The detailed shape of the hyperbolic features extending outwards from the receivers, and the parabolic cigar connecting the receivers, depends on the distribution of the effective sources and the measurement details. [NR]

Combining eqs (35) and (47), we obtain the following expression for the variation of the misfit functional  $\chi$ :

$$\delta\chi = -2\text{Re} \int_{\omega=0}^{\infty} \int_{\oplus} f g_{ik} G^\dagger(\mathbf{x}, \mathbf{x}_i) \delta\mathcal{L}_x \bar{I}(\mathbf{x}, \mathbf{x}_k) d\mathbf{x} - 2\text{Re} \int_{\omega=0}^{\infty} \int_{\oplus} f^* g_{ik}^* G^\dagger(\mathbf{x}, \mathbf{x}_k) \delta\mathcal{L}_x \bar{I}(\mathbf{x}, \mathbf{x}_i) d\mathbf{x}. \quad (48)$$

The quantities  $f g_{ik} G^\dagger(\mathbf{x}, \mathbf{x}_i)$  and  $f^* g_{ik}^* G^\dagger(\mathbf{x}, \mathbf{x}_k)$  constitute adjoint wavefields excited by adjoint sources at the receiver locations  $\mathbf{x}_i$  and  $\mathbf{x}_k$ , respectively. The time evolution of the adjoint sources is determined by the adjoint source  $f$ , that is, by the definition of the misfit or measurement functional  $\chi$ . The adjoint fields interact with the interferometric wavefields  $\bar{I}(\mathbf{x}, \mathbf{x}_k)$  and  $\bar{I}(\mathbf{x}_i, \mathbf{x})$  with foci at the other receiver, that is,  $\mathbf{x}_k$  for the adjoint field  $f g_{ik} G^\dagger(\mathbf{x}, \mathbf{x}_i)$ , and  $\mathbf{x}_i$  for the adjoint field  $f^* g_{ik}^* G^\dagger(\mathbf{x}, \mathbf{x}_k)$ . Since  $\delta\mathcal{L}_x$  is linear in the model perturbation  $\delta\mathbf{m}$ , eq. (48) can generally be written in the form

$$\delta\chi = \int_{\oplus} \mathbf{K}(\mathbf{x}) \cdot \delta\mathbf{m} d\mathbf{x}, \quad (49)$$

where  $\mathbf{K}$  is the Fréchet or sensitivity kernel. For the specific case of the acoustic wave eq. (1) with fractional perturbations  $\delta \ln \kappa$  and  $\delta \ln \rho$  we find

$$\delta\chi = \int_{\oplus} K_\kappa(\mathbf{x}) \delta \ln \kappa(\mathbf{x}) d\mathbf{x} + \int_{\oplus} K_\rho(\mathbf{x}) \delta \ln \rho(\mathbf{x}) d\mathbf{x}, \quad (50)$$

with the bulk modulus kernel

$$K_\kappa(\mathbf{x}) = -2\text{Re} \int_{\omega=0}^{\infty} \frac{\omega^2}{\kappa} f g_{ik} G^\dagger(\mathbf{x}, \mathbf{x}_i) \bar{I}(\mathbf{x}, \mathbf{x}_k) d\omega - 2\text{Re} \int_{\omega=0}^{\infty} \frac{\omega^2}{\kappa} f^* g_{ik}^* G^\dagger(\mathbf{x}, \mathbf{x}_k) \bar{I}(\mathbf{x}, \mathbf{x}_i) d\omega, \quad (51)$$

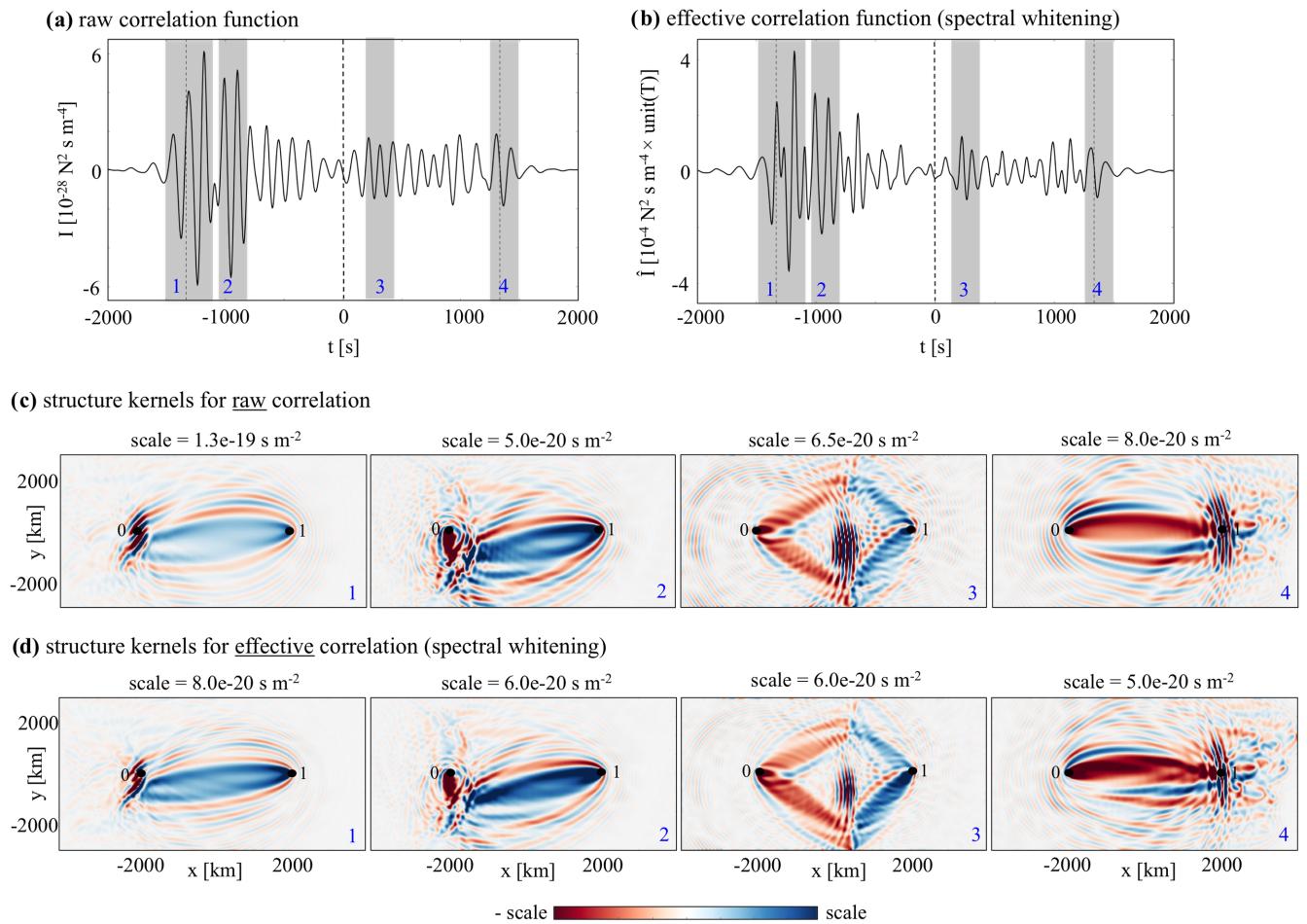
and the density kernel

$$K_\rho(\mathbf{x}) = 2\text{Re} \int_{\omega=0}^{\infty} \frac{1}{\rho} f g_{ik} \nabla G^\dagger(\mathbf{x}, \mathbf{x}_i) \cdot \nabla \bar{I}(\mathbf{x}, \mathbf{x}_k) d\omega + 2\text{Re} \int_{\omega=0}^{\infty} \frac{1}{\rho} f^* g_{ik}^* \nabla G^\dagger(\mathbf{x}, \mathbf{x}_k) \cdot \nabla \bar{I}(\mathbf{x}, \mathbf{x}_i) d\omega. \quad (52)$$

The correlation-based kernels for Earth structure are similar to those for earthquake-based measurements that have been studied extensively (e.g. Yomogida 1992; Dahlen *et al.* 2000; Friederich 2003; Sieminski *et al.* 2007). They are constructed through the interaction of a forward wavefield (the wavefield excited by the earthquake or the interferometric wavefield) and an adjoint source that propagates data residuals back into the medium. In contrast to the earthquake scenario, kernels for correlation measurements have two contributions. For one of them  $\mathbf{x}_i$  is the reference station of the interferometric wavefield, and  $\mathbf{x}_k$  is the position where the adjoint source acts. For the other contribution, the roles of  $i$  and  $k$  are interchanged. Provided that the processing is chosen such that  $g_{ik} = g_{ik}^*$ , the kernel is symmetric with respect to an interchange of  $i$  and  $k$ . A schematic illustration of the kernel construction is provided in Fig. 12.

#### 4.2.1 Example I: structure kernels for raw and processed correlations

To establish a clear thread, we illustrate the computation of bulk modulus kernels  $K_\kappa$  with the same source-receiver setup used earlier in Section 3.1 on processing and Section 4.1.2 on source kernels. We again consider station pair  $(\mathbf{x}_0, \mathbf{x}_1)$ , and for simplicity we assume that the



**Figure 13.** (a,b) Raw and processed (spectrally whitened) correlation functions, and (c,d) the corresponding structure kernels  $K_\kappa$  for different time windows indicated by grey shading and blue numbers from 1 to 4. [ER]

effective psd is frequency-independent. As measurement we choose cross-correlation traveltime differences (Luo & Schuster 1991) for which the adjoint source is derived in Appendix B2.1. A collection of sensitivity kernels for the same time windows as in Fig. 10 is presented in Fig. 13.

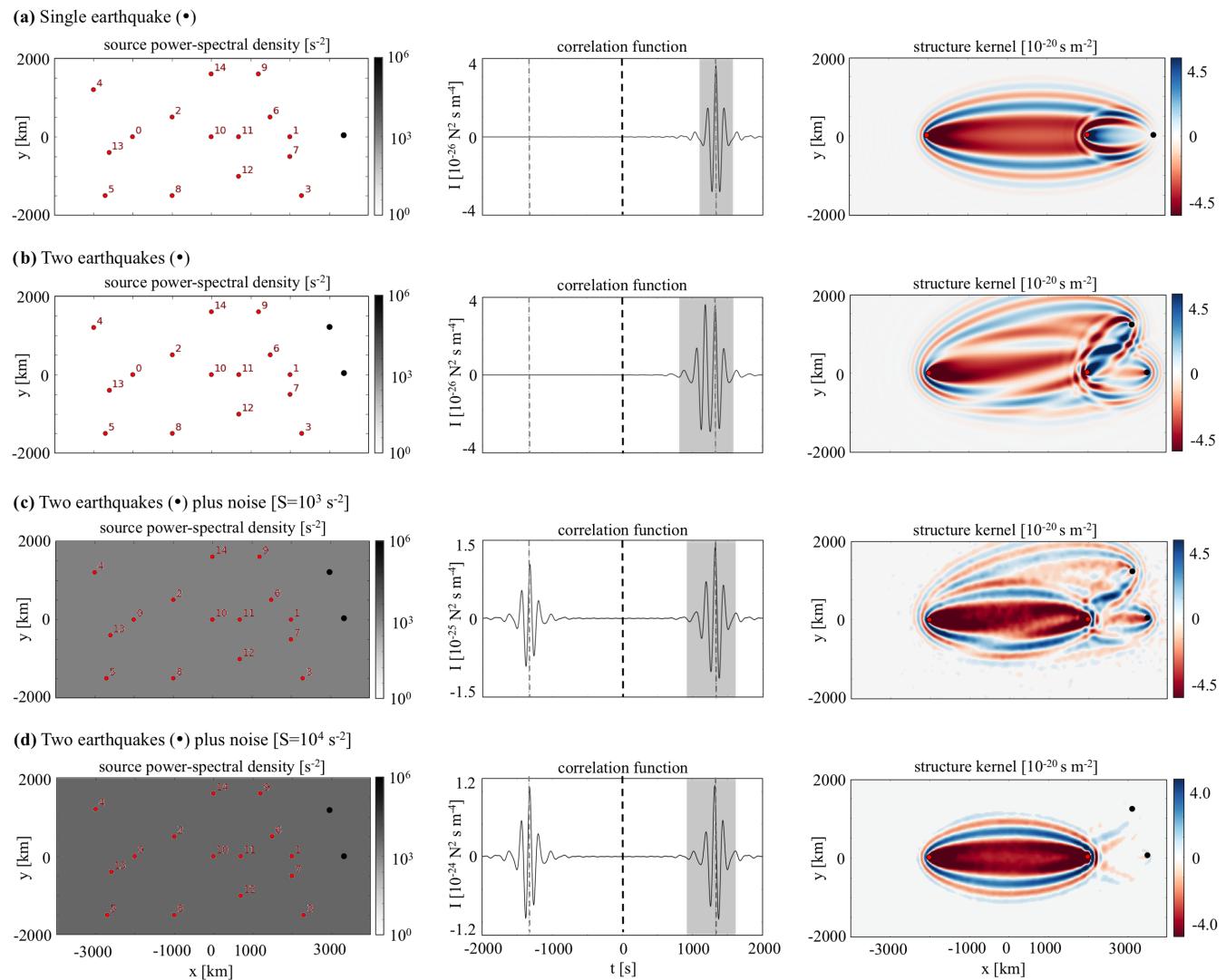
The kernels for time windows 1 and 4 around the raw (unprocessed) ballistic arrivals are dominated by a well-defined first Fresnel zone that roughly links the receivers, and resembles well-known finite-frequency kernels for earthquake tomography (e.g. Yomogida 1992; Dahlen *et al.* 2000; Dahlen & Baig 2002; Friederich 2003; Yoshizawa & Kennett 2004; Zhou *et al.* 2004). In accord with physical intuition, sensitivity in the first Fresnel zone for the causal time window 4 is negative. Increasing  $\kappa$  increases the velocity  $\sqrt{\kappa/\rho}$ , which leads to a reduced arrival time. For the acausal time window 1, sensitivity in the first Fresnel zone is positive because an increasing  $\kappa$  will lead to an increased arrival time, that is, an arrival time closer to 0.

In contrast to interferometry based on Green function retrieval, we can also extract useful information from non-ballistic arrivals that result from a heterogeneous source distribution, for example, the arrivals in time windows 2 and 3 in Fig. 13. The kernel for time window 3 deviates most from the comparatively simple kernels for ballistic waves. It extends from the receivers towards those regions where noise sources can contribute to a signal near zero lag time, that is, roughly perpendicular to the line connecting the receivers. The arrival in window 3 illuminates the model in areas to which the ballistic waves are blind.

While processing in the form of spectral whitening profoundly changes the source kernels in Fig. 10, its effects on the structure kernels are more subtle. Spectral whitening emphasizes higher frequencies, which reduces the width of the first Fresnel zone and introduces additional short-wavelength details in the kernels. Depending on the time window, spectral whitening changes the amplitudes of the kernel by 10–60 per cent. If not accounted for correctly, these are large enough to prevent the convergence of iterative optimization schemes such as quasi-Newton and L-BFGS (e.g. Nocedal & Wright 1999; Boehm *et al.* 2016).

#### 4.2.2 Example II: structure kernels for combined earthquake and noise correlations

In our developments, we have made no limiting assumptions on the nature of the sources  ${}_nF$  that excite the wavefield. They can be localized earthquakes, distributed noise sources, or combinations of both.



**Figure 14.** Structure kernels for (combined) earthquake and noise correlations. Cross-correlation traveltimes are measured in the grey-shaded time window. (a) Correlation of waveforms from a single earthquake, indicated by the black dot, leads to a one-sided correlation function. The corresponding kernel appears as the superposition of two kernels, extending from the earthquake to each of the two receivers. (b) Correlation of waveforms from two earthquakes leads to a more complex, one-sided, correlation function. The kernel connects both earthquakes to both receivers. (c,d) As the psd of the distributed noise sources increases, the correlation function becomes more symmetric, and the contribution of the earthquakes becomes less dominant. In the noise dominated scenario (d), the correlation function is nearly perfectly symmetric and the kernel simplifies to the classical cigar shape between the receiver pair. [ER]

In the special case of a single earthquake located in-line with the receiver pair, as illustrated in Fig. 14(a), we obtain a finite-frequency variant of the two-station method, previously studied by Chevrot & Zhao (2007), Kuo *et al.* (2009) and de Vos *et al.* (2013). This generalizes the traditional two-station method where information is assumed to propagate along infinitesimally thin rays along the great-circle path connecting source and receivers (e.g. Meier *et al.* 2004; Zhang *et al.* 2007; Endrun *et al.* 2008). The effective psd  $\hat{S}$ , shown in the left panel of Fig. 14(a) is point-localized at the earthquake position, which leads to a strictly one-sided correlation function. Measuring cross-correlation traveltimes, produces a sensitivity kernel that appears as the superposition of two individual kernels, each connecting a receiver to the source. Strong sensitivity is present between the source and the first receiver, thereby invalidating one of the basic assumptions of the ray-based two-station method.

As illustrated in Fig. 14(b), several earthquakes may be included. Their position is not restricted to the great-circle path that may be a poor approximation to the actual wave propagation path in 3-D heterogeneous media (Pedersen 2006; Pedersen *et al.* 2015). The use of earthquakes located off the great-circle, anticipated by Passier & Snieder (1995), has the potential to significantly increase the amount of exploitable information.

The combination of earthquakes and distributed noise sources, shown in Figs 14(c) and (d), may serve the same purpose. Instead of disregarding information from noise (as in the earthquake two-station method) or earthquakes (as in ambient noise interferometry), both types of information may be combined in order to maximize constraints on Earth structure. The relative importance of earthquake and noise data can be steered through the processing, that is, through the source correctors  $n_f$ .

## 5 EXTENSION TO ELASTIC WAVE PROPAGATION

Using the concepts introduced for the acoustic case, we now transition to elastic wave propagation and vector-valued wavefields. Following the development of an effective forward modelling theory in Section 5.1, we will derive sensitivity kernels for the effective source and structure in Section 5.2.

### 5.1 Effective forward modelling theory for the processed correlation tensor

#### 5.1.1 The unprocessed correlation tensor

In the elastic case, the frequency-domain vectorial displacement field at position  $\mathbf{x}_i$  in time interval  $n$  is given by the representation theorem

$${}_{\text{n}}\mathbf{u}(\mathbf{x}_i) = \int_{\oplus} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) {}_n\mathbf{F}(\mathbf{x}) d\mathbf{x}, \quad {}_n\mu_{\alpha}(\mathbf{x}_i) = \int_{\oplus} G_{\alpha\beta}(\mathbf{x}_i, \mathbf{x}) {}_nF_{\beta}(\mathbf{x}) d\mathbf{x}, \quad (53)$$

where  $\mathbf{G}$  is the Green tensor with components  $G_{\alpha\beta}$  for the chosen Earth model, and  ${}_n\mathbf{F}$  denotes the external force with vector components  ${}_nF_{\beta}$  during time interval  $n$ . To avoid clutter, we use the invariant (vector/tensor) notation whenever possible. From eq. (53), we deduce the correlation or interferometric tensor  ${}_{\text{n}}\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  for time interval  $n$ :

$${}_{\text{n}}\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k) = {}_{\text{n}}\mathbf{u}(\mathbf{x}_i) {}_{\text{n}}\mathbf{u}^H(\mathbf{x}_k) = \int_{\oplus} \int_{\oplus} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) {}_n\mathbf{F}(\mathbf{x}) [\mathbf{G}(\mathbf{x}_k, \mathbf{x}') {}_n\mathbf{F}(\mathbf{x}')]^H d\mathbf{x} d\mathbf{x}', \quad (54)$$

where the super-script  $H$  denotes Hermitian conjugation, that is, the complex conjugate of the transpose. Grouping together the source terms in eq. (55) and using spatial Green function reciprocity, yields

$${}_{\text{n}}\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} \int_{\oplus} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) [\mathbf{S}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k)] d\mathbf{x} d\mathbf{x}', \quad (55)$$

with the psd tensor for time interval  $n$

$${}_{\text{n}}\mathbf{S}(\mathbf{x}, \mathbf{x}') = {}_n\mathbf{F}(\mathbf{x}) {}_n\mathbf{F}^H(\mathbf{x}'). \quad (56)$$

The diagonal elements of  $\mathbf{S}$  describe the temporal correlation of neighbouring forces acting in identical direction. Similarly, the off-diagonal elements encode the temporal correlation of forces in different directions. While elaborate physical models for vertical-component noise sources and the resulting wave propagation already exist (Stutzmann *et al.* 2012; Gualtieri *et al.* 2013; Ardhuin *et al.* 2015; Farra *et al.* 2016; Gimbert & Tsai 2016), the nature of horizontal-component noise sources is less certain and still the subject of observational studies (Kurle & Widmer-Schnidrig 2006; Hadzioannou *et al.* 2012). Averaging over all time intervals, gives the raw (unprocessed) ensemble correlation tensor  $\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N {}_{\text{n}}\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  in the form

$$\underbrace{\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{interf. wavefield}} = \int_{\oplus} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) \underbrace{\left[ \int_{\oplus} \mathbf{S}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x}' \right]}_{\text{source of interf. wavefield}} d\mathbf{x}, \quad (57)$$

where  $\mathbf{S}(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_{\text{n}}\mathbf{S}(\mathbf{x}, \mathbf{x}')$  is the time-averaged psd of the sources. Eq. (57) is the elastic analogue of the scalar acoustic forward modelling eq. (8), and also has the form of a representation theorem. Taking the receiver position  $\mathbf{x}_i$  as a free variable, we can interpret  $\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  as a tensorial interferometric wavefield that is driven by the source  $\int \mathbf{S}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x}'$ . Under the assumption that neighbouring sources are on average uncorrelated in time in the sense of

$$\mathbf{S}(\mathbf{x}, \mathbf{x}') = \mathbf{S}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}'), \quad (58)$$

the forward modelling eq. (57) simplifies to

$$\underbrace{\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{interf. wavefield}} = \int_{\oplus} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) \underbrace{[\mathbf{S}(\mathbf{x}) \mathbf{G}^*(\mathbf{x}, \mathbf{x}_k)]}_{\text{source of interf. wavefield}} d\mathbf{x}, \quad (59)$$

meaning that the source of the interferometric wavefield can be computed without solving a space integral. Eq. (59) is the elastic analogue of the acoustic eq. (10). While most applications focus on the vertical-vertical component of the interferometric tensor, deterministic signals can emerge on other components as well (e.g. Campillo & Paul 2003; Stehly *et al.* 2007), and may be used to investigate seismic anisotropy and its temporal variations (e.g. Roux 2009; Durand *et al.* 2011; Riahi *et al.* 2013).

The raw, synthetic interferometric tensor  $\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  can be compared to the raw, observed interferometric tensor computed from the raw, observed, 3-component displacement recordings  ${}^o\mathbf{u}$  in time interval  $n$ :

$$\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N \mathbf{I}(\mathbf{x}_i, \mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^N {}^o\mathbf{u}(\mathbf{x}_i) {}^o\mathbf{u}^H(\mathbf{x}_k) \quad (60)$$

Provided that  $\mathbf{G}$  is the Green function of a fundamental theory in the sense of Section 2.1, there exist Earth models  $\mathbf{m}$  and source models  $\mathbf{S}$  within the space of acceptably data-fitting models  $\mathbb{P}_{\text{fit}}$  such that  ${}_n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  is equal to  ${}^n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  to within the observational uncertainties for all time intervals, that is,

$${}^n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k) = {}_n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k; \mathbf{p}) + {}_n\boldsymbol{\epsilon}_{ik}, \quad \mathbf{p} = (\mathbf{m}, \mathbf{S}) \in \mathbb{P}_{\text{fit}}. \quad (61)$$

For most seismological applications, the fundamental theory is the seismic wave equation. Effects such as attenuation, anisotropy, the Earth's rotation and self-gravitation will need to be included, depending on the requirements of a specific data set.

### 5.1.2 Admissible effective theory for processed correlations

Since the raw, observed interferometric tensor  ${}^n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  may not be well-suited for the solution of inverse problems, various processing steps are usually applied, for instance to emphasize specific seismic phases or accelerate convergence. (See Section 1.2 for a brief summary of possible options.) We thus consider a processed, observed interferometric tensor  $\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$ . The product of each individual  ${}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$  with the inverse of the corresponding  ${}^n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)$  defines the tensorial transfer coefficient or transfer matrix  ${}_n\mathbf{T}_{ik} = {}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) {}^n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k)^{-1}$  that relates raw and processed interferograms for time interval  $n$ :

$${}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = {}_n\mathbf{T}_{ik} {}^n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k). \quad (62)$$

Following the arguments introduced in Section 2.4, we can use eqs (61) and (62) to define an admissible theory for the forward modelling of processed interferograms. Indeed, multiplying eq. (61) with the transfer matrix, yields

$${}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = {}_n\mathbf{T}_{ik} {}_n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k; \mathbf{p}) + {}_n\tilde{\boldsymbol{\epsilon}}_{ik}, \quad \mathbf{p} \in \mathbb{P}_{\text{fit}}, \quad (63)$$

with the processed observational uncertainties  ${}_n\tilde{\boldsymbol{\epsilon}}_{ik} = {}_n\mathbf{T}_{ik} {}_n\boldsymbol{\epsilon}_{ik}$ . Defining the processed, synthetic interferograms as

$$\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k; \text{any } \mathbf{p}) = {}_n\mathbf{T}_{ik} {}_n\mathbf{I}(\mathbf{x}_i, \mathbf{x}_k; \text{any } \mathbf{p}) \quad (64)$$

ensures that the same suite of Earth models  $\mathbf{m}$  and source models  $\mathbf{S}$  that satisfy (61) also satisfy

$${}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = {}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k; \mathbf{p}) + {}_n\tilde{\boldsymbol{\epsilon}}_{ik}, \quad \mathbf{p} \in \mathbb{P}_{\text{fit}}. \quad (65)$$

Our forward modelling theory for processed interferograms is thus admissible in the sense of Section 2.4. Taking inspiration from eq. (13), we factorize the transfer matrix as

$${}_n\mathbf{T}_{ik} = \mathbf{g}_{ik} {}_nf + {}_n\boldsymbol{\epsilon}_{ik}, \quad (66)$$

with  $\mathbf{g}_{ik}$  the tensorial *propagation corrector*,  ${}_nf$  the scalar *source corrector* and  ${}_n\boldsymbol{\epsilon}_{ik}$  the tensorial factorization residual. In Appendix A2, we demonstrate how the factorization (66) can be constructed such that the factorization residual is minimized. Combining eqs (55), (64) and (66) yields a forward modelling equation for the processed ensemble interferometric tensor:

$$\underbrace{{}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{proc. interf. wavefield}} = \int \underbrace{\mathbf{g}_{ik} \mathbf{G}(\mathbf{x}_i, \mathbf{x})}_{\oplus \text{ eff. Green function}} \underbrace{\left[ \int \widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x}' \right]}_{\text{effective source}} d\mathbf{x} + \mathbf{E}(\mathbf{x}_i, \mathbf{x}_k) = \underbrace{{}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{effective interf. wavefield}} + \mathbf{E}(\mathbf{x}_i, \mathbf{x}_k). \quad (67)$$

Eq. (67) is the elastic counterpart of the acoustic effective forward modelling eq. (24). It contains the *effective Green function*  $\mathbf{g}_{ik} \mathbf{G}(\mathbf{x}_i, \mathbf{x})$  for the station path  $\mathbf{x}_k \rightarrow \mathbf{x}_i$ , the *effective source psd*

$$\widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{n=1}^N {}_nf {}_n\mathbf{S}(\mathbf{x}, \mathbf{x}'), \quad (68)$$

and the *forward modelling error*

$$\mathbf{E}(\mathbf{x}_i, \mathbf{x}_k) = \int \int \mathbf{G}(\mathbf{x}_i, \mathbf{x}) \left[ \frac{1}{N} \sum_{n=1}^N {}_n\mathbf{S}(\mathbf{x}, \mathbf{x}') \right] \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x} d\mathbf{x}'. \quad (69)$$

For completeness, we note that the forward modelling eq. (67) simplifies to

$$\underbrace{{}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{proc. interf. wavefield}} = \int \underbrace{\mathbf{g}_{ik} \mathbf{G}(\mathbf{x}_i, \mathbf{x})}_{\oplus \text{ eff. Green function}} \underbrace{[\widehat{\mathbf{S}}(\mathbf{x}) \mathbf{G}^*(\mathbf{x}, \mathbf{x}_k)]}_{\text{effective source}} d\mathbf{x} + \mathbf{E}(\mathbf{x}_i, \mathbf{x}_k) = \underbrace{{}_n\tilde{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)}_{\text{effective interf. wavefield}} + \mathbf{E}(\mathbf{x}_i, \mathbf{x}_k), \quad (70)$$

when the sources are on average uncorrelated. Eqs (67) and (70) constitute computable forward models for effective interferogram  $\hat{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$  that have the same structure as eqs (57) and (59) for their raw counterparts. The effective interferometric field propagates through an effective medium described by the effective Green function  $\mathbf{g}_{ik} \mathbf{G}(\mathbf{x}_i, \mathbf{x})$ . It is excited by an effective source that is determined by the effective psd  $\widehat{\mathbf{S}}$ .

The effective theory is perfect provided that processed interferograms equal effective interferograms, that is, when the forward modelling error  $\mathbf{E}(\mathbf{x}_i, \mathbf{x}_k)$  or the factorization residuals  ${}_h\mathbf{e}_{ik}$  vanish. Forward modelling errors that are large compared to the observational uncertainties indicate that the processed interferograms cannot be adequately represented as correlations, meaning that a fundamentally different forward modelling theory is required.

## 5.2 Fréchet kernels for sources and structure

The effective interferometric tensor  $\widehat{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$  is computable and can be compared to observations via a misfit or measurement functional  $\chi = \chi[\widehat{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)]$ . As in the scalar case of eq. (35), the variation of  $\chi$  induced by variations of  $\widehat{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$  can generally be written in the form of a scalar product with some tensor-valued function  $\mathbf{f}$ ,

$$\delta\chi = 2 \operatorname{Re} \int_{\omega=0}^{\infty} \delta\widehat{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) : \mathbf{f} d\omega = 2 \operatorname{Re} \int_{\omega=0}^{\infty} \mathbf{g}_{ik} \delta\bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) : \mathbf{f} d\omega, \quad (71)$$

where  $\widehat{\mathbf{I}} : \mathbf{f} = \widehat{I}_{\alpha\beta} f_{\alpha\beta}$  symbolizes the double contraction with implied summation over repeated indices. The function  $\mathbf{f}$  is determined by the specific choice of  $\chi$ . Examples are provided in Appendix B2. To simplify notation in the derivation of Fréchet kernels for structure in Section 5.2.2, we introduced in eq. (71) the interferometric wavefield

$$\bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) \left[ \int_{\oplus} \widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x}' \right] d\mathbf{x} = \mathbf{g}_{ik}^{-1} \widehat{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) \quad (72)$$

which is excited by the effective source but propagates through the original medium.

### 5.2.1 Source kernels

In the specific case where the variation of  $\chi$  is induced by a variation in the effective psd,  $\delta\widehat{\mathbf{S}}$ , we can write  $\delta\widehat{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$  explicitly as

$$\delta\bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) \left[ \int_{\oplus} \delta\widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x}' \right] d\mathbf{x}. \quad (73)$$

Inserting eq. (73) into the generic misfit variation (71) yields

$$\delta\chi = 2 \operatorname{Re} \int_{\omega=0}^{\infty} \int_{\oplus} \int_{\oplus} [\mathbf{g}_{ik} \mathbf{G}(\mathbf{x}_i, \mathbf{x}) \delta\widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k)] : \mathbf{f} d\mathbf{x}' d\mathbf{x} d\omega. \quad (74)$$

Rearranging the matrix products in eq. (74) gives the elastic analogues of eqs (37) and (38),

$$\delta\chi = \int_{\omega=0}^{\infty} \int_{\oplus} \int_{\oplus} \mathbf{K}(\mathbf{x}, \mathbf{x}') : \delta\widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') d\mathbf{x}' d\mathbf{x} d\omega, \quad (75)$$

with the space- and frequency-dependent source kernel

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = 2 \operatorname{Re} [\mathbf{g}_{ik} \mathbf{G}(\mathbf{x}_i, \mathbf{x})]^T [\mathbf{f} \mathbf{G}^*(\mathbf{x}_k, \mathbf{x}')]. \quad (76)$$

Eq. (76) can be interpreted as the product of an effective Green function  $\mathbf{g}_{ik} \mathbf{G}(\mathbf{x}_i, \mathbf{x})$  between  $\mathbf{x}_i$  and  $\mathbf{x}$ , and an adjoint field  $\mathbf{f} \mathbf{G}^*(\mathbf{x}_k, \mathbf{x}')$  between  $\mathbf{x}_k$  and  $\mathbf{x}'$ . It may be further simplified by assuming spatially uncorrelated sources.

### 5.2.2 Structure kernels

To derive Fréchet kernels for structural perturbations  $\delta\mathbf{m}$ , we first compute the variation of  $\bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$ , introduced in eq. (72):

$$\delta\bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = \int_{\oplus} \int_{\oplus} \delta\mathbf{G}(\mathbf{x}_i, \mathbf{x}) \widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x} d\mathbf{x}' + \left[ \int_{\oplus} \int_{\oplus} \delta\mathbf{G}(\mathbf{x}_k, \mathbf{x}') \widehat{\mathbf{S}}(\mathbf{x}', \mathbf{x}) \mathbf{G}^*(\mathbf{x}, \mathbf{x}_i) d\mathbf{x} d\mathbf{x}' \right]^H. \quad (77)$$

Substituting for  $\delta\mathbf{G}$  eq. (B10) from Appendix B1, the expression for  $\delta\bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$  expands to

$$\begin{aligned} \delta\bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = & - \int_{\oplus} \int_{\oplus} \int_{\oplus} [\mathbf{G}^\dagger(\mathbf{x}_i, \mathbf{x}'') \delta\mathcal{L}_{\mathbf{x}''} \mathbf{G}(\mathbf{x}'', \mathbf{x})] \widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x} d\mathbf{x}' d\mathbf{x}'' \\ & - \left[ \int_{\oplus} \int_{\oplus} \int_{\oplus} [\mathbf{G}^\dagger(\mathbf{x}_k, \mathbf{x}'') \delta\mathcal{L}_{\mathbf{x}''} \mathbf{G}(\mathbf{x}'', \mathbf{x}')] \widehat{\mathbf{S}}(\mathbf{x}', \mathbf{x}) \mathbf{G}^*(\mathbf{x}, \mathbf{x}_i) d\mathbf{x} d\mathbf{x}' d\mathbf{x}'' \right]^H. \end{aligned} \quad (78)$$

Since the wave equation operator  $\mathcal{L}_{\mathbf{x}''}$  only acts on  $\mathbf{x}''$ , all variables that are independent of  $\mathbf{x}''$  are constants for  $\mathcal{L}_{\mathbf{x}''}$ , and we can rewrite eq. (78) as

$$\begin{aligned}\delta \bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = & - \int_{\oplus} \mathbf{G}^\dagger(\mathbf{x}_i, \mathbf{x}'') \delta \mathcal{L}_{\mathbf{x}''} \left[ \int_{\oplus} \int_{\oplus} \mathbf{G}(\mathbf{x}'', \mathbf{x}) \widehat{\mathbf{S}}(\mathbf{x}, \mathbf{x}') \mathbf{G}^*(\mathbf{x}', \mathbf{x}_k) d\mathbf{x} d\mathbf{x}' \right] d\mathbf{x}'' \\ & - \left[ \int_{\oplus} \mathbf{G}^\dagger(\mathbf{x}_k, \mathbf{x}'') \delta \mathcal{L}_{\mathbf{x}''} \left[ \int_{\oplus} \int_{\oplus} \mathbf{G}(\mathbf{x}'', \mathbf{x}') \widehat{\mathbf{S}}(\mathbf{x}', \mathbf{x}) \mathbf{G}^*(\mathbf{x}, \mathbf{x}_i) d\mathbf{x} d\mathbf{x}' \right] d\mathbf{x}'' \right]^H.\end{aligned}\quad (79)$$

Recognizing two copies of the interferometric wavefield  $\bar{I}$  inside the square brackets, eq. (79) condenses to

$$\delta \bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k) = - \int_{\oplus} \mathbf{G}^\dagger(\mathbf{x}_i, \mathbf{x}) \delta \mathcal{L}_{\mathbf{x}} \bar{I}(\mathbf{x}, \mathbf{x}_k) d\mathbf{x} - \left[ \int_{\oplus} \mathbf{G}^\dagger(\mathbf{x}_k, \mathbf{x}) \delta \mathcal{L}_{\mathbf{x}} \bar{I}(\mathbf{x}, \mathbf{x}_i) d\mathbf{x} \right]^H. \quad (80)$$

It remains to insert  $\delta \bar{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_k)$  into the canonical form (71) for the misfit or measurement variation:

$$\delta \chi = -2\text{Re} \int_{\omega=0}^{\infty} \int_{\oplus} [\mathbf{g}_{ik} \mathbf{G}^\dagger(\mathbf{x}_i, \mathbf{x}) \delta \mathcal{L}_{\mathbf{x}} \bar{I}(\mathbf{x}, \mathbf{x}_k)] : \mathbf{f} d\mathbf{x} d\omega - 2\text{Re} \int_{\omega=0}^{\infty} \int_{\oplus} [\mathbf{g}_{ik} \mathbf{G}^\dagger(\mathbf{x}_k, \mathbf{x}) \delta \mathcal{L}_{\mathbf{x}} \bar{I}(\mathbf{x}, \mathbf{x}_i)] : \mathbf{f}^H d\mathbf{x} d\omega. \quad (81)$$

Since the variation of the wave equation operator  $\delta \mathcal{L}_{\mathbf{x}}$  is by definition linear in  $\delta \mathbf{m}$ , it may be written in terms of a sensitivity kernel  $\mathbf{K}(\mathbf{x})$ ,

$$\delta \chi = \int_{\oplus} \mathbf{K}(\mathbf{x}) \cdot \delta \mathbf{m}(\mathbf{x}) d\mathbf{x}. \quad (82)$$

where the form of  $\mathbf{K}(\mathbf{x})$  depends on the choice of the specific wave equation and its parametrization (e.g. Igel *et al.* 1996; Tromp *et al.* 2005; Fichtner 2010). Scalarizing eq. (82), reproduces eq. (48) for the measurement variation  $\delta \chi$  in the acoustic case.

## 6 DISCUSSION

In the following paragraphs we discuss practical and conceptual issues, such as the unavoidable joint inversion for sources and structure, the relation between the nonlinear transfer function and the transfer coefficients, the design of optimal processing schemes that eliminate forward modelling errors, and those cases where the transfer coefficient is not properly defined because of division by zero.

### 6.1 Joint inversion for and trade-offs between sources and structure

Ambient noise interferometry based on Green function retrieval rests on the assumption of homogeneously distributed sources (Wapenaar 2004; Wapenaar & Fokkema 2006), thereby seemingly eliminating the need to constrain the wavefield sources prior to an inversion for Earth structure. Generalized interferometry, as proposed in the previous paragraphs, does not enjoy the same (apparent) luxury. Since the details of correlation waveforms depend on both the source distribution and Earth structure, any inversion must constrain them jointly. As in earthquake tomography, failure to invert for the source properties will lead to incorrect inferences on Earth structure (Valentine & Woodhouse 2010), and vice versa.

Source-structure trade-offs are likely to be most important when scattering is dominant, that is, for high frequencies and long propagation distances. Tomographic inversions typically avoid this regime by design because (multiply) scattered waves are much harder to fit than transmitted waves at lower frequencies and shorter propagation paths. In fact, noise tomography has partly been developed precisely in order to achieve short propagation paths and perform structure inversions in transmission mode.

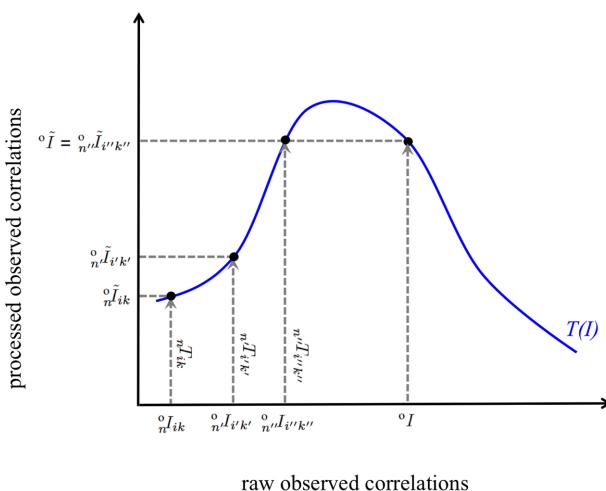
Practical methods for a joint source-structure inversion do currently not exist, but are in the process of being developed (Hanasoge 2013; Delaney *et al.* 2016; Ermert *et al.* 2016). A key element of these developments is and will be the design of misfit functionals that decouple the inverse problem as much as possible. Similar to earthquake tomography, trade-offs between Earth structure and wavefield sources will always remain. They can be quantified using, for instance, resolution analysis based on second-order adjoints (Fichtner 2015).

### 6.2 The transfer function, transfer coefficients and (non-)uniqueness

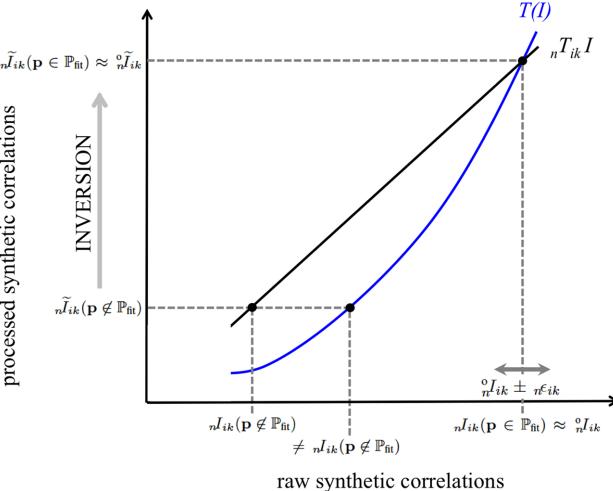
The frequency-dependent transfer coefficients  ${}_n T_{ik}$  are pointwise realizations of the transfer function  $T$  that maps a *generic* observed raw interferogram  ${}^0 I$  into its processed version  ${}^0 \tilde{I} = T({}^0 I)$ , as illustrated in Fig. 15(a). The transfer function is defined through a processing algorithm that can be applied to any raw data. For a *specific* observation  ${}^0 I_{ik}$ , we have  ${}^0 \tilde{I}_{ik} = T({}^0 I_{ik}) = {}_n T_{ik} {}^0 I_{ik}$ , where we used the subscript  $ik$  as a short-hand notation for  $(\mathbf{x}_i, \mathbf{x}_k)$ . While the transfer function  $T$  may map different  ${}^0 I$  into the same  ${}^0 \tilde{I}$ , the transfer coefficients  ${}_n T_{ik}$  uniquely link specific pairs of  ${}^0 I_{ik}$  and  ${}^0 \tilde{I}_{ik}$ , by construction.

The relation between observations and synthetics is illustrated in Fig. 15(b). When  $\mathbf{p}$  is part of the approximately data-fitting models  $\mathbb{P}_{\text{fit}}$ , that is,  ${}^0 \tilde{I}_{ik}(\mathbf{p} \in \mathbb{P}_{\text{fit}}) \approx {}^0 I_{ik}$ , the transfer coefficients directly link raw observations and synthetics. Since  ${}^0 \tilde{I}_{ik} = {}_n T_{ik} {}^0 I_{ik} = T({}^0 I_{ik})$ , and

(a) transfer function and transfer coefficients

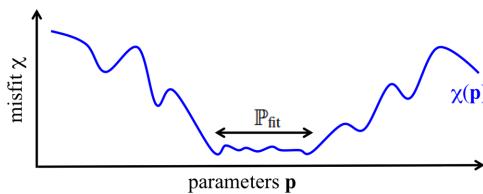


(b) linking observations and synthetics

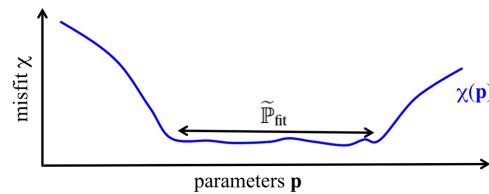


**Figure 15.** (a) Transfer coefficients are pointwise realizations of the transfer function, for example, for specific observations  $n\tilde{I}_{ik}$ ,  $n\tilde{I}_{i'k'}$ , or  $n\tilde{I}_{i''k''}$ . The transfer function  $T$ , shown as blue curve, maps a generic raw correlation  $^oI$  into a processed correlation  $^o\tilde{I}$ . It may map different  $^oI$  into the same  $^o\tilde{I}$ , thereby inducing additional non-uniqueness into the inverse problem. (b) Relation between observed and synthetics. The linear action of the transfer coefficients  $nT_{ik}$  on a raw correlation  $I$  is plotted as a black line. The potentially non-linear transfer function is plotted in blue. For data-fitting models  $\mathbf{p} \in \mathbb{P}_{\text{fit}}$ , the approximation of processed observations by processed synthetics,  $n\tilde{I}_{ik}(\mathbf{p} \in \mathbb{P}_{\text{fit}}) \approx n\tilde{I}_{ik}$  implies the approximation of raw observations by raw synthetics,  $nI_{ik}(\mathbf{p} \in \mathbb{P}_{\text{fit}}) \approx nI_{ik}$  because the transfer coefficients and the transfer function produce identical results. For non-data-fitting models  $\mathbf{p} \notin \mathbb{P}_{\text{fit}}$ , the transfer coefficients and the transfer function are not generally the same. [NR]

(a) misfit functional before processing



(b) misfit functional after processing



**Figure 16.** Processing acts to simplify this misfit surface  $\chi(\mathbf{p})$  by making specific pieces of information, for example, the traveltimes of well-defined seismic phases, more easily accessible. This pragmatic simplification mostly leads to the enlargement of the space of acceptable models, that is, of the effective null space, from the original  $\mathbb{P}_{\text{fit}}$  to  $\widetilde{\mathbb{P}}_{\text{fit}}$ . [NR]

$n\tilde{I}_{ik} = nT_{ik}nI_{ik}$  by construction, we also have  $nI_{ik}(\mathbf{p} \in \mathbb{P}_{\text{fit}}) \approx nI_{ik}$ . This means that approximately matching processed correlations implies approximately matching raw correlations.

For  $\mathbf{p} \notin \mathbb{P}_{\text{fit}}$ , the transfer coefficients  $nT_{ik}$  still define raw synthetics based on processed synthetics via  $nI_{ik}(\mathbf{p} \notin \mathbb{P}_{\text{fit}}) = nT_{ik}^{-1}n\tilde{I}_{ik}(\mathbf{p} \notin \mathbb{P}_{\text{fit}})$ . However,  $nI_{ik}(\mathbf{p} \notin \mathbb{P}_{\text{fit}})$  is not generally the correlation that reproduces  $n\tilde{I}_{ik}(\mathbf{p} \notin \mathbb{P}_{\text{fit}})$  through the application of the transfer function, that is,  $T[nI_{ik}(\mathbf{p} \notin \mathbb{P}_{\text{fit}})] \neq n\tilde{I}_{ik}(\mathbf{p} \notin \mathbb{P}_{\text{fit}})$ . This is because the transfer coefficients  $nT_{ik}$  are guaranteed to represent the transfer function  $T$  only near the actual data points  $nI_{ik} \approx nI_{ik}(\mathbf{p} \in \mathbb{P}_{\text{fit}})$ .

The fact that the transfer function  $T$  and the transfer coefficients  $nT_{ik}$  may produce different results for  $\mathbf{p} \notin \mathbb{P}_{\text{fit}}$  has no detrimental effect on the solution of an inverse problem *per se* because non-data-fitting models are mere intermediate steps towards models that explain observations acceptably well. However, by being linear, the action of the transfer coefficients may be expected to be less numerically challenging than the potentially highly non-linear transfer function.

The mapping of different  $^oI$  into the same  $^o\tilde{I}$  via the transfer function illustrated in Fig. 15(a), is a special facet of non-uniqueness induced by processing. As any operation performed on raw data, processing in the context of interferometry can at most preserve information. It follows that the space of nearly data-fitting models  $\mathbb{P}_{\text{fit}}$ , that is, the effective null space, can at best preserve its size under the action of processing. More commonly, processing will lead to the loss of information, and the effective null space will grow from  $\mathbb{P}_{\text{fit}}$  to  $\widetilde{\mathbb{P}}_{\text{fit}}$ . This loss, however, is compensated by making specific pieces of information more accessible, which often leads to a pragmatic simplification of the misfit surface. For instance, processing may enhance a specific seismic phase, which may facilitate the estimation of a propagation velocity that can be difficult to extract from raw interferograms. The relation between processing and effective null-space enlargement is shown schematically in Fig. 16.

The modification of the misfit surface changes the trajectory of iterative inversions (Valentine & Trampert 2016). Different choices concerning processing and measurements will thus lead to different models. Furthermore, models  $\mathbf{p} \in \widetilde{\mathbb{P}}_{\text{fit}}$  that explain processed observations acceptably well may not be contained in the smaller  $\mathbb{P}_{\text{fit}}$ , that is, they may not necessarily explain the raw observations.

### 6.3 The perfect processing

The quality of the effective forward modelling theories depends on the forward modelling error, which is controlled by the factorization residuals  ${}_n e_{ik}$ . The perfect effective theory has vanishing forward modelling errors caused by vanishing factorization residuals.

While the factorization residuals for a given processing scheme and data set are difficult to estimate *a priori*, we can design an optimal processing that is (i) optimally close to the original one, and (ii) has zero factorization residuals. All necessary elements for the design of the optimal processing have already been developed in the previous sections, and it proceeds as follows:

First, we execute, as usual, a given processing scheme. This yields preliminary processed interferograms  ${}^0 \tilde{I}_{ik}^{\text{prelim}}$ , and factorized transfer coefficients  ${}_n T_{ik}^{\text{prelim}} = {}_n f g_{ik} + {}_n e_{ik}$ . Second, we define the optimal transfer coefficient as  ${}_n T_{ik}^{\text{opt}} = {}_n f g_{ik}$ , that is, we omit the factorization error. Third, we define the optimal processing through the action of the optimal transfer function,  ${}^0 \tilde{I}_{ik} = {}_n T_{ik}^{\text{opt}} {}_n I_{ik}$ .

Since the factorization residuals  ${}_n e_{ik}$  are already designed to be optimally small, the optimal processing  ${}_n T_{ik}^{\text{opt}}$  is as close as possible to the original processing  ${}^0 \tilde{I}_{ik}^{\text{prelim}}$ . By construction, the optimal processing has no factorization residual, and the forward modelling error vanishes. The effectiveness of this optimal design is likely to be dependent on the specifics of the original processing scheme and the data set, and will be investigated in future studies.

### 6.4 The transfer coefficient for vanishing raw correlations

The transfer coefficient  ${}_n T_{ik} = {}^0 \tilde{I}_{ik} / {}^0 I_{ik}$  is formally not defined when the raw correlation vanishes for a specific frequency, that is, when  ${}^0 I_{ik} = 0$ . While  ${}^0 I_{ik}$  is never exactly zero in practice, it is still meaningful to set  ${}_n T_{ik} = 0$  when  $|{}^0 I_{ik}|$  drops below the instrumental noise level.

When only few time windows  $n$  are affected, the effect will average out. Otherwise, when  ${}^0 I_{ik} \approx 0$  for most time windows, the propagation corrector  $g_{ik}$  and the synthetic  ${}^0 \hat{I}_{ik}$  will be nearly zero for this frequency and receiver pair. Two scenarios must then be considered: (i) The raw correlation  ${}^0 I_{ik}$  is also close to zero. In this case, the observations can by design be explained by the effective synthetics. (ii) When the raw correlation  ${}^0 I_{ik}$  is significantly different from zero, spurious information in the form of a non-zero amplitude has been introduced by the processing. This information cannot and should not be explained by the effective synthetics that are by definition. Consequently, the forward modelling error will be large.

### 6.5 Homogenizing noise sources through processing

An important corollary of our work is the fact that no processing scheme is able to produce a homogeneous effective source from a heterogeneous original source. The spatial psd distributions  ${}_n S(\mathbf{x})$  appear under the space integral in eq. (8) and are therefore not accessible by any time-domain operation on the left-hand side. Processing can merely balance the sources  ${}_n S(\mathbf{x})$  for individual time windows via the processing-derived source correctors  ${}_n f$  that produce an effective individual source  ${}_n \hat{S}(\mathbf{x}) = {}_n f {}_n S(\mathbf{x})$  and an effective ensemble source  $\hat{S}(\mathbf{x}) = \sum_n {}_n f {}_n S(\mathbf{x})$ . Regions that are practically source-free will therefore remain source-free, irrespective of the processing applied to the raw data.

## 7 CONCLUSIONS AND OUTLOOK

In the previous sections we have developed a forward and inverse modelling framework for interferometry by correlation that goes beyond traditional Green function recovery. Our theory meets the following objectives: (i) Allow for heterogeneously distributed sources of continuous or transient nature, which includes noise sources and earthquakes. (ii) Properly accounts for any type of linear or nonlinear processing that may be applied to render the correlations more suitable for inversion. (iii) Enables the exploitation of seemingly unphysical waveforms that result from heterogeneous noise sources and that cannot be used in interferometry based on Green function retrieval. (iv) Operate for any type of medium, including a 3-D elastic, heterogeneous Earth with attenuation and anisotropy. (v) Unify the earthquake-based two-station method and ambient noise interferometry. (vi) Establish the theoretical foundation of interferometric full-waveform inversion for sources and Earth structure.

Our theory is based on the formulation of an effective forward model that links effective wavefield sources and effective wave propagation to synthetic interstation correlations. The extent to which the effective sources and effective propagation differ from the true physical sources and propagation is determined by the processing applied to the raw data.

The effective forward model may not be perfect. A forward modelling error, induced and controlled by the data processing, determines if the processed correlations can actually be interpreted as proper correlations, that is, as being the result of some effective sources and some effective wave propagation. If the forward modelling error is significantly larger than plausible differences between observations and synthetics, the correlations must either be excluded from the inversion or the processing scheme must be changed.

The application of adjoint techniques to the effective forward model allows us to derive finite-frequency Fréchet kernels for the sources of the wavefield and Earth structure. Using examples with a 2-D, homogeneous, unbounded medium, we demonstrate that processing must be taken into account properly in order to ensure meaningful inferences on sources and structure.

Since our developments do not make any restrictive assumptions on the nature of the wavefield sources, they can be naturally applied to both earthquake and noise data, either separately or combined. This allows us, for instance, to locate earthquakes using interstation correlations, to bridge the gap between the earthquake-based two-station method and noise correlations, and to eliminate the strict need to remove earthquake signals from noise recordings prior to the computation of correlation functions.

While our work is focused on interferometry by correlation, we acknowledge that alternative approaches have been developed in recent years; mostly in order to overcome the deficiencies of Green function retrieval by correlation. These include interferometry by deconvolution (e.g. Snieder & Šafak 2006; Vasconcelos & Snieder 2008a,b), multidimensional deconvolution (e.g. Wapenaar *et al.* 2008; Wapenaar & van der Neut 2010), and repeated correlation (e.g. Stehly *et al.* 2008). A generalization of these approaches, along the lines presented above, is possible, beyond the scope of this, but the subject of a follow-up publication.

## ACKNOWLEDGEMENTS

AF would like to thank Michel Campillo, Shravan Hanasoge, Paula Koelemeijer, Dimitri Komatitsch, Korbinian Sager, Jeannot Trampert, Victor Tsai, Arie Verdel and the Stanford class on ‘Seismic Imaging using Earthquake and Ambient Noise Data’ for many fruitful discussions and inspiration. Much of this manuscript was written during AF’s stay at Stanford University. The hospitality of Biondo Biondi, Simon Klemperer, Gregory Beroza, Jon Claerbout and many others is very kindly appreciated. All authors would like to thank the editor Jean Virieux for his work, as well as the reviewers Kees Wapenaar and Yann Capdeville who provided numerous constructive comments and suggestions. Furthermore, the COST project TIDES is gratefully acknowledged for providing an excellent platform for discussions on the topic presented here. This research was supported by ETH Zurich, by the Swiss National Supercomputing Center (CSCS) in the form of the GeoScale and CH1 projects, by the Swiss National Science Foundation (SNF) under grant 200021\_149143, by the Netherlands Organisation for Scientific Research (VIDI grant 864.11.008), and by Shell through the project ‘Boosting full-waveform inversion’. Software to reproduce figures labelled [ER] can be found on the web pages of the ETH Computational Seismology Group at <http://www.cos.ethz.ch/software.html>.

## REFERENCES

- Aki, K., 1957. Space and time spectra of stationary stochastic waves, with special reference to microtremors, *Bull. Earthq. Res. Inst., Univ. Tokyo*, **35**, 415–457.
- Aki, K. & Richards, P., 2002. *Quantitative Seismology*, University Science Books.
- Alkhalifah, T., 2000. An acoustic wave equation for anisotropic media, *Geophysics*, **65**, 1239–1250.
- Ardhuin, F., Stutzmann, E., Schimmel, M. & Mangeney, A., 2011. Ocean wave sources of seismic noise, *J. geophys. Res.*, **116**, doi:10.1029/2011JC006952.
- Ardhuin, F., Gualtieri, L. & Stutzmann, E., 2015. How ocean waves rock the earth: two mechanisms explain microseisms with periods 3 to 300 s, *Geophys. Res. Lett.*, **42**(3), 765–772.
- Backus, G.E., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.*, **67**, 4427–4440.
- Baig, A.M., Campillo, M. & Brenguier, F., 2009. Denoising seismic noise cross correlations, *J. geophys. Res.*, **114**, doi:10.1029/2008JB006085.
- Barnes, C. & Charara, M., 2008. Full-waveform inversion results when using acoustic approximations instead of elastic medium, *SEG Expanded Abstr.*, **78**, 1895–1899.
- Basini, P., Nissen-Meyer, T., Boschi, L., Casarotti, E., Verbeke, J., Schenk, O. & Giardini, D., 2013. The influence of nonuniform ambient noise on crustal tomography in Europe, *Geochem. Geophys. Geosys.*, **14**, 1471–1492.
- Bath, M., 1968. *Mathematical Aspects of Seismology*, Elsevier Publishing Company.
- Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**, 1239–1260.
- Bodin, T., Capdeville, Y., Romanowicz, B. & Montagner, J.-P., 2015. Interpreting radial anisotropy in global and regional tomographic models, in *The Earth’s Heterogeneous Mantle*, pp. 105–144, eds Khan, A. & Deschamps, F., Springer.
- Boehm, C., Hanzich, M., de la Puente, J. & Fichtner, A., 2016. Wavefield compression or adjoint methods in full-waveform inversion, *Geophysics*, **81**, R385–R397.
- Bogey, C. & Bailly, C., 2004. A family of low dispersive and low dissipative explicit schemes for flow and noise computations, *J. Comp. Phys.*, **194**, 194–214.
- Boschi, L., Weemstra, C., Verbeke, J., Ekström, G., Zunino, A. & Giardini, D., 2013. On measuring surface wave phase velocity from station-station cross-correlation of ambient signal, *Geophys. J. Int.*, **192**, 346–358.
- Boué, P., Poli, P., Campillo, M., Pedersen, H., Briand, X. & Roux, P., 2013. Teleseismic correlations of ambient noise for deep global imaging of the Earth, *Geophys. J. Int.*, **194**, 844–848.
- Bucher, R. & Misra, D., 2002. A synthesizable VHDL model for the exact solution for three-dimensional hyperbolic positioning system, *VLSI Des.*, **15**, 507–520.
- Bussat, S. & Kugler, S., 2011. Offshore ambient-noise surface-wave tomography above 0.1 Hz and its applications, *Leading Edge*, **30**, 514–524.
- Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse coda, *Science*, **299**, 547–549.
- Cance, P. & Capdeville, Y., 2015. Validity of the acoustic approximation for elastic waves in heterogeneous media, *Geophysics*, **80**, T161–T173.
- Capdeville, Y., Guillot, L. & Marigo, J.J., 2010. 2-D nonperiodic homogenization to upscale elastic media for P-SV waves, *Geophys. J. Int.*, **182**, 903–922.
- Capdeville, Y., Stutzmann, E., Montagner, J.-P. & Wang, N., 2013. Residual homogenization for seismic forward and inverse problems in layered media, *Geophys. J. Int.*, **194**, 470–487.
- Capdeville, Y., Zhao, M. & Cupillard, P., 2015. Fast Fourier homogenization for elastic wave propagation in complex media, *Wave Motion*, **50**, 170–186.
- Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances, *Pure appl. Geophys.*, **168**, 1527–1552.
- Chen, P., Zhao, L. & Jordan, T.H., 2007. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.*, **97**, 1094–1120.
- Chevrot, S. & Zhao, L., 2007. Multicale finite-frequency Rayleigh wave tomography of the Kaapvaal craton, *Geophys. J. Int.*, **169**, 201–215.
- Claerbout, J.F., 1968. Synthesis of a layered medium from its acoustic transmission response, *Geophysics*, **33**, 264–269.
- Cole, S.P., 1995. Passive seismic and drill-bit experiments using 2-D arrays, *PhD thesis*, The Stanford Exploration Project, Stanford University.

- Cupillard, P., 2008. Simulation par la méthode des éléments spectraux des formes d'onde obtenues par corrélation de bruit sismique, *PhD thesis*, Institut de Physique du Globe de Paris.
- Cupillard, P. & Capdeville, Y., 2010. On the amplitude of surface waves obtained by noise correlation and the capability to recover the attenuation: a numerical approach, *Geophys. J. Int.*, **181**, 1687–1700.
- Cupillard, P., Stehly, L. & Romanowicz, B., 2011. The one-bit noise correlation: a theory based on the concepts of coherent and incoherent noise, *Geophys. J. Int.*, **184**, 1397–1414.
- Curtis, A. & Halliday, D., 2010. Directional balancing for seismic and general wavefield interferometry, *Geophysics*, **75**, doi:10.1190/1.3298736.
- Dahlen, F., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes – I. Theory, *Geophys. J. Int.*, **141**, 157–174.
- Dahlen, F.A. & Baig, F.A., 2002. Fréchet kernels for body wave amplitudes, *Geophys. J. Int.*, **150**, 440–466.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton Univ. Press.
- Daskalakis, E., Evangelidis, C.P., Garnier, J., Melis, N.S., Papanicolaou, G. & Tsogka, C., 2016. Robust seismic velocity change estimation using ambient noise recordings, *Geophys. J. Int.*, **205**, 1926–1936.
- Delaney, E., Ermert, L., Sager, K., Krtski, A., Bussat, S. & Fichtner, A., 2016. Removal of traveltime errors in time-lapse passive seismic monitoring induced by non-stationary noise sources, *Geophysics*, in press.
- de la Puente, J., Dumbser, M., Käser, M. & Igel, H., 2007. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured methods – IV. Anisotropy, *Geophys. J. Int.*, **169**, 1210–1228.
- de Ridder, S.A.L. & Biondi, B.L., 2015. Ambient seismic noise tomography at Ekofisk, *Geophysics*, **80**, B167–B176.
- de Ridder, S.A.L., Biondi, B.L. & Clapp, R.G., 2014. Time-lapse seismic noise correlation tomography at Valhall, *Geophys. Res. Lett.*, **41**, 6116–6122.
- de Vos, D., Paulssen, H. & Fichtner, A., 2013. Finite-frequency sensitivity kernels from two-station surface wave measurements, *Geophys. J. Int.*, **194**, 1042–1049.
- Dumbser, M., Käser, M. & de la Puente, J., 2007. Arbitrary high order finite volume schemes for seismic wave propagation on unstructured meshes in 2D and 3D, *Geophys. J. Int.*, **171**, 665–694.
- Durand, S., Montagner, J.-P., Roux, P., Brenguier, F., Nadeau, R.M. & Ricard, Y., 2011. Passive monitoring of anisotropy change associated with the Parkfield 2004 earthquake, *Geophys. Res. Lett.*, **38**, doi:10.1029/2011GL047875.
- Duvall, T.L., Birch, A.C. & Gizon, L., 2006. Direct measurement of travel-time kernels for helioseismology, *Astrophys. J.*, **646**, 553–559.
- Endrun, B., Meier, T., Lebedev, S., Bonhoff, M., Stavrakakis, G. & Harjes, H.-P., 2008. S velocity structure and radial anisotropy in the Aegean region from surface wave dispersion, *Geophys. J. Int.*, **174**, 593–616.
- Ermert, L., Villasenor, A. & Fichtner, A., 2016. Cross-correlation imaging of ambient noise sources, *Geophys. J. Int.*, **204**, 347–364.
- Fang, B.T., 1990. Simple solutions for hyperbolic and related position fixes, *IEEE Trans. Aerosp. Electr. Sys.*, **26**, 748–753.
- Farra, V., Stutzmann, E., Gualtieri, L. & Arduin, F., 2016. Ray-theoretical modeling of secondary microseism p waves, *Geophys. J. Int.*, **206**, 1730–1739.
- Fichtner, A., 2010. *Full Seismic Waveform Modelling and Inversion*, Springer.
- Fichtner, A., 2014. Source and processing effects on noise correlations, *Geophys. J. Int.*, **197**, 1527–1531.
- Fichtner, A., 2015. Source-structure trade-offs in ambient noise correlations, *Geophys. J. Int.*, **202**, 678–694.
- Fichtner, A. & Igel, H., 2008. Efficient numerical surface wave propagation through the optimization of discrete crustal models – a technique based on non-linear dispersion curve matching (DCM), *Geophys. J. Int.*, **173**, 519–533.
- Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology – I. Theory, *Phys. Earth planet. Inter.*, **157**, 86–104.
- Fichtner, A., Kennett, B. L.N., Igel, H. & Bunge, H.-P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.*, **175**, 665–685.
- 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**, 1703–1725.
- Fichtner, A., Kennett, B.L.N. & Trampert, J., 2013a. Separating intrinsic and apparent anisotropy, *Phys. Earth planet. Int.*, **219**, 11–20.
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villasenor, A., 2013b. Multi-scale full waveform inversion, *Geophys. J. Int.*, **194**, 534–556.
- Fletcher, R., Du, X. & Fowler, P.J., 2008. A new pseudo-acoustic wave equation for TI media, *SEG Expanded Abstr.*, **78**, 2082–2086.
- Friederich, W., 2003. The S-velocity structure of the East Asian mantle from inversion of shear and surface waveforms, *Geophys. J. Int.*, **153**, 88–102.
- Froment, B., Campillo, M., Roux, P., Gouédard, P., Verdel, A. & Weaver, R.L., 2010. Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations, *Geophysics*, **75**, SA85–SA93.
- Gee, L.S. & Jordan, T.H., 1992. Generalized seismological data functionals, *Geophys. J. Int.*, **111**, 363–390.
- Gimbert, F. & Tsai, V.C., 2016. Predicting short-period wind-wave generated seismic noise in coastal regions, *Earth Planet. Sci. Lett.*, **426**, 280–292.
- Gizon, L. & Birch, A.C., 2002. Time-distance helioseismology: the forward problem for random distributed sources, *Astrophys. J.*, **571**, 966–986.
- Godin, O.A., 2009. Emergence of deterministic Green's functions from noise generated by finite random sources, *Phys. Rev. E*, **80**, doi:10.1103/PhysRevE.80.066605.
- Gokhberg, A. & Fichtner, A., 2016. Full-waveform inversion on heterogeneous HPC systems, *Comput. Geosci.*, **89**, 260–268.
- Groos, J.C., Bussat, S. & Ritter, J. R.R., 2012. Performance of different processing schemes in seismic noise cross-correlations, *Geophys. J. Int.*, **188**, 498–512.
- Gualtieri, L., Stutzmann, E., Capdeville, Y., Arduin, F.M., Schimmel, A.M. & Morelli, A., 2013. Modeling secondary microseismic noise by normal mode summation, *Geophys. J. Int.*, **193**, 1732–1745.
- Hadzioannou, C., Gaebler, P., Wassermann, J. & Igel, H., 2012. Examining ambient noise using co-located measurements of rotational and translational motion, *J. Seismol.*, **16**, 787–796.
- Halliday, D. & Curtis, A., 2008. Seismic interferometry, surface waves and source distribution, *Geophys. J. Int.*, **175**, 1067–1087.
- Halliday, D. & Curtis, A., 2009. Seismic interferometry of scattered surface waves in attenuative media, *Geophys. J. Int.*, **178**, 419–446.
- Hanasoge, S.M., 2013. Measurements and kernels for source-structure inversions in noise tomography, *Geophys. J. Int.*, **192**, 971–985.
- Hanasoge, S.M. & Branicki, M., 2013. Interpreting cross-correlations of one-bit filtered noise, *Geophys. J. Int.*, **195**, 1811–1830.
- Hanasoge, S.M., Birch, A., Gizon, L. & Tromp, J., 2011. The adjoint method applied to time-distance helioseismology, *Astrophys. J.*, **738**, doi:10.1088/0004–637X/738/1/100.
- Haned, A., Stutzmann, E., Schimmel, M., Kiselev, S., Davaille, A. & Yelles-Chaouche, A., 2016. Global tomography using seismic hum, *Geophys. J. Int.*, **204**, 1222–1236.
- Harmon, N., Rychert, C. & Gerstoft, P., 2010. Distribution of noise sources for seismic interferometry, *Geophys. J. Int.*, **183**, 1470–1484.
- Hasselmann, K., 1963. A statistical analysis of the generation of microseisms, *Rev. Geophys.*, **1**(2), 177–210.
- Hillers, G., Campillo, M., Lin, Y.-Y., Ma, K.-F. & Roux, R., 2012. Anatomy of the high-frequency ambient seismic wave field at the TCDP borehole, *J. geophys. Res.*, **117**, doi:10.1029/2011JB008999.
- Huang, H.-H., Lin, F.-C., Tsai, V.C. & Koper, K.D., 2015. High-resolution probing of inner core structure with seismic interferometry, *Geophys. Res. Lett.*, **42**(24), 10 622–10 630.
- Igel, H., 2016. *Computational Seismology: A Practical Introduction*, Oxford Univ. Press.
- Igel, H., Djikpesse, H. & Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, *Geophys. J. Int.*, **124**, 363–371.
- Kennett, B.L.N., 2001. *The Seismic Wavefield I. Introduction and Theoretical Development*, Cambridge Univ. Press.
- Kimman, W. & Trampert, J., 2010. Approximations in seismic interferometry and their effects on surface waves, *Geophys. J. Int.*, **182**, 461–476.

- Kristekova, M., Kristek, J., Moczo, P. & Day, S.M., 2006. Misfit criteria for quantitative comparison of seismograms, *Bull. seis. Soc. Am.*, **96**, 1836–1850.
- Kuo, B.-Y., Chi, W.-C., Lin, C.-R., Chang, E.T.-Y., Collins, J. & Liu, C.-S., 2009. Two-station measurement of Rayleigh-wave phase velocities or the Huatung basin, the westernmost Philippine Sea, with OBS: implications for regional tectonics, *Geophys. J. Int.*, **179**, 1859–1869.
- Kurrale, D. & Widmer-Schnidrig, R., 2006. Spatiotemporal features of the Earth's background oscillations observed in central Europe, *Geophys. Res. Lett.*, **33**(24), L24304, doi:10.1029/2006GL028429.
- Larose, E., Derode, A., Campillo, M. & Fink, M., 2004. Imaging from one-bit correlations of wideband diffuse wave fields, *J. Appl. Phys.*, **95**, 8393–8399.
- Lin, F.-C., Moschetti, M.P. & Ritzwoller, M.H., 2008. Surface wave tomography of the western United States from ambient noise: Rayleigh and Love wave phase velocity maps, *Geophys. J. Int.*, **173**, 281–298.
- Lin, F.-C., Tsai, V.C., Schmandt, B., Duputel, Z. & Zhan, Z., 2013. Extracting seismic core phases with array interferometry, *Geophys. Res. Lett.*, **40**, 1049–1053.
- Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.*, **96**, 2383–2397.
- Liu, X. & Ben-Zion, Y., 2013. Estimating correlations of neighbouring frequencies in ambient seismic noise, *Geophys. J. Int.*, **206**, 1065–1075.
- Lobkis, O.I. & Weaver, R.L., 2001. On the emergence of the Green's function in the correlations of a diffuse field, *J. acoust. Soc. Am.*, **110**, 3011–3017.
- Longuet-Higgins, M.S., 1950. A theory of the origin of microseisms, *Phil. Trans. R. Soc. Lond. A*, **243**(857), 1–35.
- Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltimes inversion, *Geophysics*, **56**, 645–653.
- Malcolm, A.E., Scales, J. & van Tiggelen, B.A., 2004. Extracting the Green function from diffuse, equipartitioned waves, *Phys. Rev. E*, **70**, doi:10.1103/PhysRevE.70.015601.
- Marelli, S., Maurer, H. & Manukyan, E., 2012. Validity of the acoustic approximation in full-waveform seismic crosshole tomography, *Geophysics*, **77**, R129–R139.
- Meier, T., Dietrich, K., Stockhert, B. & Harjes, H.-P., 2004. One-dimensional models of shear wave velocity for the eastern Mediterranean obtained from the inversion of Rayleigh wave phase velocities and tectonic implications, *Geophys. J. Int.*, **156**, 45–58.
- Moczo, P., Kristek, J., Vavrycuk, V., Archuleta, R. & Halada, L., 2002. 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli, *Bull. seis. Soc. Am.*, **92**, 3042–3066.
- Mordret, A., Landès, M., Shapiro, N., Singh, S., Roux, P. & Barkved, O., 2013. Near-surface study at the Valhall oil field from ambient noise surface wave tomography, *Geophys. J. Int.*, **193**(3), 1627–1643.
- Mordret, A., Landès, M., Shapiro, N., Singh, S. & Roux, P., 2014. Ambient noise surface wave tomography to determine the shallow shear velocity structure at Valhall: depth inversion with a neighbourhood algorithm, *Geophys. J. Int.*, **198**(3), 1514–1525.
- Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.*, **100**, 12 431–12 447.
- Nakata, N., Chang, J.P., Lawrence, J.F. & Boué, P., 2015. Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *J. geophys. Res.*, **120**, 1159–1173.
- Nawa, K., Suda, N., Fukao, Y., Sato, T., Aoyama, Y. & Shibuya, K., 1998. Incessant excitation of the Earth's free oscillations, *Earth Planets Space*, **50**, 3–8.
- Nishida, K., 2013. Global propagation of body waves revealed by cross-correlation analysis of seismic hum, *Geophys. Res. Lett.*, **40**, 1691–1696.
- Nishida, K., 2014. Source spectra of seismic hum, *Geophys. J. Int.*, **199**(1), 416–429.
- Nishida, K. & Montagner, J.-P., 2009. Global surface wave tomography using seismic hum, *Science*, **326**, 5949.
- Nocedal, J. & Wright, S.J., 1999. *Numerical Optimization*, Springer.
- Operto, S., Virieux, J., Ribodetti, A. & Anderson, J.E., 2009. Finite-difference frequency-domain modeling of viscoacoustic wave propagation in 2D tilted transversely isotropic (TTI) media, *Geophysics*, **74**, T75–T95.
- Parker, R.L., 1977. Understanding inverse theory, *Annu. Rev. Earth Planet. Sci.*, **5**, 35–64.
- Passier, M.L. & Snieder, R.K., 1995. Using differential waveform data to retrieve local S velocity structure or path-averaged S velocity gradients, *J. geophys. Res.*, **100**, 24 061–24 078.
- Pedersen, H.A., 2006. Impacts of non-plane waves on two-station measurements of phase velocities, *Geophys. J. Int.*, **65**, 279–287.
- Pedersen, H.A., Boué, P., Poli, P. & Colombi, A., 2015. Arrival angle anomalies of Rayleigh waves observed at a broadband array: a systematic study based on earthquake data, full waveform simulations and noise correlations, *Geophys. J. Int.*, **203**, 1626–1641.
- Peter, D. et al., 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.*, **186**, 721–739.
- Peterson, J., 1993. Observations and modeling of seismic background noise, *USGS Open File Report*, 93–322, 94 pp.
- Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, **167**, 495–503.
- Poli, P., Campillo, M., Pedersen, H. & LAPNET Working Group, 2012. Body-wave imaging of Earth's mantle discontinuities from ambient seismic noise, *Science*, **38**, 1063–1065.
- Poli, P., Thomas, C., Campillo, M. & Pedersen, H., 2015. Imaging the D'' reflector with noise correlations, *Geophys. Res. Lett.*, **42**, 60–65.
- Reading, A.M., Koper, K.D., Gal, M., Graham, L.S., Tkalcic, H. & Hemer, M.A., 2014. Dominant seismic noise sources in the southern ocean and west pacific, 2000–2012, recorded at the Warramunga seismic array, Australia, *Geophys. Res. Lett.*, **41**(10), 3455–3463.
- Rhie, J. & Romanowicz, B., 2006. A study of the relation between ocean storms and the Earth's hum, *Geochem. Geophys. Geosyst.*, **7**(10), doi:10.1029/2006GC001274.
- Riahi, N., Bokelmann, G., Sala, P. & Saenger, E.H., 2013. Time-lapse analysis of ambient surface wave anisotropy: a three-component array study above an underground gas storage, *J. geophys. Res.*, **118**, 5339–5351.
- Roux, P., 2009. Passive seismic imaging with directive ambient noise: application to surface waves and the San Andreas Fault in Parkfield, CA, *Geophys. J. Int.*, **179**, 367–373.
- Rudin, W., 1966. *Real and Complex Analysis*, McGraw-Hill.
- Sabra, K.G., Gerstoft, P., Roux, P. & Kuperman, W.A., 2005. Surface wave tomography from microseisms in Southern California, *Geophys. Res. Lett.*, **32**, doi:10.1029/2005GL023155.
- Sambridge, M.S. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, **40**, doi:10.1029/2000RG000089.
- Sambridge, M.S., Bodin, T., Gallagher, K. & Tkalcic, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A*, **371**, doi:10.1098/rsta.2011.0547.
- Saygin, E. & Kennett, B. L.N., 2012. Crustal structure of Australia from ambient seismic noise tomography, *J. geophys. Res.*, **117**, doi:10.1029/2011JB008403.
- Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.*, **130**, 497–505.
- Schimmel, M., Stutzmann, E. & Gallart, J., 2011. Using instantaneous phase coherence for signal extraction from ambient noise data at a local to a global scale, *Geophys. J. Int.*, **184**, 494–506.
- Schulte-Pelkum, V., Earle, P.S. & Vernon, F.L., 2004. Strong directivity of ocean-generated seismic noise, *Geochem. Geophys. Geosyst.*, **5**(3), doi:10.1029/2003GC000520.
- Schuster, G.T., Hu, J., Sheng, J. & Rickett, J., 2004. Interferometric/daylight seismic imaging, *Geophys. J. Int.*, **157**, 838–852.
- Seats, K.J., Lawrence, J.F. & Prieto, G.A., 2012. Improved ambient noise correlation functions using Welch's method, *Geophys. J. Int.*, **188**, 513–523.
- Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**, doi:10.1029/2004GL019491.

- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M., 2005. High resolution surface wave tomography from ambient seismic noise, *Science*, **307**, 1615–1618.
- Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007. Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.*, **168**, 1153–1174.
- Sigloch, K., McQuarrie, N. & Nolet, G., 2008. Two-stage subduction history under north america inferred from multiple-frequency tomography, *Nat. Geosc.*, **1**, doi:10.1038/ngeo231.
- Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E*, **69**, doi:10.1103/PhysRevE.69.046610.
- Snieder, R. & Šafak, E., 2006. Extracting the building response using seismic interferometry: Theory and application to the Millikan Library in Pasadena, California, *Bull. seism. Soc. Am.*, **96**, 586–598.
- Stehly, L. & Cupillard, P., 2016. Modulating the distribution of noise sources that contribute to seismic-noise correlations, *Geophys. J. Int.*, in press.
- Stehly, L., Campillo, M. & Shapiro, N.M., 2007. Travelttime measurements from noise correlation: stability and detection of instrumental time-shifts, *Geophys. J. Int.*, **171**, 223–230.
- Stehly, L., Campillo, M., Froment, B. & Weaver, R.L., 2008. Reconstructing Green's function by correlation of the coda of the correlation (C3) of ambient seismic noise, *J. geophys. Res.*, **113**, doi:10.1029/2008JB005693.
- Stehly, L., Fry, B., Campillo, M., Shapiro, N.M., Guilbert, J., Boschi, L. & Giardini, D., 2009. Tomography of the Alpine region from observations of seismic ambient noise, *Geophys. J. Int.*, **178**, 338–350.
- Stehly, L., Cupillard, P. & Romanowicz, B., 2011. Towards improving ambient noise tomography using simultaneously curvelet denoising filters and SEM simulations of seismic ambient noise, *C. R. Geosc.*, **343**, 591–599.
- Stutzmann, E., Arduin, F., Schimmel, M., Mangeney, A. & Patau, G., 2012. Modelling long-term seismic noise in various environments, *Geophys. J. Int.*, **191**, 707–722.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based upon spectral-element and adjoint methods, *Geophys. J. Int.*, **180**, 433–462.
- Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure appl. Geophys.*, **128**, 365–399.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, 2nd edn, Society for Industrial and Applied Mathematics.
- Traer, J. & Gerstoft, P., 2014. A unified theory of microseisms and hum, *J. geophys. Res.*, **119**(4), 3317–3339.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Tromp, J., Luo, Y., Hanasoge, S. & Peter, D., 2010. Noise cross-correlation sensitivity kernels, *Geophys. J. Int.*, **183**, 791–819.
- Tsai, V.C., 2009. On establishing the accuracy of noise tomography travelttime measurements in a realistic medium, *Geophys. J. Int.*, **178**, 1555–1564.
- Tsai, V.C., 2010. The relationship between noise correlation and the Green's function in the presence of degeneracy and the absence of equipartition, *Geophys. J. Int.*, **182**, 1509–1514.
- Tsai, V.C., 2011. Understanding the amplitudes of noise correlation measurements, *J. geophys. Res.*, **116**, doi:10.1029/2011JB008483.
- Valentine, A.P. & Trampert, J., 2016. The impact of approximations and arbitrary choices on geophysical images, *Geophys. J. Int.*, **204**, 59–73.
- Valentine, A.P. & Woodhouse, J.H., 2010. Reducing errors in seismic tomography: combined inversion for sources and structure, *Geophys. J. Int.*, **180**, 847–857.
- van Driel, M., Krischer, L., Stähler, S.C., Hosseini, K. & Nissen-Meyer, T., 2015. Instaseis: instant global seismograms based on a broadband waveform database, *Solid Earth*, **6**, 701–717.
- van Leeuwen, T. & Mulder, W.A., 2010. A correlation-based misfit criterion for wave-equation travelttime tomography, *Geophys. J. Int.*, **182**, 1383–1394.
- van Vleck, J.H. & Middleton, D., 1966. The spectrum of clipped noise, *Proc. IEEE*, **54**, 2–19.
- Vasconcelos, I. & Snieder, R., 2008a. Interferometry by deconvolution, Part 1 – Theory for acoustic waves and numerical examples, *Geophysics*, **73**, S129–S141.
- Vasconcelos, I. & Snieder, R., 2008b. Interferometry by deconvolution, Part 2 – Theory for elastic waves and application to drill-bit seismic imaging, *Geophysics*, **73**, S115–S128.
- Verbeke, J., Boschi, L., Stehly, L., Kissling, E. & Michelini, A., 2012. High-resolution Rayleigh-wave velocity maps of central Europe from a dense ambient-noise data set, *Geophys. J. Int.*, **188**, 1173–1187.
- Wang, M. & Xu, S., 2015. Finite-difference time dispersion transforms for wave propagation, *Geophysics*, **80**, WD19–WD25.
- Wapenaar, K., 2004. Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation, *Phys. Rev. Lett.*, **93**, doi:10.1103/PhysRevLett.93.254301.
- Wapenaar, K. & Fokkema, J., 2006. Green's function representations for seismic interferometry, *Geophysics*, **71**, SI33–SI46.
- Wapenaar, K. & van der Neut, J., 2010. A representation for Greens function retrieval by multidimensional deconvolution, *J. acoust. Soc. Am.*, **128**, 366–371.
- Wapenaar, K., van der Neut, J. & Ruigrok, E., 2008. Passive seismic interferometry by multidimensional deconvolution, *Geophysics*, **73**, A51–A56.
- Welch, P.D., 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time-averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoust.*, **15**, 70–73.
- Woodard, M.F., 1997. Implications of localized, acoustic absorption for heliotomographic analysis of sunspots, *Astrophys. J.*, **485**, 890–894.
- Yanovskaya, T.B., Koroleva, T. & Lyskova, E., 2016. Effect of earthquakes on ambient noise surface wave tomography in upper-mantle studies, *Geophys. J. Int.*, **205**, 1208–1220.
- Yao, H. & van der Hilst, R.D., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet, *Geophys. J. Int.*, **179**, 1113–1132.
- Yomogida, K., 1992. Fresnel zone inversion for lateral heterogeneities in the Earth, *Pure appl. Geophys.*, **138**, 391–406.
- Yoshizawa, K. & Kennett, B. L.N., 2004. Multi-mode surface wave tomography for the Australian region using a 3-stage approach incorporating finite-frequency effects, *J. geophys. Res.*, **109**, doi:10.1029/2002JB002254.
- Zhan, Z., Tsai, V.C. & Clayton, R.W., 2013. Spurious velocity changes caused by temporal variations in ambient noise frequency content, *Geophys. J. Int.*, **194**, 1574–1581.
- Zhang, X., Paulsen, H., Lebedev, S. & Meier, 2007. Surface wave tomography of the Gulf of California, *Geophys. Res. Lett.*, **34**, doi:10.1029/2007GL030631.
- Zheng, Y., Shen, W., Zhou, L., Yang, Y., Xie, Z. & Ritzwoller, M., 2011. Crust and uppermost mantle beneath the North China Craton, north-eastern China, and the Sea of Japan from ambient noise tomography, *J. geophys. Res.*, **116**, doi:10.1029/2011JB008637.
- Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.*, **158**, 142–168.

## SUPPORTING INFORMATION

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

### GI1

#### GI\_documentation.pdf figures\_input\_files

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: OPTIMAL FACTORIZATION OF THE TRANSFER COEFFICIENT

In the following paragraphs, we derive a scheme for the computation of the transfer coefficient factors such that the factorization error is minimal. We start with the simpler scalar acoustic case and then use similar concepts for the vectorial elastic scenario.

### A1 The scalar (acoustic) case

We consider the factorization of the transfer coefficient  ${}_n T_{ik}$ , first introduced in eq. (13),

$${}_n T_{ik} = {}_n f g_{ik} + {}_n e_{ik}, \quad (\text{A1})$$

where we try to find optimal source correction factors  ${}_n f$  and optimal propagation correction factors  $g_{ik}$  such that the error terms  ${}_n e_{ik}$  are as small as possible. As demonstrated in Section 2.5.2, the forward modelling error of the effective theory is minimal when the empirical expectation of the factorization residuals is equal to zero:

$$e_{ik} = \frac{1}{N} \sum_{n=1}^N {}_n e_{ik} = 0, \quad \text{for all fixed } i, k, \quad [\text{First optimality condition}]. \quad (\text{A2})$$

To ensure that not only the average but also the individual factorization errors are small, we furthermore request that the receiver-pair-averaged errors be minimal:

$$\sum_{i,k} |{}_n e_{ik}|^2 = \text{minimum}, \quad \text{for all fixed } n. \quad [\text{Second optimality condition}]. \quad (\text{A3})$$

With the optimality conditions (A2) and (A3), the factorization  ${}_n f g_{ik}$  is still under-determined up to a complex-valued factor. We may therefore put an additional constraint, and we choose

$$\frac{1}{N} \sum_{n=1}^N {}_n f = 1, \quad [\text{Subsidiary condition}]. \quad (\text{A4})$$

Condition (A4) ensures that the average source correction is equal to 1, which means that average amplitude changes of the interferometric wavefield cannot result from a mere scaling of the source. To determine the propagation corrections  $g_{ik}$ , we sum eq. (A1) over  $n$ . This yields

$$T_{ik} = g_{ik} + e_{ik}, \quad (\text{A5})$$

where we defined the time-averaged transfer coefficients

$$T_{ik} = \frac{1}{N} \sum_{n=1}^N {}_n T_{ik}. \quad (\text{A6})$$

From the first optimality condition (A2) combined with eq. (A5), we deduce that the optimal propagation correction factors are

$$g_{ik} = T_{ik}. \quad (\text{A7})$$

With the propagation correction factors determined, we consider the second optimality condition in order to find the optimal source correction factors. For this, we write eq. (A3) in the form of a misfit functional  ${}_n L$  that we wish to minimize:

$${}_n L = \sum_{i,k} |{}_n e_{ik}|^2 = \sum_{i,k} [{}_n T_{ik} - {}_n f {}_n g_{ik}] [{}_n T_{ik} - {}_n f {}_n g_{ik}]^*. \quad (\text{A8})$$

Requiring the derivative of  ${}_n L$  with respect to  ${}_n f$  to vanish, yields the normal equation

$$0 = \text{Re} {}_n f \sum_{i,k} |g_{ik}|^2 - \text{Re} \sum_{i,k} {}_n T_{ik} g_{ik}^*. \quad (\text{A9})$$

It follows that the real part of the optimal source correction factor  ${}_n f$  is given by

$$\operatorname{Re} {}_n f = \frac{\operatorname{Re} \sum_{i,k} {}_n T_{ik} g_{ik}^*}{\sum_{i,k} |g_{ik}|^2}. \quad (\text{A10})$$

Obviously, our optimality requirements only determine the real part of  ${}_n f$ . The imaginary parts of  ${}_n f$  are so far undetermined up to the requirement that they sum to zero, which comes from the subsidiary condition (A4). However, an additional constraint comes from the requirement that an effective psd for a specific time window,  ${}_n f {}_n S(\mathbf{x}, \mathbf{x}')$  be physically meaningful. For  $\mathbf{x} = \mathbf{x}'$  the phase of  ${}_n S(\mathbf{x}, \mathbf{x})$  vanishes. To preserve this property in the effective psd, we set the imaginary parts of the source correction factors  ${}_n f$  to 0.

## A2 The vectorial (elastic) case

To derive the optimal factorization for the elastic case, we translate the approach used in the previous paragraphs. In the interest of a condensed notation, we introduce the scalar product between two matrices  $\mathbf{A}$  and  $\mathbf{B}$  as  $\mathbf{A} : \mathbf{B} = \sum_{i,k=1}^3 A_{ik} B_{ik}$ . In analogy to eqs (A2) and (A4), we request that the time-averaged factorization residuals vanish component-wise for each receiver pair,

$$\frac{1}{N} \sum_{n=1}^N {}_n \mathbf{e}_{ik} = \mathbf{0}, \quad (\text{A11})$$

and that the time-averaged source correctors are equal to 1,

$$\frac{1}{N} \sum_{n=1}^N {}_n f = 1. \quad (\text{A12})$$

Averaging the factorization  ${}_n \mathbf{T}_{ik} = \mathbf{g}_{ik} {}_n f + {}_n \mathbf{e}_{ik}$  (eq. 66) over all time intervals then yields the tensorial propagation corrector

$$\mathbf{T}_{ik} = \frac{1}{N} \sum_{n=1}^N {}_n \mathbf{T}_{ik} = \mathbf{g}_{ik}. \quad (\text{A13})$$

To minimize the factorization residuals  ${}_n \mathbf{e}_{ik}$ , we consider the misfit functional

$${}_n L = \sum_{i,k} |{}_n \mathbf{e}_{ik}|^2 = \sum_{i,k} {}_n \mathbf{e}_{ik} : {}_n \mathbf{e}_{ik}^* = [{}_n \mathbf{T}_{ik} - {}_n f \mathbf{g}_{ik}] : [{}_n \mathbf{T}_{ik} - {}_n f \mathbf{g}_{ik}]^*. \quad (\text{A14})$$

Forcing the derivative of  ${}_n L$  with respect to the source corrector  ${}_n f$  to zero, yields again a normal equation

$$0 = \operatorname{Re} {}_n f \sum_{i,k} |\mathbf{g}_{ik}|^2 - \operatorname{Re} \sum_{i,k} {}_n \mathbf{T}_{ik} : \mathbf{g}_{ik}^*. \quad (\text{A15})$$

Rearranging eq. (A15) gives an equation for each source corrector  ${}_n f$  as a function of the transfer matrix  ${}_n \mathbf{T}_{ik}$  and the previously computed propagation corrector  $\mathbf{g}_{ik}$ :

$${}_n f = \frac{\operatorname{Re} \sum_{i,k} {}_n \mathbf{T}_{ik} : \mathbf{g}_{ik}^*}{\sum_{i,k} |\mathbf{g}_{ik}|^2}. \quad (\text{A16})$$

As in the scalar case we note that the imaginary part, and therefore the phase, of the source correctors are not constrained by the minimization procedure. The physically most meaningful solution is again to require the imaginary part of  ${}_n f$  to vanish, thereby ensuring that the phase of the original psd  $S$  equals the phase of the effective psd  $\widehat{S}$ .

## APPENDIX B: GREEN FUNCTION VARIATION AND ADJOINT SOURCES

### B1 The variation of the Green function due to variations in model parameters

We derive an explicit expression of the variation  $\delta G_{ij}$  of the Green function for an elastic medium. The corresponding expression for scalar wavefields in acoustic media follows as a special case. Our derivation rests on the adjoint method, that has been described in more detail and analysed from different perspectives by various authors (e.g. Tarantola 1988; Fichtner *et al.* 2006; Liu & Tromp 2006; Plessix 2006; Chen 2011).

The frequency domain Green function  $\mathbf{G}_j(\mathbf{m}; \mathbf{x}, \boldsymbol{\xi})$  with vector components  $G_{ij}(\mathbf{m}; \mathbf{x}, \boldsymbol{\xi})$  is the solution of the wave equation when the right-hand side is a point-localized force at position  $\boldsymbol{\xi}$  acting in  $j$ -direction, that is,

$$\mathcal{L}_x(\mathbf{m}) \mathbf{G}_j(\mathbf{m}; \mathbf{x}, \boldsymbol{\xi}) = \mathbf{e}_j \delta(\mathbf{x} - \boldsymbol{\xi}). \quad (\text{B1})$$

The generic forward modelling operator  $\mathcal{L}_x$  is a linear differential operator that acts on the spatial variable  $\mathbf{x}$  and depends on the model parameters  $\mathbf{m}$ . In the specific case of an anisotropic medium without rotation and self-gravitation, eq. (B1) becomes (e.g. Dahlen & Tromp 1998; Kennett 2001; Aki & Richards 2002)

$$\mathcal{L}_x(\mathbf{m}) \mathbf{G}_j(\mathbf{m}; \mathbf{x}, \boldsymbol{\xi}) = -\omega^2 \rho(\mathbf{x}) \mathbf{G}_j(\mathbf{m}; \mathbf{x}, \boldsymbol{\xi}) - \nabla \cdot [\mathbf{C}(\mathbf{x}) : \nabla \mathbf{G}_j(\mathbf{m}; \mathbf{x}, \boldsymbol{\xi})] = \mathbf{e}_j \delta(\mathbf{x} - \boldsymbol{\xi}), \quad (\text{B2})$$

with the fourth-order elastic tensor  $\mathbf{C}$ . Other forms are possible, depending on the physics that needed to be modelled. The first variation of the  $i$ -component of  $\mathbf{G}_j$  with respect to the model parameters  $\mathbf{m}$  is

$$\delta G_{ij}(\mathbf{x}, \xi) = \int_{\mathbf{x}' \in \oplus} \mathbf{e}_i \cdot \delta \mathbf{G}_j(\mathbf{x}', \xi) \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}', \quad (\text{B3})$$

where we omitted  $\mathbf{m}$  in the notation to avoid clutter. To eliminate  $\delta \mathbf{G}_j$  from eq. (B3), we use the variation of eq. (B1)

$$\delta \mathcal{L}_{\mathbf{x}'} \mathbf{G}_j(\mathbf{x}', \xi) + \mathcal{L}_{\mathbf{x}'} \delta \mathbf{G}_j(\mathbf{x}', \xi) = \mathbf{0}. \quad (\text{B4})$$

Multiplying eq. (B4) by the test field  $\mathbf{u}^\dagger(\mathbf{x}')$ , integrating over space, and adding the result to eq. (B3), gives

$$\delta G_{ij}(\mathbf{x}, \xi) = \int_{\mathbf{x}' \in \oplus} \mathbf{e}_i \cdot \delta \mathbf{G}_j(\mathbf{x}', \xi) \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}' + \int_{\mathbf{x}' \in \oplus} \mathbf{u}^\dagger(\mathbf{x}') \cdot [\delta \mathcal{L}_{\mathbf{x}'} \mathbf{G}_j(\mathbf{x}', \xi)] d\mathbf{x}' + \int_{\mathbf{x}' \in \oplus} \mathbf{u}^\dagger(\mathbf{x}') \cdot [\mathcal{L}_{\mathbf{x}'} \delta \mathbf{G}_j(\mathbf{x}', \xi)] d\mathbf{x}'. \quad (\text{B5})$$

We can rearrange the third term in eq. (B5) using the adjoint  $\mathcal{L}_{\mathbf{x}'}^\dagger$  of  $\mathcal{L}_{\mathbf{x}'}$ , defined by  $\int \mathbf{u}^\dagger(\mathbf{x}') \cdot [\mathcal{L}_{\mathbf{x}'} \delta \mathbf{G}_j(\mathbf{x}', \xi)] d\mathbf{x}' = \int \delta \mathbf{G}_j(\mathbf{x}', \xi) \cdot [\mathcal{L}_{\mathbf{x}'}^\dagger \mathbf{u}^\dagger(\mathbf{x}')] d\mathbf{x}'$ :

$$\delta G_{ij}(\mathbf{x}, \xi) = \int_{\mathbf{x}' \in \oplus} \delta \mathbf{G}_j(\mathbf{x}', \xi) \cdot [\mathbf{e}_i \delta(\mathbf{x} - \mathbf{x}') + \mathcal{L}_{\mathbf{x}'}^\dagger \mathbf{u}^\dagger(\mathbf{x}')] d\mathbf{x}' + \int_{\mathbf{x}' \in \oplus} \mathbf{u}^\dagger(\mathbf{x}') \cdot [\delta \mathcal{L}_{\mathbf{x}'} \mathbf{G}_j(\mathbf{x}', \xi)] d\mathbf{x}'. \quad (\text{B6})$$

Specifying the so far arbitrary test function  $\mathbf{u}^\dagger(\mathbf{x}')$  to be the solution of the *adjoint equation*

$$\mathcal{L}_{\mathbf{x}'}^\dagger \mathbf{u}^\dagger(\mathbf{x}') = -\mathbf{e}_i \delta(\mathbf{x} - \mathbf{x}'), \quad (\text{B7})$$

determines  $\mathbf{u}^\dagger(\mathbf{x}')$  to be the negative adjoint Green function with unit force in  $i$ -direction at position  $\mathbf{x}$ , that is,

$$\mathbf{u}^\dagger(\mathbf{x}') = -\mathbf{G}_i^\dagger(\mathbf{x}', \mathbf{x}). \quad (\text{B8})$$

Inserting eqs (B7) and (B8) into eq. (B6), the variation of the Green function simplifies to

$$\delta G_{ij}(\mathbf{x}, \xi) = - \int_{\mathbf{x}' \in \oplus} \mathbf{G}_i^\dagger(\mathbf{x}', \mathbf{x}) \cdot [\delta \mathcal{L}_{\mathbf{x}'} \mathbf{G}_j(\mathbf{x}', \xi)] d\mathbf{x}'. \quad (\text{B9})$$

Eq. (B9) gives the variation of the Green function in terms of a scalar product of the Green function itself and its adjoint. This integral kernel of the scalar product is the variation of the forward modelling operator  $\delta \mathcal{L}_{\mathbf{x}}$ . Defining the Green tensor as  $\mathbf{G} = (\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3)$  and  $\delta \mathcal{L}_{\mathbf{x}} \mathbf{G} = (\delta \mathcal{L}_{\mathbf{x}} \mathbf{G}_1, \delta \mathcal{L}_{\mathbf{x}} \mathbf{G}_2, \delta \mathcal{L}_{\mathbf{x}} \mathbf{G}_3)$ , we obtain a tensorial version of eq. (B9),

$$\delta \mathbf{G}(\mathbf{x}, \xi) = - \int_{\mathbf{x}' \in \oplus} [\mathbf{G}^\dagger(\mathbf{x}', \mathbf{x})]^T \delta \mathcal{L}_{\mathbf{x}} \mathbf{G}(\mathbf{x}', \xi) d\mathbf{x}'. \quad (\text{B10})$$

Replacing all vectorial and tensorial quantities in the previous derivation by scalars lead to the analogue of eq. (B10) for acoustic wave propagation, used in Section 4.2.

## B2 Adjoint sources for traveltime and energy measurements

In the following we derive the adjoint sources for finite-frequency traveltime and waveform energy measurements, used in the numerical examples of Section 4.

### B2.1 Cross-correlation travelttime measurement

Following work by Luo & Schuster (1991) and Dahlen *et al.* (2000), we define the travelttime shift  $T$  between a processed, observed interferogram  $\tilde{I}(t)$  and a synthetic, effective interferogram  $\hat{I}(t)$  as the time where the time-domain correlation of both reaches its global maximum:

$$\chi = T = \arg \max \int_{-\infty}^{\infty} \hat{I}(\tau) w(\tau - t) \tilde{I}(\tau - t) d\tau. \quad (\text{B11})$$

The time window  $w(t)$  may be used to focus on a specific seismic phase. For  $T > 0$ , the synthetic  $\hat{I}(t)$  is delayed relative to the observation  $\tilde{I}(t)$ , and vice versa. The definition of  $T$  implies that the time derivative of (B11) vanishes for  $t = T$ :

$$0 = \int_{-\infty}^{\infty} \hat{I}(\tau) \frac{d}{dt} [w(\tau - t) \tilde{I}(\tau - t)]_{t=T} d\tau. \quad (\text{B12})$$

While the traveltime shift  $T$  is only defined implicitly through eq. (B12), its variation  $\delta T$  with respect to the model parameters  $\delta \mathbf{m}$  can be expressed explicitly with the help of implicit function differentiation:

$$\delta T = \frac{\int_{-\infty}^{\infty} \delta \widehat{I}(\tau) \frac{d}{d\tau} [w(\tau - t) \circ \widetilde{I}(\tau - t)]_{t=T} d\tau}{\int_{-\infty}^{\infty} \frac{d}{d\tau} \widehat{I}(\tau) \frac{d}{d\tau} [w(\tau - t) \circ \widetilde{I}(\tau - t)]_{t=T} d\tau}. \quad (\text{B13})$$

Making the common assumption that the observed waveform is a time-shifted copy of the synthetic waveform (e.g. Dahlen *et al.* 2000), that is,  $w(\tau - T) \circ \widetilde{I}(\tau - T) = w(\tau) \widehat{I}(\tau)$ , eq. (B13) simplifies to

$$\delta T = \frac{\int_{-\infty}^{\infty} \delta \widehat{I}(\tau) \frac{d}{d\tau} [w(\tau) \widehat{I}(\tau)] d\tau}{\int_{-\infty}^{\infty} \frac{d}{d\tau} \widehat{I}(\tau) \frac{d}{d\tau} [w(\tau) \widehat{I}(\tau)] d\tau}. \quad (\text{B14})$$

Invoking Plancherel's theorem, we can write eq. (B14) in the frequency domain

$$\delta T = -i \frac{\int_{-\infty}^{\infty} \omega [w(\omega) * \widehat{I}(\omega)]^* \delta \widehat{I}(\omega) d\omega}{\int_{-\infty}^{\infty} \omega^2 [w(\omega) * \widehat{I}(\omega)]^* \widehat{I}(\omega) d\omega}. \quad (\text{B15})$$

Eq. (B15) can be brought into the canonical form of eq. (35), that is,  $\delta \chi = \delta T = 2\text{Re} \int_0^\infty f \delta \widehat{I} d\omega$ , when the adjoint source  $f$  is defined as

$$f(\omega) = -i \frac{\omega [w(\omega) * \widehat{I}(\omega)]^*}{\int_{-\infty}^{\infty} \omega^2 [w(\omega) * \widehat{I}(\omega)]^* \widehat{I}(\omega) d\omega}, \quad (\text{B16})$$

where  $*$  denotes a frequency-domain convolution. The approximation leading from eq. (B13) to eq. (B14) makes the adjoint source and the resulting sensitivity kernels effectively data-free. While being convenient for our numerical examples, the validity of this approximation must be checked on a case-by-case basis in real-data applications.

## B2.2 Waveform energy measurement

We define the waveform energy  $E$  as the  $L_2$  norm of a windowed portion of the synthetic interferogram, that is,

$$\chi = E = \frac{1}{2} \int_{-\infty}^{\infty} [w(t) \widehat{I}(t)]^2 dt. \quad (\text{B17})$$

Taking the variation of eq. (B17), and again using Plancherel's theorem to transform to the frequency domain, we find

$$\delta E = \int_{-\infty}^{\infty} w^2(t) \widehat{I}(t) \delta \widehat{I}(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} [w(\omega) * w(\omega) * \widehat{I}(\omega)]^* \delta \widehat{I}(\omega) d\omega. \quad (\text{B18})$$

We obtain the canonical form  $\delta \chi = \delta E = 2\text{Re} \int_0^\infty f \delta \widehat{I} d\omega$  when the adjoint source  $f$  is defined as

$$f(\omega) = \frac{1}{\pi} [w(\omega) * w(\omega) * \widehat{I}(\omega)]^*, \quad (\text{B19})$$

which is simply the frequency-domain version of the doubly windowed and time-reversed interferogram.