

The exponentiated phase measurement, and objective-function hybridization for adjoint waveform tomography

Yanhua O. Yuan,¹ Ebru Bozdağ²,³ Caio Ciardelli,³ Fuchun Gao⁴ and F. J. Simons¹

¹*Department of Geosciences, Princeton University, Princeton, NJ 08544, USA*

²*Department of Geophysics, Colorado School of Mines, Golden, CO 80401, USA. E-mail: bozdag@mines.edu*

³*Department of Geophysics, Institute of Astronomy, Geophysics and Atmospheric Sciences, SP 05508, Brazil*

⁴*TOTAL EP Research and Technology, Houston, TX 77002, USA*

Accepted 2020 January 31. Received 2020 January 17; in original form 2019 April 26

SUMMARY

Seismic tomography has arrived at the threshold of the era of big data. However, how to extract information optimally from every available time-series remains a challenge; one that is directly related to the objective function chosen as a distance metric between observed and synthetic data. Time-domain cross-correlation and frequency-dependent multitaper traveltime measurements are generally tied to window selection algorithms in order to balance the amplitude differences between seismic phases. Even then, such measurements naturally favour the dominant signals within the chosen windows. Hence, it is difficult to select all usable portions of seismograms with any sort of optimality. As a consequence, information ends up being lost, in particular from scattered waves. In contrast, measurements based on instantaneous phase allow extracting information uniformly over the seismic records without requiring their segmentation. And yet, measuring instantaneous phase, like any other phase measurement, is impeded by phase wrapping. In this paper, we address this limitation by using a complex-valued phase representation that we call ‘exponentiated phase’. We demonstrate that the exponentiated phase is a good substitute for instantaneous-phase measurements. To assimilate as much information as possible from every seismogram while tackling the non-linearity of inversion problems, we discuss a flexible hybrid approach to combine various objective functions in adjoint seismic tomography. We focus on those based on the exponentiated phase, to take into account relatively small-magnitude scattered waves; on multitaper measurements of selected surface waves; and on cross-correlation measurements on specific windows to select distinct body-wave arrivals. Guided by synthetic experiments, we discuss how exponentiated-phase, multitaper and cross-correlation measurements, and their hybridization, affect tomographic results. Despite their use of multiple measurements, the computational cost to evaluate gradient kernels for the objective functions is scarcely affected, allowing for issues with data quality and measurement challenges to be simultaneously addressed efficiently.

Key words: Inverse theory; Time-series analysis; Seismic tomography.

1 INTRODUCTION

Driven in part by recent advances in numerical wave-propagation algorithms and high-performance computing, full-waveform inversion has been used to study Earth’s interior at a variety of scales (e.g. Fichtner *et al.* 2008; Virieux & Operto 2009; Tape *et al.* 2009; Prieux *et al.* 2013a, b; Borisov *et al.* 2016). The quality and resolution with which seismic tomography manages to image the 3-D structure of Earth’s interior rely on its capability to match the individual observed seismograms—and on overall geographic data coverage (e.g. Nolet *et al.* 2019). Despite being at the threshold of a new era of big data in exploration and earthquake seismology, a potentially vast amount of the information contained in seismograms remains insufficiently usable in our inversions, due to measurement and data assimilation difficulties. The success of wave-equation-based tomography techniques is inextricably reliant on the nature of the misfit functions that measure the distance between the forward-computed wavefield and the observations (e.g. Modrak & Tromp 2016), which drive the inversion process. For instance, reducing traveltime-based misfit functions (Luo & Schuster 1991; Phillips & Fehler 1991; Marquering *et al.* 1999; Dahlen *et al.* 2000; Rawlinson & Sambridge 2003; Zhou *et al.* 2004; Liu & Tromp 2006; Tape 2009; Burdick *et al.* 2010) is a popular way to map wavespeed heterogeneity

within the Earth. To make these measurements, seismic phases with similar shapes are to be selected, for example using explicit measures of coherence (e.g. Maggi *et al.* 2009), or increasingly, based on machine-learning approaches (e.g. Chen *et al.* 2017).

The careful selection of these phases is essential to maximizing the information entering the inversion procedure. It is difficult to find an optimal set of window-selection parameters by which to include as much signal as possible from all given seismograms while avoiding noisy portions. Conservative approaches might exclude potentially useful information; more liberal selection criteria could incorporate segments that reveal themselves as insignificant for the inversion process. Either way, the inclusion of ill-fitting data portions could drive the inversion to non-convergence. Scattered waves, in particular, are rarely included by such automated procedures, yet they are crucial to resolving slow-wavespeed anomalies (Rickers *et al.* 2012), due to the wave phenomenon known as ‘wavefront healing’ (Nolet & Dahlen 2000; Hung *et al.* 2001). To conclude: data selection, the choice of misfit function and the judicious combination of multiple observables are as important now as ever. Next, we report on two particular aspects of these vital steps in the tomographic waveform-inversion process.

1.1 Importance and challenges of using phase information

Various approaches to maximize extraction of information from seismograms have been the subject of previous studies. Fichtner *et al.* (2008) discussed full-waveform inversion based on the separation of phase and amplitude through time–frequency analysis (Kristeková *et al.* 2006). Their method provides a natural way to include all necessary information from seismograms while reducing the non-linearity caused by the interaction of phase and amplitude. Bozdağ *et al.* (2011) proposed tomography based on the use of ‘instantaneous phase’ and envelopes. Avoiding additional processing when compared to time-domain approaches, the instantaneous phase has been shown to hold an advantage over cross-correlation in capturing traveltimes delays, as discussed by Djebbi & Alkhalifah (2014). Rickers *et al.* (2012) used instantaneous-phase measurements successfully to image mantle plumes, which are difficult to resolve by cross-correlation traveltimes tomography.

Like all the other phase measurements, calculating instantaneous phase is hampered by ‘cycle skipping’, with the phase wrapped after one cycle. Phase unwrapping is not trivial, especially for complicated or noisy signals—a sticky problem (Pritt & Ghiglia 1998). An alternative idea is to measure phase implicitly from complex-valued signals. To estimate phase in the frequency domain, Shin *et al.* 2003 calculated the imaginary part of a complex-valued wavefield and its derivative with respect to each frequency. Working in the time domain, Sguazzero *et al.* (1987) introduced a normalized complex correlation to estimate stacking velocities for multichannel seismic surveys. Luo *et al.* (2018) discussed using a normalized complex form of phase combined with a damping method to tackle the phase wrapping problem.

Along the same lines, we use an implicit approach to extract phase information in the time domain, one that we call ‘exponentiated phase’. Given the analytic representation of a real-valued signal, after stripping its amplitude, what remains is a complete representation of its phase information. We show that the exponentiated phase and the adjoint source needed to incorporate it into full-waveform inversion are naturally connected to the instantaneous-phase methods introduced by Bozdağ *et al.* (2011). However, in contrast to the measurement challenges of instantaneous phase, the exponentiated phase utilizes the original phase information from the analytical signal, and measures phase uniquely, consistently and robustly. Also, it is extremely straightforward to apply the exponentiated phase method to any type of signal with complex mixtures of waves, rendering it applicable to handle currently available big datasets. However, we show that the exponentiated-phase misfit function, when the phase difference is too large, does not preserve the desirable convexity of unwrapped instantaneous-phase and cross-correlation misfit functions. To address this limitation, and to balance the trade-offs between convergence and model recovery, we propose to combine exponentiated-phase with other phase measurements such as cross-correlation traveltimes and multitaper phase. In the following section, we discuss our ‘objective-function hybridization’ approach.

1.2 Combining multiple observation types

Broadband seismograms typically contain multiple seismic events with distinct spectral-temporal features. To characterize these features, one can apply hierarchical, multiscale strategies by progressively incorporating data filtered from long to short wavelengths, either through conventional Fourier transformation (Bunks *et al.* 1995) or by multiresolution wavelet decomposition (Yuan & Simons 2014; Yuan *et al.* 2015). Alternatively, one can use different objective functions to capture distinct seismic phases (Maraschini *et al.* 2010; Bozdağ *et al.* 2011). Traveltimes tomography (see, e.g. Nolet 2008) can behave robustly while measuring only relatively short-period body waves. To characterize dispersive body and surface waves, multitaper approaches (Zhou *et al.* 2004, 2005; Tape *et al.* 2010) have been eminently suitable. And as discussed here, the exponentiated phase is capable of including various small-magnitude scattered waves in addition to their primary dominant phases without windowing data into small segments.

To optimize the information that can be extracted from every seismic wiggle, we propose combining various objective functions. At the long periods, we count on the relatively good agreement for major phases to readily take into account scattered waves. At the shorter periods, the time-domain selection allows for the identification of prominent body-wave arrivals without complicating their measurement due to non-linearity. Objective-function hybridization embraces data-quality and measurement challenges in a way that moves seismic tomography towards using any and all information that can be mined from our datasets.

In the literature (e.g. Roy 1971), approaches to blending different objective functions encompass both the hierarchical (Haimes & Li 1988) or the simultaneous (Sabri & Beamon 2000). In this paper, we focus on the latter, as it largely leaves the computational cost in calculating gradients unaffected, compared to when a single objective function is involved. Furthermore, no measurement order is imposed.

1.3 Main components of the paper

We review and numerically compare misfit functions based on various traveltime or phase measurements (we do not strictly distinguish traveltime, t , and phase, ϕ , information as they are connected for a particular frequency via the relation $\Delta\phi = -\omega\Delta t$), including cross-correlation traveltime, frequency-dependent traveltime, and time-dependent instantaneous phase measurements. We show the importance of using the temporal phase for scattered waves through balancing waves with different magnitudes, and the challenge of addressing cycle skips in computing it. We discuss how the alternative, time-continuous phase measurement—the exponentiated phase—successfully circumvents phase-wrapping, while representing structural information similarly from the viewpoint of small-perturbation theory. Leaning on both theoretical derivations and numerical examples, we highlight the connections between the exponentiated phase and the instantaneous phase (Bozdağ *et al.* 2011; Rickers *et al.* 2012).

We test the performance of different phase-based measurements in a synthetic model. The differences in their sensitivity kernels inspire a joint inversion of various objective functions, and the results clearly show great improvement of the inverted models. We test our phase measurements on a real dataset from one event recorded by the Incorporated Research Institutions for Seismology (IRIS) Global Seismographic Network (GSN) by computing its ‘event kernel’, the necessary first step for global adjoint seismic tomography, which ultimately will involve all available earthquakes and recording stations.

In short, we show how, by combining a multitude of different seismic measurements designed to capture all reliable information from recorded seismograms including body waves, surface waves and even interfering phases, and incorporating it into a single objective function, the resolution of global seismic models (e.g. French & Romanowicz 2015; Bozdağ *et al.* 2016) will be improved—without much affecting the computational cost in the adjoint seismic tomography workflow.

2 OBJECTIVE-FUNCTION HYBRIDIZATION

Broad-band seismograms comprise multiple seismic phases (e.g. body-wave arrivals such as direct P and S waves and their reflected, refracted and converted phases, surface waves, etc.) with different dominant periods. Overall, the period range of interest varies from exploration to earthquake seismology. In exploration settings, generally seismic frequencies higher than 1 Hz are of interest, while in earthquake (continental to global-scale) seismology body waves up to 1 Hz dominant frequency, longer-period surface waves and even longer-period normal modes are considered. In the experiment with real earthquake data, we focus on records within the period range of 17–250 s, the band used in our own first examples of global adjoint tomography models (e.g. Bozdağ *et al.* 2016), but these frequency restrictions are not limiting.

2.1 Heuristic motivation

Fig. 1 compares a 180-min-long radial-component observed seismogram from the September 3, 2010 New Zealand earthquake ($M_w = 7$, depth 12 km) recorded at station KBL in Afghanistan, over different period bands. The synthetic seismogram was calculated for the Centroid Moment Tensor (CMT) mechanism 201009031635A by the global spectral-element seismic wave propagation solver SPECFEM3D.GLOBE (Komatitsch & Tromp 2002a, b) in a rotating, elliptical, self-gravitating (under the Cowling approximation) Earth, accommodating the effects of topography and seafloor bathymetry, ocean loading, and a global adjoint tomography model for 3-D crust and mantle structure, GLAD-M15 (Bozdağ *et al.* 2016).

As seen in the figure, synthetic seismograms explain the observations adequately at the long periods, but mismatches are increasingly apparent in the shorter-period bands, where waveform complexity significantly increases. Different segments or portions of the data tend to be sensitive to different types and domains of subsurface structure in seismic tomography. The frequency-dependent misfits shown in these records reveal the vast amount of information that remains to be assimilated by ‘full’-waveform inversions, and the progressive waveform complexity hints at the difficulties likely to be encountered in attempting to do so, as the inverse problem will become significantly non-linear. No single waveform misfit metric is expected to carry the inversions through the iterations and across the frequency bands, hence our goal is to derive strategies that are flexible in their choice of objective functions and adept at handling their combination, thereby ultimately aiming to increase the quantity of information that can be extracted from every seismic trace.

2.2 Hierarchical optimization

A common way to address the non-linearity of full-waveform inversion is to use a hierarchical approach (Roy 1971; Haines & Li 1988) with multiple objective functions. The inverse problem starts with a linearization, generally focusing on certain specific major phases, followed by another objective function that allows for the extraction of additional information. Especially in the case of earthquake seismology, traveltime/phase misfit is the more commonly considered objective function, as it is more linearly related to wavespeed than amplitude. Although not utilized in this work, seismic amplitudes do provide good constraints on Earth’s structure (both elastic and anelastic)—together with phases or time-shifts (e.g. Laske & Masters 1996; Selby & Woodhouse 2000; Dalton & Ekström 2006; Zhu *et al.* 2013; Karaoğlu & Romanowicz 2018). In the hierarchical method, the optimal solution to the most preferred objective function is used as a constraint for the next optimization using a new misfit function, and so on, for a number of levels.

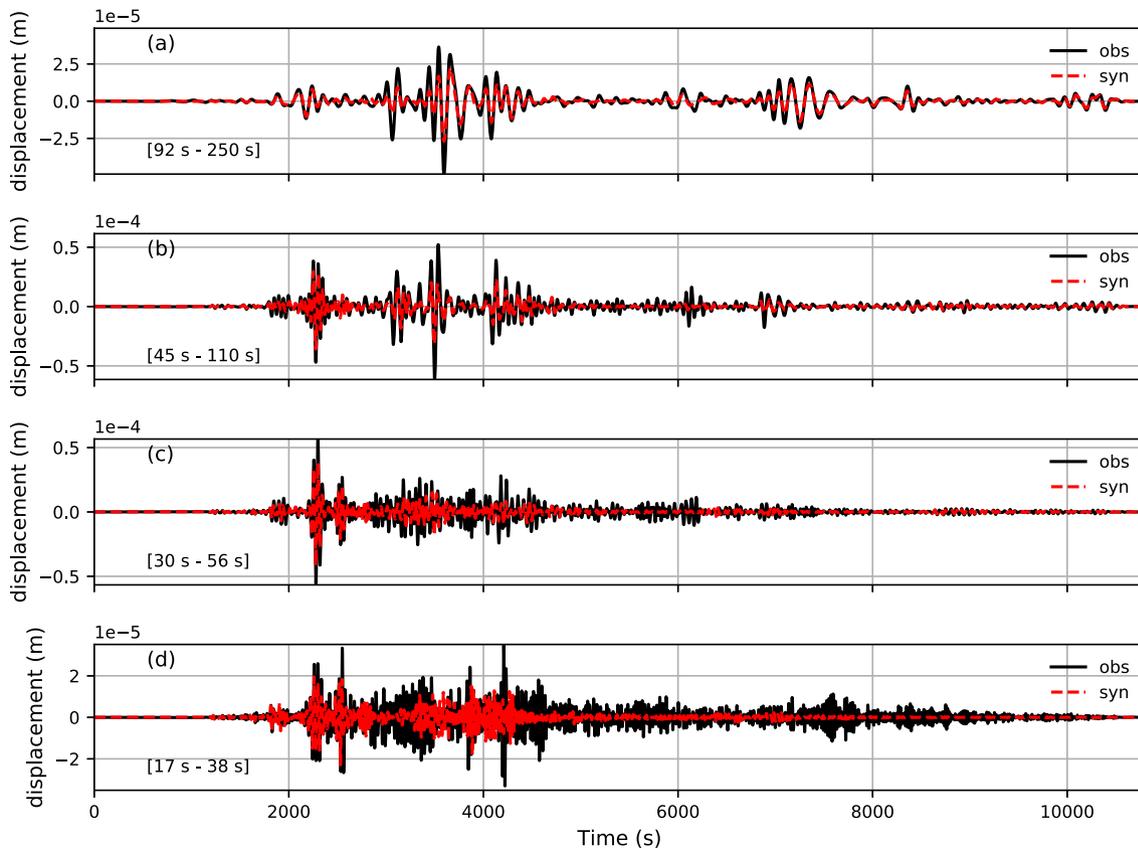


Figure 1. An example of observed and synthetic waveforms filtered to different period bands: radial-component seismograms, for a station within the Global Seismographic Network. The solid black lines are the observations, the red traces the computed predictions. The panels show different period ranges arranged from top to bottom, as the labels indicate. Improving the misfit at the short periods appears as important as it is challenging.

The order in which the objective functions are presented to the optimization suite is fruitfully determined based on physical assumptions. For example: early-arriving phases are inverted first to correct for shallow structure before involving later signals (Sheng *et al.* 2006); near-offsets capturing reflected energy give way to far-offset traces that record transmitted and converted waves; starting from the low frequencies, seismograms are gradually incorporated into the inversion by filtering to progressively higher frequencies to retrieve subsurface features at incrementally finer scales (Bunks *et al.* 1995; Akçelik *et al.* 2002; Operto *et al.* 2004; Yuan & Simons 2014); phase measurements are dominantly sensitive to elastic parameters, while the amplitude mismatch can be used to further constrain both elastic and anelastic properties.

2.3 Simultaneous optimization

A multi-objective optimization problem can also be solved using a simultaneous approach (Sabri & Beamon 2000). In contrast to the hierarchical process, no strict order of the objective functions is required, which renders the procedure attractive from the standpoint of computational efficiency. Only one adjoint simulation is needed to evaluate the misfit-gradient kernels when various objectives are combined into a single, hybrid, objective function.

A generic expression for the simultaneous combination of objective functions into $\chi_H(\mathbf{m})$, a hybrid metric on the model space containing the parameters \mathbf{m} , can be written as:

$$\chi_H(\mathbf{m}) = \sum_c^{N_c} w_c \frac{\chi_c(\mathbf{m})}{\sigma_c^2}, \quad (1)$$

where w_c is a weighting factor to balance the relative importance of measurement classes through their individual objective functions $\chi_c(\mathbf{m})$. Our expression contains an explicit term representing the measurement uncertainty, σ_c . In the following pages we will discuss how to select or measure these parameters, and illustrate our choices through numerical experiments.

3 OBJECTIVE FUNCTIONS AND THEIR ADJOINT SOURCES

We begin by briefly revisiting the traveltimes/phase measurements that have been widely used in waveform-based seismic tomography to date. Our purpose is to evaluate their scope of application and their contribution towards the assimilation of ‘all’ the available information from the seismograms. We call upon numerical examples to demonstrate how to make these measurements and construct their adjoint sources, and discuss limitations and challenges.

3.1 Cross-correlation traveltimes [CC]

In regional and global seismology, the use of cross-correlation to measure the traveltimes of compressional (P) and shear (S) body waves has long been the cornerstone in the tomographic mapping of elastic wavespeed variations in the crust and mantle. In exploration geophysics, traveltimes tomography or refraction tomography, using first arrivals picked manually or automatically (e.g. Gelchinsky & Shtivelman 1983), similarly often provide initial velocity models for subsequent pre-stack depth migration or full-waveform inversion.

Cross-correlation traveltimes Δt and amplitude $\Delta \ln A$ measurements are carried out in the time domain on the seismograms, and typically target specific seismic phases via time-window selection, whether based on preset criteria (e.g. Maggi *et al.* 2009), or, lately, from machine learning (e.g. Chen *et al.* 2017). Considering only the cross-correlation traveltimes, we measure the time delay between a synthetic, $s(t)$ and an observation, $d(t)$, over a window of length T , by when their waveform cross-correlation function is maximized,

$$\Delta t = \arg \max_{\tau} \int_0^T s(t + \tau) d(t) dt. \quad (2)$$

A positive Δt indicates that the data waveform $d(t)$ is advanced relative to the synthetic $s(t)$, meaning that the wavespeed model that generates the synthetics is slower than the truth. Hence, to reduce the mismatch between $d(t)$ and $s(t)$, one needs to increase the wavespeeds in the model. In the optimization step, the wavespeed parameters are tuned until our objective function, the sum of squares of traveltimes shifts for all sources s and receivers r , in other words:

$$\chi_{CC} = \frac{1}{2} \sum_{s,r} [\Delta t_{s,r}]^2, \quad (3)$$

is minimized. From the Fréchet derivatives of the terms in the cross-correlation traveltimes misfit function in eq. (3), as shown by various authors (e.g. Luo & Schuster 1991; Marquering *et al.* 1999; Dahlen *et al.* 2000), Tromp *et al.* (2005) derive an adjoint source in the form, for each earthquake source (omitting the subscript s),

$$f_{CC}^\dagger(\mathbf{x}, t) = \sum_r \Delta t_r \left[\frac{\partial_t s_r(T-t)}{\int_0^T \partial_t^2 s_r(t) s_r(t) dt} \right] \delta(\mathbf{x} - \mathbf{x}_r). \quad (4)$$

This virtual source takes care of injecting the cross-correlation traveltimes misfit information into the adjoint wave-propagation simulations that drive the inversion, for all receivers r .

To maximize the information entering the tomographic inversion workflow to achieve optimum illumination, window selection is key. But optimal design of selection criteria and parameters is not trivial. Only clearly identifiable phases with similar waveforms in both observed and computed seismograms can be selected for the cross-correlation traveltimes measurement. Therefore, many useful signals including various interfering waves, which are valuable to constrain small-scale structure, are often removed during automated data selection procedures. Even when the signals are successfully picked, the majority of the selected information will not be fully utilized, as cross-correlation only measures a single value in a window that will be primarily influenced by the largest-amplitude pulse.

On the one hand, coupled with a robust window selection algorithm, cross-correlation traveltimes tomography will behave almost linearly and perform excellently for well-recognized and non-dispersive body waves. On the other hand, the tomographic resolution that it can ultimately achieve will be limited, as a significant amount of information in the waveforms will be either missed, not adequately considered, or bluntly lumped together.

3.2 Multitaper frequency-dependent traveltimes [MT]

For strongly dispersive waves such as surface waves, frequency-dependent measurements of traveltimes and amplitude relations are commonly made using a multitaper technique (Laske & Masters 1996; Zhou *et al.* 2005; Hjörleifsdóttir 2007), which has been successful in revealing structural information in tomographic models beyond what can be achieved by typical applications of time-domain cross-correlation methods.

The frequency-dependent (at angular frequency ω) traveltimes measurement, $\Delta t(\omega)$, and the amplitude measurement, $\Delta \ln A(\omega)$, are attributes of a frequency-domain transfer function between the data and the synthetics. This transfer function is estimated via the multitaper method, which, prior to Fourier transformation, tapers the time-domain records $s(t)$ and $d(t)$ using a set $h_k^{T\Omega}(t)$ of prolate spheroidal functions (Slepian 1978; Thomson 1982; Simons 2010). These are uniquely defined for a particular window length T and a desired frequency half-bandwidth Ω , and commonly, $k = 1, \dots, [T\Omega/\pi] - 1$. Details on the measurement can be found in Appendix A1.

Limiting ourselves to the traveltimes measurements $\Delta t(\omega)$ —for considerations on amplitude, see Zhou *et al.* (2004) and Tape (2009)—our misfit function summing over all frequencies ω , sources s and receivers r , is

$$\chi_{\text{MT}} = \frac{1}{2} \sum_{s,r} \sum_{\omega} W_t(\omega) [\Delta t_{s,r}(\omega)]^2, \quad (5)$$

with a certain weighting function $W_t(\omega)$. As shown in Appendix A1, the adjoint source for the frequency-dependent time-delay misfit function for a single event is given by

$$f_{\text{MT}}^{\dagger}(\mathbf{x}, t) = \sum_r \sum_k h_k^{T\Omega}(t) \mathcal{F}^{-1} \left\{ W_t(\omega) \Delta t_{s,r}(\omega) \left[\frac{i}{\omega} \frac{s_k^*(\omega)}{\sum_{k'} s_{k'}(\omega) s_{k'}^*(\omega)} \right] \right\} (t) \delta(\mathbf{x} - \mathbf{x}_r), \quad (6)$$

where \mathcal{F}^{-1} is the inverse Fourier transform operator, and the s_k denote the tapered modeled traces, omitting for clarity the identifying subscripts s and r .

As with the cross-correlation method, application of the multitaper method also demands window selections. However, the multitaper approach is capable of extracting more information from each selected window.

3.3 Instantaneous phase [IP]

Bypassing the need for restrictive window selection and Fourier transformation required for the multitaper method, Bozdağ *et al.* (2011) discussed a method to isolate phase directly in the time domain. With $s_a(t)$ and $\mathcal{H}\{s(t)\}$ denoting the analytic signal and the Hilbert transform of a real signal $s(t)$, $\phi_s(t)$ and $E_s(t)$ stand for its instantaneous phase and amplitude or envelope, following

$$s_a(t) = s(t) + i\mathcal{H}\{s(t)\} = E_s(t) e^{i\phi_s(t)}. \quad (7)$$

It is worthwhile to note that here we follow the definition of the analytic signal in Oppenheim & Lim (1981), which has a different sign in the imaginary part of eq. (15) in Bozdağ *et al.* (2011). We have

$$\phi_s(t) = \arctan \frac{\mathcal{H}\{s(t)\}}{s(t)}, \quad \text{and} \quad E_s(t) = \sqrt{s^2(t) + \mathcal{H}^2\{s(t)\}}. \quad (8)$$

The instantaneous-phase misfit relies on the measurement between data $d(t)$ and synthetics $s(t)$,

$$\Delta\phi(t) = \phi_d(t) - \phi_s(t), \quad (9)$$

once again summing over all sources s and receivers r , integrated over a time window T ,

$$\chi_{\text{IP}} = \frac{1}{2} \sum_{s,r} \int_0^T [\Delta\phi_{s,r}(t)]^2 dt. \quad (10)$$

The adjoint source for an individual event can be written, without the subscripts, as

$$f_{\text{IP}}^{\dagger}(t) = \sum_r \left[\frac{\Delta\phi(t)}{E_s^2(t)} \mathcal{H}\{s(t)\} + \mathcal{H} \left\{ \frac{\Delta\phi(t)}{E_s^2(t)} s(t) \right\} \right] \delta(\mathbf{x} - \mathbf{x}_r). \quad (11)$$

See Bozdağ *et al.* (2011) and Yuan *et al.* (2015) for the equivalent expressions regarding the envelope.

As can be seen in eq. (11), the instantaneous-phase adjoint source takes into account amplitude effects by a normalization involving the square of the signal envelopes. Sensitivity kernels for the high-amplitude and low-amplitude segments within a single seismogram remain amplitude-independent. Practitioners are recommended to add a ‘water level’ term ϵ , some small percentage of its peak amplitude (e.g. 0.1–5 per cent), to the denominator, to avoid numerical instabilities and to regularize noisy portions.

3.4 Adjoint-source heuristics

In the previous paragraphs we (re-)introduced three different types of waveform-based time or phase measurements that relate observed and synthetic traces: cross-correlation traveltimes (CC), multitaper traveltimes (MT) and instantaneous phase (IP). We discussed objective functions based upon them, and the adjoint sources that they lead to. In this section we use a simple example to highlight similarities and discrepancies between these three approaches, with an eye towards evaluating their suitability for capturing small-magnitude signals in the seismogram.

The solid line in Fig. 2(a) shows an ‘observed’ seismic trace, and the dashed line a synthetic prediction. Both are modeled using SPECFEM_2D. The purple shadow reveals a selective time window. In Fig. 2(b-I), CC corresponds to a constant time delay between the observed and the predicted seismograms within the selected window, representative of the time-shift of the peak pulse around the center of the window. Fig. 2(b-II) shows the adjoint source for the CC traveltimes misfit function, computed using eq. (4). In Fig. 2(c-I), the multitaper method calculates traveltimes delays between the observed and the synthetic seismograms at different frequencies within the selected time window. Instead of displaying the measurements in the time domain, the MT traveltimes information is plotted in the frequency domain (the horizontal axis is the period from 30 to 3 s for this example, and the vertical axis is the time-shift in seconds at each frequency or period).

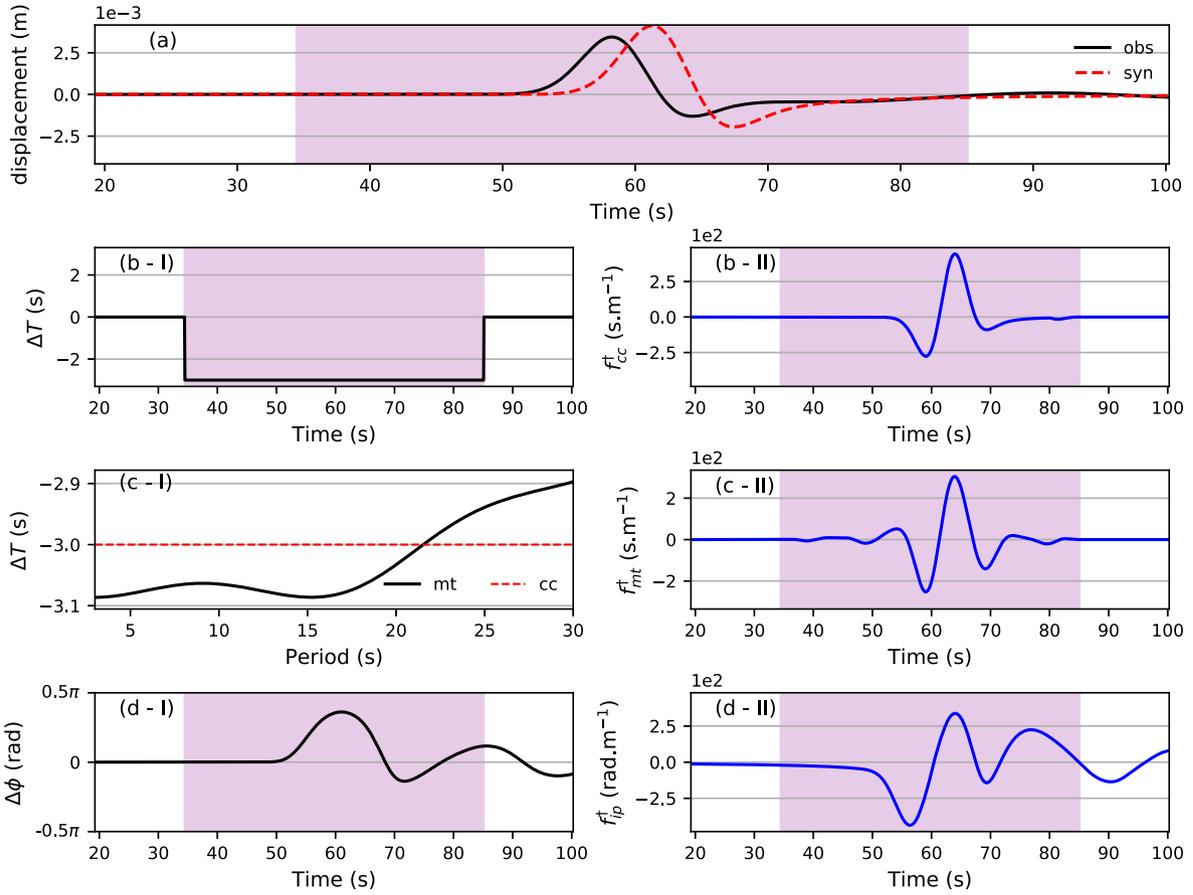


Figure 2. Similarities and discrepancies between cross-correlation (CC) time delay, frequency-dependent multitaper traveltime (MT) and instantaneous phase (IP) measurements between seismic traces. (a) An example of ‘observed’ (solid black line) and ‘synthetic’ (dashed red line) seismograms used to study different traveltime/phase measurements and their adjoint sources. The time window within which CC and MT measurements will be made is shaded purple. (b-I) The CC time delay between data and synthetics is constant over the time window. (b-II) The adjoint source for the CC traveltime. (c-I) MT traveltime difference measures variable time-shifts at different periods inside the purple time window. The MT traveltime information is plotted in the frequency domain (the purple time window is not rendered here) and the horizontal axis ranges linearly over periods from 30 down to 3 s. (c-II) The MT adjoint source is subtly different from the CC adjoint source in (b-II) due to the frequency-varying time delays. (d-I) Time-domain IP difference between the waveforms captures their phase discrepancies over the entire trace, without window selection. (d-II) The IP adjoint source naturally integrates information from the main signal and from subsequent minor arrivals into a single measurement.

What we observe is that MT computes relatively larger time delays at low frequencies and smaller shifts at higher frequencies. The MT adjoint source in Fig. 2(c-II) has a fairly good agreement with the adjoint source in Fig. 2(b-II) that uses the constant CC traveltime quantity, but it includes richer high-frequency information due to the frequency-dependent traveltime variations contained in the seismograms. Different from CC and MT, IP does not require strict window selection or segmentation of traces, thus we calculate the time-continuous phase differences (in terms of phase angles: $\Delta\phi = -\omega\Delta t$ for a specific frequency) between the observed and the synthetic waveforms in the entire time domain, which we show in Fig. 2(d-I). The adjoint source of IP in Fig. 2(d-II) is calculated using eq. (11), which involves a normalization term defined by the squared envelope of the synthetics, leading to relatively balanced sensitivities for small- and large-magnitude signals, thus both the main and the following minor signals are naturally included in one adjoint source.

3.5 Instantaneous phase: measurement challenges

Time-continuous phase measurements are appealing for many reasons discussed above. The measuring of instantaneous phase, however, can be challenging in practice. First, referring to eq. (8), the division causes numerical instability when signal magnitude is small, which requires regularization. In addition, the measured instantaneous phase can be discontinuous due to the non-uniqueness of the arc-tangent function in eq. (8), which is defined in the interval of $(-\pi/2, \pi/2]$. When signals cross zero, the measured phase wraps back onto the $(-\pi/2, \pi/2]$ interval. Instead of measuring the two individual phases separately (which involves two potentially unstable divisions and inverse tangents) and then subtracting them, one can preempt phase wrapping somewhat by alternatively computing the equivalent to eqs (8)–(9),

$$\Delta\phi(t) = \arctan \left[\frac{s(t) \mathcal{H}\{d(t)\} - d(t) \mathcal{H}\{s(t)\}}{d(t) s(t) + \mathcal{H}\{d(t)\} \mathcal{H}\{s(t)\}} \right]. \quad (12)$$

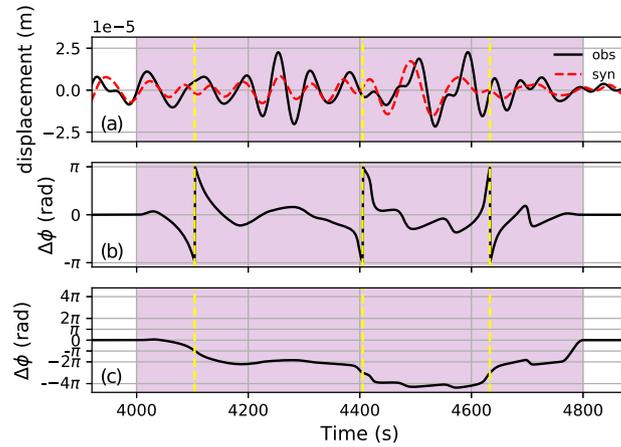


Figure 3. Phase wrapping and phase unwrapping. (a) A radial-component seismogram (solid) recorded at station NOQE from the August 25, 2008 Xizang earthquake ($M_w = 6.7$, depth 17 km), and the synthetic trace (dashed). Both data are filtered within the 45–110 s period band. (b) Their instantaneous phase difference shows clear phase jumps from $-\pi$ to π marked by yellow lines. (c) Phase measurements after regularized unwrapping.

The phase-wrapping condition is less likely to be triggered by the single expression above, since $d(t)$ and $s(t)$ would have to be identical, crossing zero in close proximity to ensure that both signals enter or exit a new cycle before their phase difference exceeds $\pm\pi/2$ ($\pm\pi$ in this example as we use the two-argument arc-tangent). To further improve the phase calculation, Rickers *et al.* (2012) suggested taking the multiplication of the complex conjugate of the synthetic analytic signal with the observed analytic signal normalized by their amplitudes (their eq. 8), and using the arc-sine to replace the arc-tangent, as the arc-sine function does not switch sign as soon as the value grows out of the interval limits.

We use a real example to demonstrate a situation where phase jumps occur while calculating phase difference using eq. (12). The black line in Fig. 3(a) depicts a radial-component displacement seismogram from the Global Seismographic Network for the August 25, 2008 Xizang Tibet earthquake ($M_w = 6.7$, depth 17 km, CMT 200808251322A) recorded at station NOQE; the red line displays the seismic trace computed using SPECFEM3D.GLOBE. The purple window is a window selected using FLEXWIN. In Fig. 3(b), we observed two types of phase jumps (locations marked by yellow lines). The first type happens after one of the signals has one more zero crossing, which brings it into a new cycle while the other signal is still in the old cycle. This causes their phase difference to exceed the π limit and leads to the phase jump. The first two phase jumps in Fig. 3(b) at 4104 and 4405 s are examples of this type of discontinuity. The second type of phase jump occurs when two signals have zero crossings at similar times, but enter into opposite cycles. The phase jump at 4633 s in Fig. 3(b) shows such an example when the observed data switches from negative to positive while the synthetic data goes from positive to negative. We successively unwrap these phase jumps by shifting all signals after the first discontinuity point by -2π , then shifting all signals after the second discontinuity point by -2π , and finally shifting all signals after the third discontinuity point by $+2\pi$. The unwrapped phase measurement is displayed in Fig. 3(c). Through this example, we aim to demonstrate the challenges related to measuring IP and the non-local effects of the unwrapping process. These issues are aggravated in the presence of noise, because even small perturbations in amplitude can cause phase jumps and therefore significant fluctuations in the phase estimate, which will then propagate to affect the entire measurement.

4 EXPONENTIATED PHASE [EP]

We normalize the analytic signal underlying the phase calculation by its corresponding envelope, to yield a new expression which we term the ‘exponentiated phase’ (EP),

$$\tilde{s}(t) = \frac{s(t)}{E_s(t)} + i \frac{\mathcal{H}\{s(t)\}}{E_s(t)} = e^{i\phi_s(t)}, \quad (13)$$

which defines phase completely, but implicitly.

4.1 Misfit function and adjoint source

We next define the ‘exponentiated-phase misfit function’ as the difference between the observed and the synthetic normalized analytic signals, denoted $\tilde{d}(t)$ and $\tilde{s}(t)$, as in eq. (13), summed over all sources s and receivers r ,

$$\chi_{\text{EP}} = \frac{1}{2} \sum_{s,r} \int_0^T \|\tilde{d}(t) - \tilde{s}(t)\|^2 dt = \frac{1}{2} \sum_{s,r} \int_0^T [\|\Delta R(t)\|^2 + \|\Delta I(t)\|^2] dt, \quad (14)$$

where $\Delta R(t)$, the difference in the real part, and $\Delta I(t)$, the difference in the imaginary part, are given by

$$\Delta R(t) = \frac{d(t)}{E_d(t)} - \frac{s(t)}{E_s(t)}, \quad \text{and} \quad \Delta I(t) = \frac{\mathcal{H}\{d(t)\}}{E_d(t)} - \frac{\mathcal{H}\{s(t)\}}{E_s(t)}. \quad (15)$$

Gradient-based methods require the derivative of the misfit function in eq. (14),

$$\delta\chi_{\text{EP}} = - \sum_{s,r} \int_0^T \left[\Delta R(t) \delta\left(\frac{s(t)}{E_s(t)}\right) + \Delta I(t) \delta\left(\frac{\mathcal{H}\{s(t)\}}{E_s(t)}\right) \right] dt. \quad (16)$$

Using the product rule of differentiation, the definition eq. (8), and the fact that the Hilbert transform is anti-self adjoint and commutes with the derivative operator (see also Yuan *et al.* 2015), we rewrite the derivative of χ_{EP} in eq. (16) as

$$\begin{aligned} \delta\chi_{\text{EP}} = & \sum_{s,r} \int_0^T \left[\Delta I(t) \frac{s(t) \mathcal{H}\{s(t)\}}{E_s^3(t)} - \Delta R(t) \frac{[\mathcal{H}\{s(t)\}]^2}{E_s^3(t)} \right] \delta s(t) dt \\ & + \sum_{s,r} \int_0^T \left[\mathcal{H} \left\{ \Delta I(t) \frac{s^2(t)}{E_s^3(t)} - \Delta R(t) \frac{s(t) \mathcal{H}\{s(t)\}}{E_s^3(t)} \right\} \right] \delta s(t) dt. \end{aligned} \quad (17)$$

To numerically evaluate the gradient for an event located at \mathbf{x}_s in eq. (14), the adjoint sources to be back-propagated can be expressed as

$$f_{\text{EP}}^\dagger(\mathbf{x}, t) = \sum_r \left[\Delta I(t) \frac{s(t) \mathcal{H}\{s(t)\}}{E_s^3(t)} - \Delta R(t) \frac{[\mathcal{H}\{s(t)\}]^2}{E_s^3(t)} + \mathcal{H} \left\{ \Delta I(t) \frac{s^2(t)}{E_s^3(t)} - \Delta R(t) \frac{s(t) \mathcal{H}\{s(t)\}}{E_s^3(t)} \right\} \right] \delta(\mathbf{x} - \mathbf{x}_r). \quad (18)$$

Similarly to eq. (11), we add a regularization term ϵ to the denominator to avoid numerical instabilities. In addition to being used as absolute measurements in eq. (14), the exponentiated phase can be used in a differential manner for pairs of signals, adopting the approach of *double-difference* adjoint seismic tomography introduced by Yuan *et al.* (2016). For additional information, we recommend reading Appendix A2.

4.2 Pros and cons of the exponentiated phase

Before applying the exponentiated-phase measurements on a real dataset, we illustrate its advantages and disadvantages through the analysis of simple synthetic signals.

The first example, shown in Fig. 4, uses a Gaussian signal, d_1 (blue), and a noisy trace, d_2 (orange) by adding 1 per cent random noise (green) generated from a uniform distribution $[-0.01, 0.01]$. The example is designed to test the robustness of IP and EP to the presence of noise. For these datasets, we calculate IP in Fig. 4(b): IP of d_1 does not have phase jumps and is used as a reference, but we experience multiple phase wraps while measuring IP of d_2 . The unwrapping procedure removes the discontinuities by scanning and shifting phases by $\pm\pi$, but it also results in non-local effects to the entire phase measurements (green). In contrast, EP does not show this limitation: Fig. 4(c)–(d) calculate consistent real and imaginary components of EP, with and without noise. Through this simple example, we further demonstrate the challenges of measuring IP robustly: the presence of only 1 per cent noise (maximum noise level) leads to significant changes in IP, preventing its straightforward application to real problems. On the other hand, the robustness of EP makes it practical in handling currently available big datasets characterized by noisy and complex wave trains.

A comprehensive analysis of the relationships between IP and EP can also be found in Appendix A3 and we summarize here: EP is similar to unwrapped IP because both measure continuous phase, and it is also similar to wrapped phase as neither distinguishes the number of ‘cycles’ while characterizing phase. That being said, EP may not preserve the good convexity of the unwrapped IP misfit function. To show this, we offer another example in Fig. 5 to analyse the shape of the misfit function of IP and EP with respect to time-shift. In Fig. 5(a) we use a simple Ricker wavelet (in black, wavelet temporal resolution $T_R = 0.45$ s) and gradually shift it beyond the wavelet resolution (examples in colours); in Fig. 5(b)–(d) we calculate the IP and EP (real and imaginary part, respectively) phase change due to the time-shift, which provide the curves of their misfit functions on the right panel. A much smaller time-shift interval is used to get smooth misfit curves and the sampled examples on the left panel are marked by circles. To have all misfit functions at a comparable level for clear comparison, we normalize their misfit curves by their maximum values in Fig. 5(e). As clearly demonstrated, EP does not show the good convexity of CC and IP after the time-shift exceeds one period. It is worth mentioning that at this point, IP starts to have phase jumps, but here we only show the best case where the phase is carefully unwrapped and no noise is present. The misfit curves before normalization at small time-shift in Fig. 5(f) tell us that IP and EP are identical when the time-shift is within a quarter period (the detailed mathematical derivations can be found in Appendix A3). Through these two examples, we learn that EP has both pros and cons compared to IP: although EP is able to handle the discontinuities while measuring phase, the misfit function shows non-convexity at large phase difference. The awareness of its limitation motivates us to combine EP with CC and MT to strike a good balance between model recovery and convergence when the initial model is not adequate. However, as suggested in Fig. 5(f), the misfit functions could be at different orders of magnitude, which requires proper weighting to ensure each of them contribute equally in the hybridization.

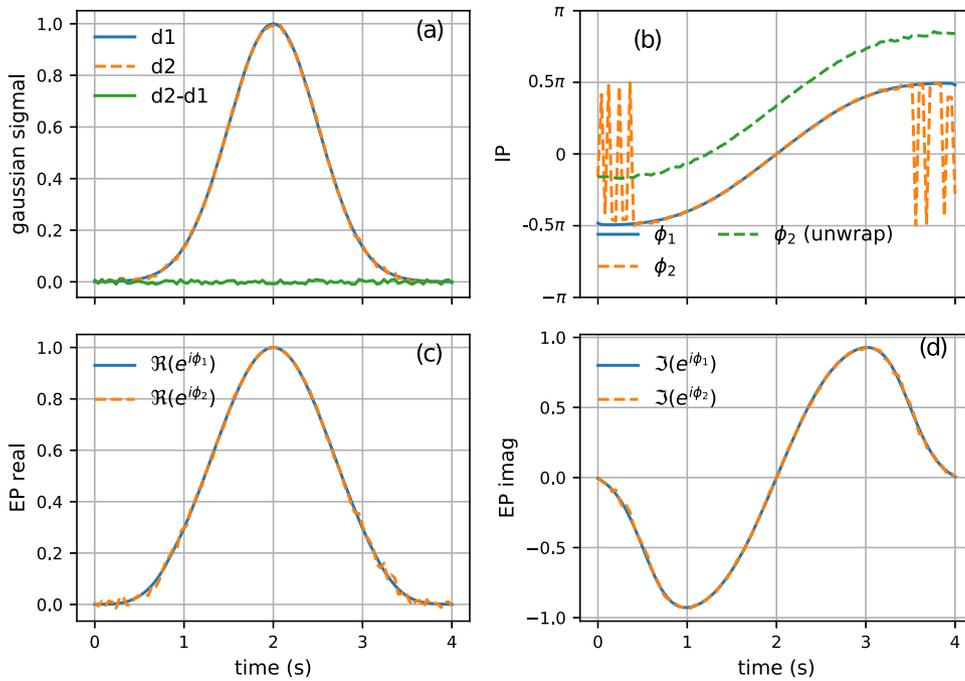


Figure 4. A synthetic test to compare the robustness to noise of instantaneous-phase (IP) and exponentiated-phase (EP) measurements, showing that IP can be very sensitive to noise, resulting in non-local effects to the entire phase measurement. (a) A simple Gaussian signal, d_1 , without noise and d_2 after the addition of 1 per cent random noise in green. (b) The instantaneous phase, ϕ_1 , for the noise-free data d_1 , and ϕ_2 , the instantaneous phase for the noisy record d_2 , which exhibits discontinuities in phase. The unwrapped phase (in green), removes the discontinuities, but is nevertheless significantly shifted from the noise-free phase ϕ_1 . The real (c) and imaginary (d) components of the complex EP for d_1 and d_2 show that EP is more robust to noise.

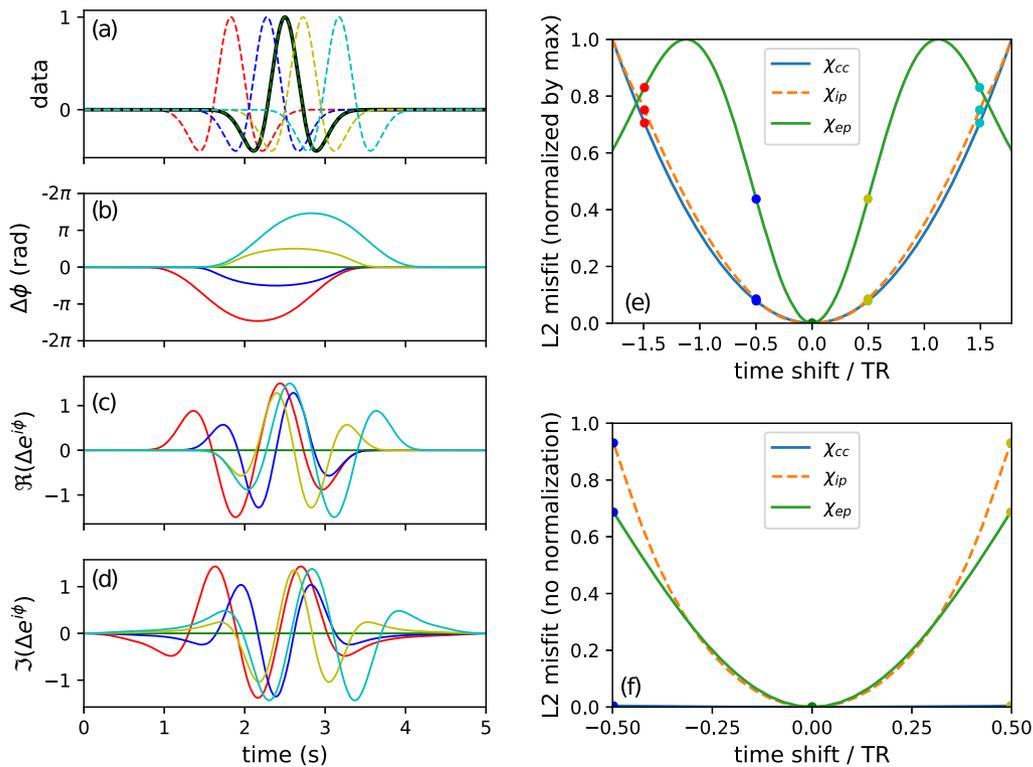


Figure 5. A synthetic test of Ricker signals with different time-shifts (left-hand panel) to compare the convergence behaviour of CC, IP and EP (right-hand panel). (a) An ‘observed’ trace in black and the time-shifted ‘synthetics’ in colours. (b) Their instantaneous phase difference measurements after the unwrapping process. (c) and (d) Their exponentiated phase difference measurements in real and imaginary parts. (e) Normalized misfit function curves of CC, IP (noise-free, unwrapped) and EP with time-shifts beyond one period demonstrate EP does not have the good convexity of IP and CC at large time/phase differences. (f) The unnormalized misfit function curves show that EP and IP are identical when time-shifts are within a quarter period.

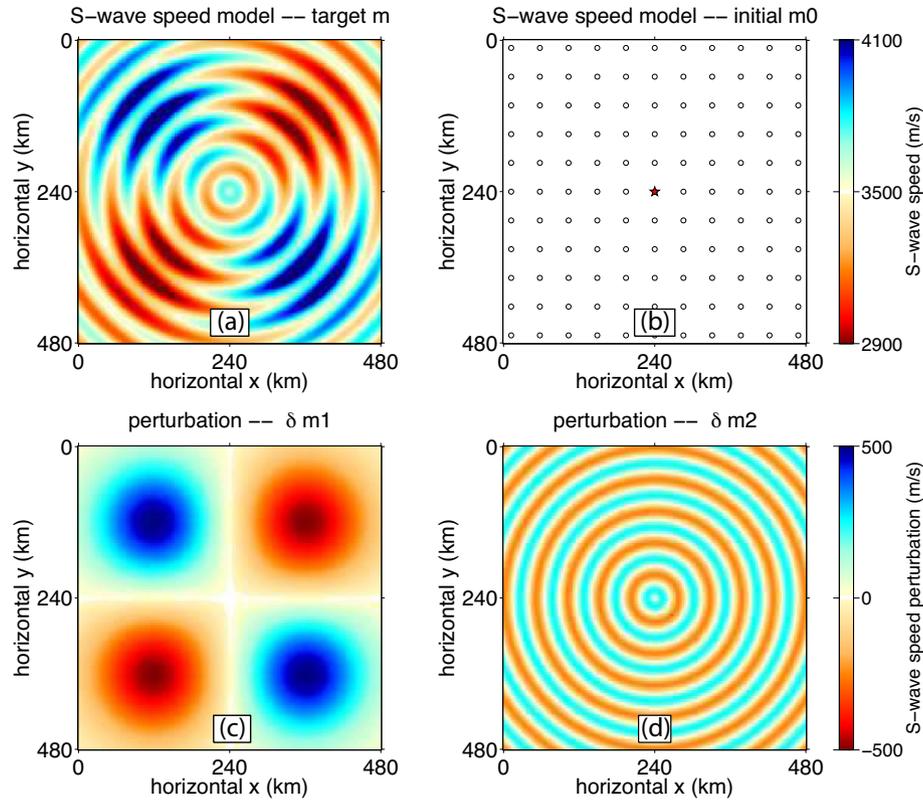


Figure 6. Wavespeed and acquisition geometry for our synthetic experiments. (a) The target model. (b) A homogeneous starting model, where we denote the source by a star and receivers by open circles. (c) Long-wavelength and (d) short-wavelength features which sum together to the target model in (a).

5 NUMERICAL EXPERIMENTS

We conducted numerical experiments using a 2-D synthetic model. We solve the wave equation at finite frequencies in a zero-thickness surface model, using the ‘membrane wave’ as an analog to seismic surface waves (Peter *et al.* 2007). We created a target model of shear wavespeeds V_s as shown in Fig. 6(a) by adding the long-wavelength velocity variations of Fig. 6(c) and the short-wavelength perturbations of 6(d) to the homogeneous initial model in Fig. 6(b). The source and receiver configuration used in this synthetic example is shown in Fig. 6(b), where open circles represent receiver arrays, equidistant at 45 km spacing, and the central red star denotes a single seismic source.

5.1 An experiment with a good initial model

In the first experiment, we started with a reasonably good initial model: a homogeneous layer with $V_s = 3500 \text{ m s}^{-1}$, as in Fig. 6(b). We used the target model in Fig. 6(a) to numerically generate synthetic observations, and the homogeneous model to simulate initial-model synthetics. We computed their cross-correlation (CC), multitaper (MT) and exponentiated (EP) time and phase differences, as well as their corresponding adjoint sources at each receiver location. We iteratively computed the event kernels to form the gradients by back-propagating these adjoint sources from all receivers, and updated the velocity model until it converged, or until the saturation of the misfit function reduction. The inversion results with CC (Fig. 7a) and MT (Fig. 7b) recover the long-wavelength structure, where the direct waves are dominant, but fail to see any of the small-scale information. In contrast, the inversion with EP (Fig. 7c) reveals significant fine structure in the true model, picking up both the main phases and the small-magnitude scattered waves generated by the small-scale model perturbations. In this noiseless synthetic experiment, we used a water level of 0.1 per cent of the maximum amplitude in the measured data, but readers are encouraged to look up the additional inversion examples shown in Fig. A4 in Appendix A3, where we further discuss the choice of water level.

5.2 An experiment with a poor initial model

In the second experiment, we start with a relatively poor initial model: a homogeneous layer with shear wavespeed of 3000 m s^{-1} . Similarly to the previous experiment, we used the same termination criteria and performed iterative inversions using CC, MT and EP individually. In this experiment, both CC (Fig. 8a) and MT (Fig. 8b) did converge to recover the long-wavelength structure. However, EP (Fig. 8c) failed while trying to recover the small-scale heterogeneities, likely because of the non-convexity of the EP misfit function discussed previously. In

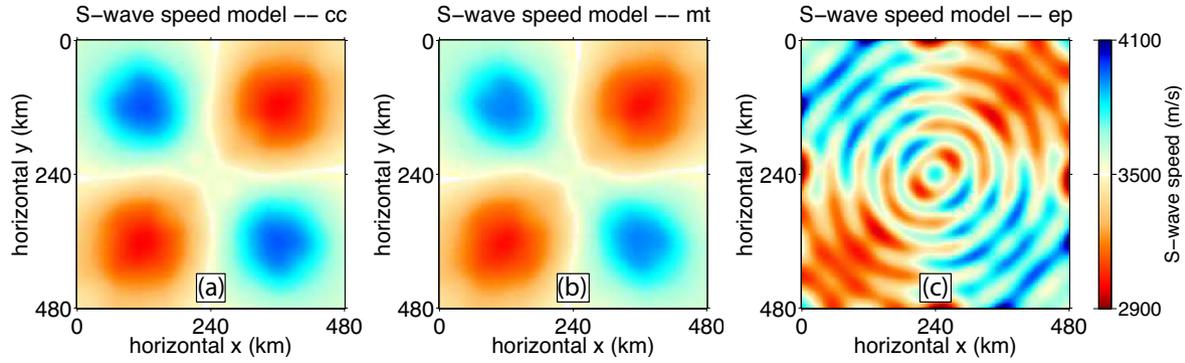


Figure 7. Inversion results using three different types of measurements starting from a relatively good homogeneous model with $V_S = 3500 \text{ m s}^{-1}$, via (a) cross-correlation (CC), (b) multitaper traveltimes (MT) and (c) exponentiated phase (EP) measurements.

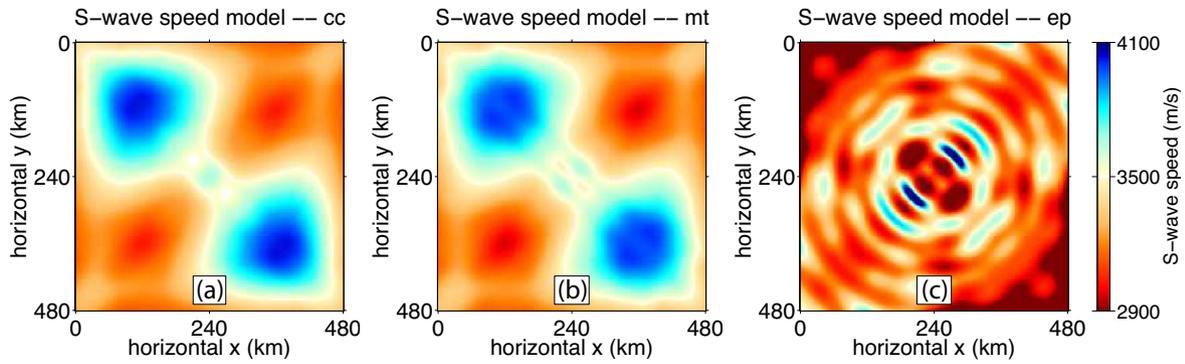


Figure 8. Inversion results as in Fig. 7, but starting from a relatively poor homogeneous model with $V_S = 3000 \text{ m s}^{-1}$, via (a) cross-correlation (CC), (b) multitaper traveltimes (MT) and (c) exponentiated phase (EP) measurements. Despite the poor initial model, CC and MT still successfully resolve the long-wavelength structure in Fig. 6(c). However, the EP approach fails to converge due to the non-convexity of its misfit function for large phase differences.

the following section, we will use this experiment to explore the hybridization approach by taking advantage of CC and MT to exploit their merits of good convergence, and of EP for its capability to recover fine structure.

5.3 Objective-function hybridization

The simultaneous hybridization approach allows us to combine each individual type of measurement into an overall misfit function whose gradient can be computed at once, without the cost of evaluating all of the individual gradients separately. We rewrite the generic expression of eq. (1) to hybridize the three specific CC, MT and EP measurement classes,

$$\chi_H(\mathbf{m}) = w_{CC} \frac{\chi_{CC}(\mathbf{m})}{\sigma_{CC}^2} + w_{MT} \frac{\chi_{MT}(\mathbf{m})}{\sigma_{MT}^2} + w_{EP} \frac{\chi_{EP}(\mathbf{m})}{\sigma_{EP}^2}, \quad (19)$$

where σ_{CC} , σ_{MT} and σ_{EP} are measurement uncertainties, to be estimated from the predicted and observed waveforms; and w_{CC} , w_{MT} and w_{EP} are additional weighting constants that serve to balance the relative importance of each measurement type in the hybridization.

The CC measurement returns the time-shift by which the observed and modeled waveforms are most alike, but even shifted by Δt and scaled by $\Delta \ln A$, the synthetic waveforms will not perfectly match the observations. The remaining waveform residual provides an estimate of the CC measurement error. We use the approach advocated by Tape *et al.* (2010), determining our σ_T^2 from their eqs (A6) and (A7), adopting $\sigma_{CC} = \sigma_T$. The MT method provides multiple approximately uncorrelated measurements, and we estimate the measurement uncertainty σ_{MT} via the jackknife approach (Efron & Stein 1981; Thomson & Chave 1991). As to EP, since we have been arguing that it provides a nearly complete description of all portions of the signal, we maintained $\sigma_{EP} = 1$ throughout the experiment.

Since MT does not provide any additional structural information compared to CC in this experiment, we omitted MT measurements in the hybridization, $w_{MT} = 0$. As suggested by Fig. 5(f), the misfit values of CC and EP could be orders of magnitude different. Proper weighting ensures the inversion does not focus on one misfit function at the expense of the other. Here we are able to perform a series of inversions to find the optimal weight. We conducted an iterative binary search for w_{CC} between 0 and 1 ($w_{CC} + w_{EP} = 1$). The gallery of results using the hybrid misfit function is shown in Fig. 9, clearly demonstrating the effect of weighting on the inversions. Visually, $w_{CC} = 0.98$ gives the best result. Plotting data and model misfits in Fig. 10 confirms that $w_{CC} = 0.98$ is best. With the standard deviations of CC and EP measurements very different, their misfit values are not of the same order, which likely results in the large weight in this experiment. For the first-iteration misfits $\chi_{CC}(m_0)$ and $\chi_{EP}(m_0)$, the weighted misfits for CC and EP are $0.98 \times 1.2 = 1.18$ and $0.02 \times 56.58 = 1.13$, respectively. This suggests

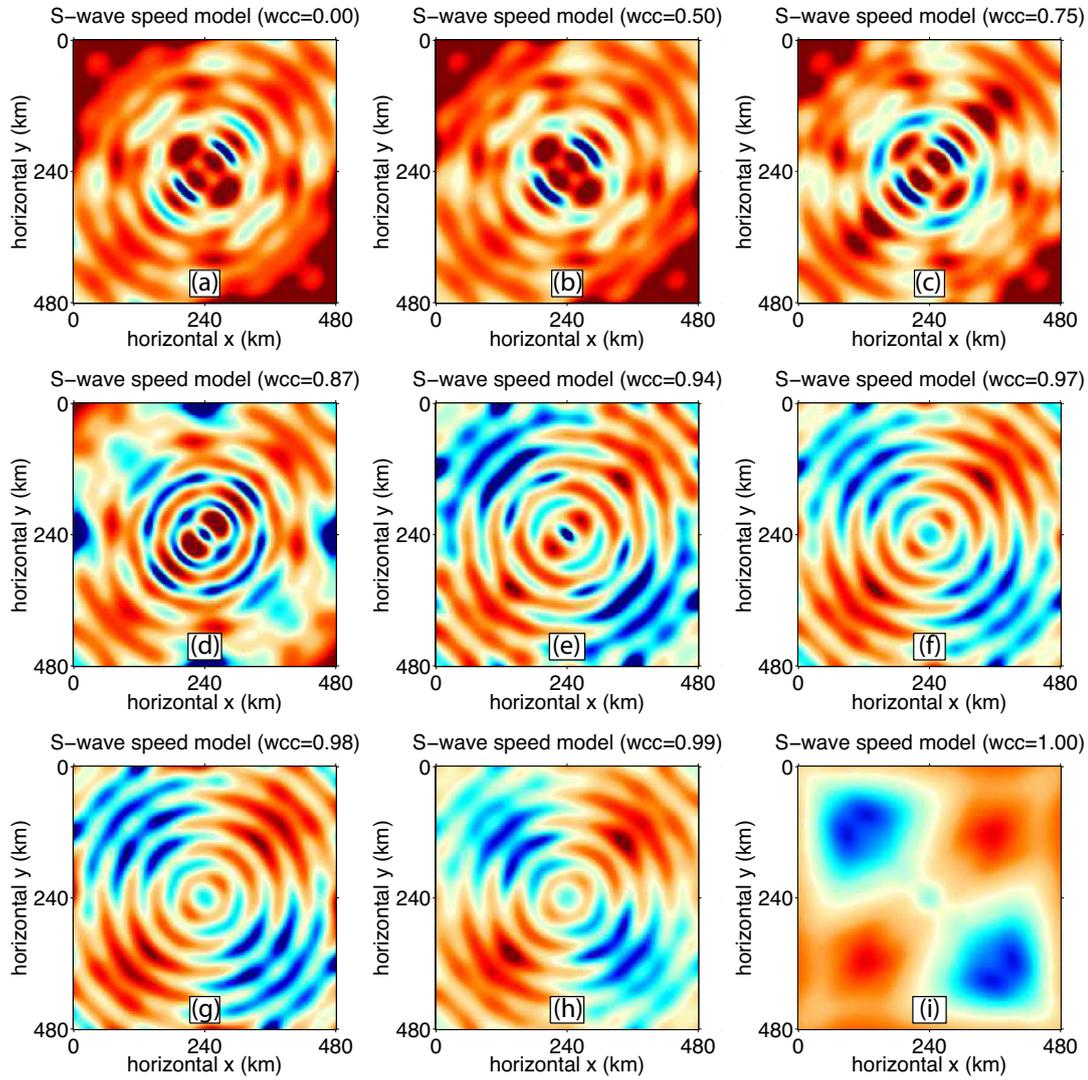


Figure 9. Inversion experiments with varying w_{CC} ($w_{EP} = 1 - w_{CC}$) by hybridizing CC and EP. Each plot is the final inverted velocity model. Using $w_{CC} = 0.98$ obtains the result that is most faithful to the known truth in this synthetic experiment.

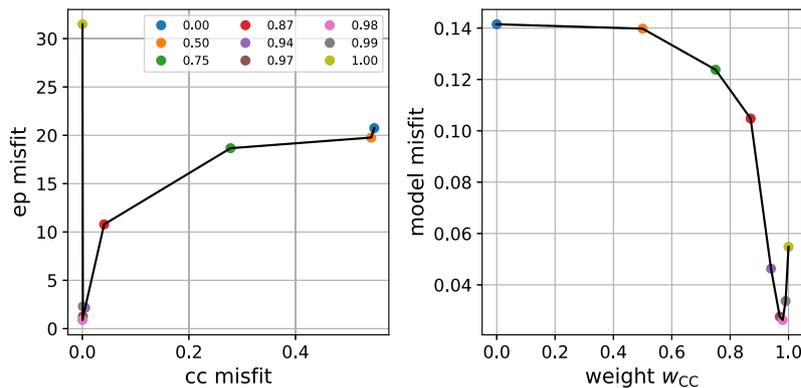


Figure 10. Data and model misfit evolution for the experiments in Fig. 9. The graph of CC versus EP data misfit is shown on the left, and the model misfit against the weights used in the inversion is shown on the right for the final models, illustrating the optimality of the weight $w_{CC} = 0.98$ for the misfit hybridization.

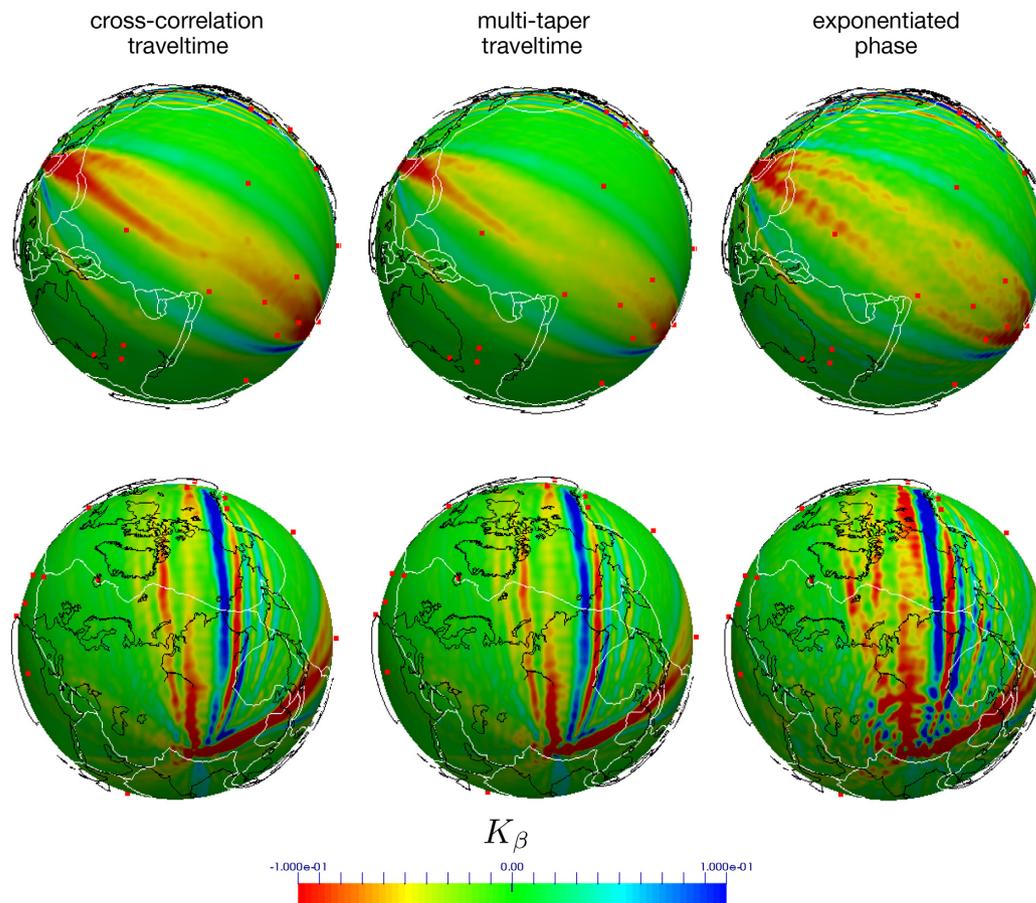


Figure 11. Map views taken at 120 km depth of 3-D global shear wavespeed event kernels K_β for CC, MT and EP measurements presented from two different viewing angles. The kernels are computed using data from the August 25, 2008 Xizang earthquake ($M_w = 6.7$, depth 17 km) filtered within the 45–110 s period band. The event kernels were smoothed and normalized.

that we get the best results when CC and EP contribute similarly to the total misfit. A real experiment could consider the scale-length of heterogeneities to be resolved while determining the measurement weights. With a poor initial model, hybridization with weights properly balancing each measurement is more advantageous than using either objective function alone.

6 DISCUSSION AND CONCLUSIONS

Seismic tomography relies on being able to extract a maximum of information from every seismogram while managing data quality and measurement challenges. In this work, we discussed various objective functions, and their combination, to benefit full-waveform tomography. We focused on seismic measurements of traveltimes and phase. We reviewed and numerically compared several phase-based measurements including cross-correlation (CC), multitaper (MT) and instantaneous phase (IP). Evaluating their scope of application and exploring their limitations, we showed the importance and challenges of using time and phase information from the entire seismogram, and the necessity to combine different measurements through objective-function hybridization.

We introduced a new type of phase measurement, the exponentiated phase (EP), an implicit device to calibrate phase information in the time domain for which the selection of windows targeting each phase, required for the CC and MT approaches, is no longer necessary. We compared the advantages and disadvantages of EP compared to IP in terms of measurement stability and inversion convergence. The enhanced measurement stability of EP will facilitate progress towards using all of the main attributes of the available seismograms in real-world applications, and an awareness of its limitation motivated our notions of objective-function hybridization. We propose to combine EP measurements with classical CC approaches to capture short-wavelength body waves, and with MT for dispersive waves.

Our synthetic experiments demonstrate the advantages and limitations of each measurement type. EP is able to reveal detailed structural information but is less suited for characterizing large phase differences between observations and synthetics, while CC and MT are nicely complemented by EP in robustly characterizing short-wavelength features from seismic data.

To gain better insight into how EP behaves when applied to real data, we computed global event kernels by CC, MT and EP measurements using teleseismic seismograms recorded by the Global Seismographic Network. In Fig. 11 we present sample horizontal cross-sections of

shear wavespeed global event kernels at 120 km depth computed for CC, MT and EP misfits. We used data from the 25 August 2008 Xizang earthquake ($M_w = 6.7$, depth=17 km, CMT 200808251322A) recorded by 200 stations distributed worldwide. Synthetics were calculated using the 3-D global wave propagation solver SPECSEM3D_GLOBE (Komatitsch & Tromp 2002a, b) using the 3-D global model GLAD-M15 (Bozdağ *et al.* 2016). We filtered both observed and synthetic data within the period range of 45–110 s and used FLEXWIN (Maggi *et al.* 2009) to pick measurement windows. We adjusted FLEXWIN to favour large windows in order to work on wave trains where CC and MT measurements may recover larger-amplitude signals. We used a water level of 1 per cent of the maximum amplitude over the EP measurement windows, and the kernels were smoothed by a Gaussian operator with a halfwidth of 80 and 25 km in the horizontal and vertical directions, respectively. We observe that the EP measurements behave robustly in the 3-D global setting. The EP kernel represents structural variations consistently with those of CC and MT, but it reveals richer structural information off the dominant path, where the EP measurements better emphasize the effect of scattered waves.

Future work can use CC and MT to measure readily identifiable body- and surface-wave arrivals. Once the major phases are in good agreement EP measurements made on larger time windows then can be included to take the small-magnitude scattered waves into account. Alternatively, simultaneous objective-function hybridization is promising to improve convergence through joint gradients, whose calculation does not affect the computational cost of numerical simulations, given that multiple measurements are combined at the stage of the construction of adjoint sources.

Our discussions have been limited to waveform metrics that primarily capture elastic properties of the subsurface. We look forward to the further extension of our methods to allow for the measurement and hybridization of the counterparts of these phase measurements, including cross-correlation amplitude, multitaper amplitude and waveform envelopes, and to their use for a second-stage inversion for the elastic and anelastic properties of the subsurface.

ACKNOWLEDGEMENTS

This paper grew out of early discussions with Jeannot Trampert on combining different phase measurements in adjoint tomography. We thank Ryan Modrak and Jeroen Tromp for fruitful discussions. We also thank Ridvan Orsvuran for computing the global kernels. The open-source spectral-element software package SPECSEM and the automated window-selection package FLEXWIN are freely available from the Computational Infrastructure for Geodynamics (www.geodynamics.org). We gratefully acknowledge the computational resources provided by the Princeton Institute for Computational Science & Engineering (PICSciE). CC thanks the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2018/04918-6) for providing the financial support for his PhD studies, and FJS thanks the U.S. National Science Foundation (EAR 1736046), the King Abdullah University of Science and Technology (OSR-2016-CRG5-2970-01) and TOTAL EP Research and Technology for financial support, as well as the Institute for Advanced Study in Princeton for a quiet yet stimulating work environment during 2018–2019. The authors thank the Global Seismographic Network (GSN) for making data available. We would like to thank the editor, Dr Martin Schimmel, and two anonymous reviewers for their careful reading and constructive suggestions, which have helped improve the manuscript. Our computer codes are available from <https://github.com/yanhuay/seisEP>.

REFERENCES

- Akçelik, V., Biros, G. & Ghattas, O., 2002. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, in *Proc. Conf. Supercomputing*, ACM/IEEE, doi:10.1109/SC.2002.10002.
- Borisov, D., Modrak, R., Rusmanugroho, H., Yuan, Y.O., Gao, F., Simons, F.J. & Tromp, J., 2016. Spectral-element based 3D elastic full waveform inversion of surface waves in the presence of complex topography using envelope-based misfit function, *SEG Tech. Prog. Expanded Abstracts*, pp. 1211–1215, doi:10.1190/segam2016-13843759.1.
- Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.*, **185**(2), 845–870, doi:10.1111/j.1365-246X.2011.04970.x.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.*, **207**(3), 1739–1766, doi:10.1093/gji/ggw356.
- Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multi-scale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473, doi:10.1190/1.1443880.
- Burdick, S., *et al.*, 2010. Model update January 2010: Upper mantle heterogeneity beneath North America from traveltime tomography with global and USArray Transportable Array data, *Seismol. Res. Lett.*, **81**(5), 689–693, doi:10.1785/gssrl.81.5.689.
- Chen, Y., Hill, J., Lei, W., Lefebvre, M., Tromp, J., Bozdağ, E. & Komatitsch, D., 2017. Automated time-window selection based on machine learning for full-waveform inversion, in *SEG Tech. Prog. Expanded Abstracts*, pp. 1604–1609, doi:10.1190/segam2017-17734162.1.
- Dahlen, F.A., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes — I. Theory, *Geophys. J. Int.*, **141**(1), 157–174, doi:10.1046/j.1365-246X.2000.00070.x.
- Dalton, C.A. & Ekström, G., 2006. Global models of surface wave attenuation, *J. geophys. Res.*, **111**(B5), B05317, doi:10.1029/2005JB003997.
- Djebbi, R. & Alkhalifah, T., 2014. Traveltime sensitivity kernels for wave equation tomography using the unwrapped phase, *Geophys. J. Int.*, **197**(2), 975–986, doi:10.1093/gji/ggu025.
- Efron, B. & Stein, C., 1981. The jackknife estimate of variance, *Ann. Stat.*, **9**(3), 586–596, doi:10.1214/aos/1176345462.
- 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**(2), 665–685, doi:10.1111/j.1365-246X.2008.03923.x.
- French, S.W. & Romanowicz, B., 2015. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, **525**, 95–99, doi:10.1038/nature14876.
- Gelchinsky, B.T. & Shtivelman, V., 1983. Automatic picking of first arrivals and parameterization of traveltime curves, *Geophys. Prospect.*, **31**(6), 915–928, doi:10.1111/j.1365-2478.1983.tb01097.x.
- Haimes, Y.Y. & Li, D., 1988. Hierarchical multiobjective analysis for large-scale systems: review and current status, *Automatica*, **24**(1), 53–69, doi:10.1016/0005-1098(88)90007-6.

- Hjörleifsdóttir, V., 2007. Earthquake source characterization using 3D numerical modeling, *PhD thesis*, California Institute of Technology, Pasadena, CA.
- Hung, S.-H., Dahlen, F.A. & Nolet, G., 2001. Wavefront healing: a banana-doughnut perspective, *Geophys. J. Int.*, **146**, 289–312, doi:10.1046/j.1365-246x.2001.01466.x.
- Karaoğlu, H. & Romanowicz, B., 2018. Inferring global upper-mantle shear attenuation structure by waveform tomography using the spectral element method, *Geophys. J. Int.*, **213**(3), 1536–1558, doi:10.1093/gji/ggy030.
- Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation — I. Validation, *Geophys. J. Int.*, **149**(2), 390–412, doi:10.1046/j.1365-246X.2002.01653.x.
- Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation — II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophys. J. Int.*, **150**(1), 303–318, doi:10.1046/j.1365-246X.2002.01716.x.
- Kristeková, M., Kristek, J., Moczo, P. & Day, S.M., 2006. Misfit criteria for quantitative comparison of seismograms, *Bull. seism. Soc. Am.*, **96**(5), 1836–1850, doi:10.1785/0120060012.
- Laske, G. & Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data, *J. geophys. Res.*, **101**(B7), 16059–16075, doi:10.1029/96JB00526.
- Liu, Q. & Tromp, J., 2006. Finite-frequency sensitivity kernels based upon adjoint methods, *Bull. seism. Soc. Am.*, **96**(6), 2383–2397, doi:10.1785/0120060041.
- Luo, J., Wu, R.-S. & Gao, F., 2018. Time-domain full waveform inversion using instantaneous phase information with damping, *J. Geophys. Eng.*, **15**(3), 1032–1041, doi:10.1088/1742-2140/aaa984.
- Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics*, **56**(5), 654–663, doi:10.1190/1.1443081.
- Maggi, A., Tape, C., Chen, M., Chao, D. & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, **178**(1), 257–281, doi:10.1111/j.1365-246X.2009.04099.x.
- Maraschini, M., Ernst, F., Foti, S. & Socco, L.V., 2010. A new misfit function for multimodal inversion of surface waves, *Geophysics*, **75**(4), G31–G43, doi:10.1190/1.3436539.
- Marquering, H., Dahlen, F.A. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency travel times: the banana-doughnut paradox, *Geophys. J. Int.*, **137**(3), 805–815, doi:10.1046/j.1365-246x.1999.00837.x.
- Modrak, R. & Tromp, J., 2016. Seismic waveform inversion best practices: regional, global and exploration test cases, *Geophys. J. Int.*, **206**(3), 1864–1889, doi:10.1093/gji/ggw202.
- Nolet, G., 2008. *A Breviary for Seismic Tomography*, Cambridge Univ. Press, Cambridge, UK.
- Nolet, G. & Dahlen, F.A., 2000. Wave front healing and the evolution of seismic delay times, *J. geophys. Res.*, **105**(B8), 19043–19054, doi:10.1029/2000JB900161.
- Nolet, G. et al., 2019. Imaging the Galápagos mantle plume with an unconventional application of floating seismometers, *Sci. Rep.*, **9**, 1326, doi:10.1038/s41598-018-36835-w.
- Operto, S., Ravaut, C., Imbrota, L., Virieux, J., Herrero, A. & Aversana, P.D., 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study, *Geophys. Prospect*, **52**(6), 625–651, doi:10.1111/j.1365-2478.2004.00452.x.
- Oppenheim, A.V. & Lim, J.S., 1981. The importance of phase in signals, *Proc. IEEE*, **69**(5), 529–541, doi:10.1109/PROC.1981.12022.
- Peter, D., Tape, C., Boschi, L. & Woodhouse, J.H., 2007. Surface wave tomography: global membrane waves and adjoint methods, *Geophys. J. Int.*, **171**(3), 1098–1117, doi:10.1111/j.1365-246X.2007.03554.x.
- Phillips, W.S. & Fehler, M.C., 1991. Traveltime tomography: a comparison of popular methods, *Geophysics*, **56**(10), 1639–1649, doi:10.1190/1.1442974.
- Prieux, V., Brossier, R., Operto, S. & Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: imaging compressional wave speed, density and attenuation, *Geophys. J. Int.*, **194**(3), 1640–1664, doi:10.1093/gji/ggt177.
- Prieux, V., Brossier, R., Operto, S. & Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 2: imaging compressive-wave and shear-wave velocities, *Geophys. J. Int.*, **194**(3), 1665–1681, doi:10.1093/gji/ggt178.
- Pritt, M.D. & Ghiglia, D.C., 1998. *Two-Dimensional Phase Unwrapping. Theory, Algorithms, and Software*, John Wiley, New York, NY.
- Rawlinson, N. & Sambridge, M., 2003. Seismic traveltime tomography of the crust and lithosphere, *Adv. Geophys.*, **46**, 81–198, doi:10.1016/S0065-2687(03)46002-0.
- Rickers, F., Fichtner, A. & Trampert, J., 2012. Imaging mantle plumes with instantaneous phase measurements of diffracted waves, *Geophys. J. Int.*, **190**(1), 650–664, doi:10.1111/j.1365-246X.2012.05515.x.
- Roy, B., 1971. Problems and methods with multiple objective functions, *Math. Prog.*, **1**, 239–266, doi:10.1007/BF01584088.
- Sabri, E.H. & Beamon, B.M., 2000. A multi-objective approach to simultaneous strategic and operational planning in supply chain design, *Omega*, **28**(5), 581–598, doi:10.1016/S0305-0483(99)00080-8.
- Selby, N.D. & Woodhouse, J.H., 2000. Controls on Rayleigh wave amplitudes: attenuation and focusing, *Geophys. J. Int.*, **142**(3), 933–940, doi:10.1046/j.1365-246x.2000.00209.x.
- Sguazzero, P., Vesnaver, A. & Kindelan, M., 1987. Stacking velocity analysis on complex traces, in *SEG Tech. Prog. Expanded Abstracts*, pp. 420–422, doi:10.1190/1.1891980.
- Sheng, J., Leeds, A., Buddensiek, M. & Schuster, G.T., 2006. Early arrival waveform tomography on near-surface refraction data, *Geophysics*, **71**(1), U47–U57, doi:10.1190/1.2210969.
- Shin, C. et al., 2003. Traveltime calculations from frequency-domain downward-continuation algorithms, *Geophysics*, **68**(4), 1380–1388, doi:10.1190/1.1598131.
- Simons, F.J., 2010. Slepian functions and their use in signal estimation and spectral analysis, in *Handbook of Geomathematics*, Chap. 30, pp. 891–923, doi:10.1007/978-3-642-01546-5_30, eds Freeden, W., Nashed, M.Z. & Sonar, T., Springer, Heidelberg, Germany.
- Slepian, D., 1978. Prolate spheroidal wave functions, Fourier analysis and uncertainty — V: the discrete case, *Bell Syst. Tech. J.*, **57**, 1371–1429, doi:10.1002/j.1538-7305.1978.tb02104.x.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the Southern California crust, *Science*, **325**, 988–992, doi:10.1126/science.1175298.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.*, **180**(1), 433–462, doi:10.1111/j.1365-246X.2009.04429.x.
- Tape, C.A., 2009. Seismic tomography of southern California using adjoint methods, *PhD thesis*, Calif. Inst. Technol., Pasadena, CA.
- Thomson, D.J., 1982. Spectrum estimation and harmonic analysis, *Proc. IEEE*, **70**(9), 1055–1096, doi:10.1109/PROC.1982.12433.
- Thomson, D.J. & Chave, A.D., 1991. Jackknifed error estimates for spectra, coherences, and transfer functions, in *Advances in Spectrum Analysis and Array Processing*, Vol. 1, Chap. 2, pp. 58–113, ed. Haykin, S., Prentice-Hall, Englewood Cliffs, NJ.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216, doi:10.1111/j.1365-246X.2004.02453.x.
- Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26, doi:10.1190/1.3238367.
- Yuan, Y.O. & Simons, F.J., 2014. Multiscale adjoint waveform-difference tomography using wavelets, *Geophysics*, **79**(3), WA79–WA95, doi:10.1190/GEO2013-0383.1.
- Yuan, Y.O., Simons, F.J. & Bozdağ, E., 2015. Multiscale adjoint tomography for surface and body waves, *Geophysics*, **80**(5), R281–R302, doi:10.1190/GEO2014-0461.1.
- Yuan, Y.O., Simons, F.J. & Tromp, J., 2016. Double-difference adjoint seismic tomography, *Geophys. J. Int.*, **206**(3), 1599–1618, doi:10.1093/gji/ggw233.

Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.*, **158**(1), 142–168, doi:10.1111/j.1365-246X.2004.02324.x.

Zhou, Y., Dahlen, F.A. & Nolet, G., 2005. Finite-frequency effects in global surface-wave tomography, *Geophys. J. Int.*, **163**(3), 1087–1111, doi:10.1111/j.1365-246X.2005.02780.x.

Zhu, H., Bozdağ, E., Duffy, T.S. & Tromp, J., 2013. Seismic attenuation beneath Europe and the North Atlantic: Implications for water in the mantle, *Earth planet. Sci. Lett.*, **381**, 1–11, doi:10.1016/j.epsl.2013.08.030.

APPENDIX: SUPPLEMENTARY MATERIAL

In this Appendix we provide additional detail on the material presented in this paper. Appendix A1 is modelled after the material presented by Yuan *et al.* (2016), which in turn borrows heavily from work presented elsewhere (notably Laske & Masters 1996; Hjörleifsdóttir 2007), rewritten here for completeness without any claim to originality. Appendix A2 provides the double-difference versions of the exponentiated-phase equations, and Appendix A3 illustrates the equivalence of a modified instantaneous-phase measurement to that of the complex exponentiated phase.

A1 Multitaper phase measurements for adjoint tomography

With the multitaper technique the frequency-dependent traveltime anomaly $\Delta t(\omega)$ and amplitude anomaly $\Delta \ln A(\omega)$ can be linked to a complex frequency-domain ‘transfer function’ that maps the Fourier-domain synthetic traces $s(\omega)$ to the observations to $d(\omega)$ by modulating traveltime and amplitude using the measurements Δt and $\Delta \ln A$ in the following manner:

$$d(\omega) = \mathcal{T}(\omega)s(\omega) = e^{\Delta \ln A(\omega) - i\omega \Delta t(\omega)} s(\omega). \quad (\text{A1})$$

Here, we only consider the frequency-dependent traveltime measurements $\Delta t(\omega)$. For those interested in using $\Delta \ln A(\omega)$ in the adjoint method, we recommend reading Zhou *et al.* 2004 and Tape 2009. The above linear mapping can be optimally estimated with the multitaper method, the asterisk denoting complex conjugation,

$$T(\omega) = \frac{\sum_k d_k(\omega) s_k^*(\omega)}{\sum_k s_k(\omega) s_k^*(\omega)} = \underset{\mathcal{T}(\omega)}{\operatorname{argmin}} \sum_k [d_k(\omega) - \mathcal{T}(\omega) s_k(\omega)]^2, \quad (\text{A2})$$

where $s_k(\omega)$ and $d_k(\omega)$ are frequency-domain versions of $s_k(t)$ and $d_k(t)$, approximately uncorrelated versions of $s(t)$ and $d(t)$ obtained after individually tapering them by multiplication with an orthonormal set of $k = 1, \dots, \lfloor T\Omega/\pi \rfloor - 1$ prolate spheroidal functions $h_k^{T\Omega}(t)$, designed for a particular window length T and an angular frequency half-bandwidth Ω , in the manner of Thomson (1982).

To optimize the multitaper-traveltime weighted misfit function of eq. (5), we need access to its derivative

$$\delta \chi_{\text{MT}} = \sum_{s,r} \sum_{\omega} W_t(\omega) \Delta t_{s,r}(\omega) \delta \Delta t_{s,r}(\omega), \quad (\text{A3})$$

where $\delta \Delta t_{s,r}(\omega)$ denotes the perturbation of the frequency-dependent traveltime measurement due to the model perturbation, for a particular source s and receiver r . In the linear regime, when the approximation $d_k(\omega) \approx s_k(\omega) + \delta s_k(\omega)$ is valid, we can rewrite eq. (A2) as

$$T(\omega) = \frac{\sum_k [s_k(\omega) + \delta s_k(\omega)] s_k^*(\omega)}{\sum_k s_k(\omega) s_k^*(\omega)} = 1 + \frac{\sum_k \delta s_k(\omega) s_k^*(\omega)}{\sum_k s_k(\omega) s_k^*(\omega)}. \quad (\text{A4})$$

Taking the first-order Taylor expansion of eq. (A1) and combining it with the above eq. (A4), we can link the traveltime perturbation $\delta \Delta t(\omega)$, dropping the subscripts, to the wavefield perturbation $\delta s_k(\omega)$ via

$$\delta \Delta t(\omega) = \operatorname{Re} \left\{ \frac{i \sum_k \delta s_k(\omega) s_k^*(\omega)}{\sum_k s_k(\omega) s_k^*(\omega)} \right\} = \operatorname{Re} \left\{ \sum_k \delta s_k(\omega) p_k(\omega) \right\}, \quad (\text{A5})$$

introducing the frequency-dependent partial traveltime derivative

$$p_k(\omega) = \frac{i}{\omega} \frac{s_k^*(\omega)}{\sum_k s_k(\omega) s_k^*(\omega)}. \quad (\text{A6})$$

Using the above relations (A5)–(A6), we recast the Fréchet derivative defined in eq. (A3) as, again using the subscripts,

$$\delta \chi_{\text{MT}} = \sum_{s,r} \operatorname{Re} \left(\sum_{\omega} W_t(\omega) \Delta t_{s,r}(\omega) \sum_k p_k(\omega) \delta s_k(\omega) \right) \quad (\text{A7})$$

$$= \sum_{s,r} \sum_k \int h_k^{T\Omega}(t) \mathcal{F}^{-1} \left\{ W_t(\omega) \Delta t_{s,r}(\omega) p_k(\omega) \right\} (t) \delta s(t) dt, \quad (\text{A8})$$

after transforming the sum over frequencies to the time-domain using the inverse Fourier operator \mathcal{F}^{-1} . From this follows the final expression for the adjoint source of the frequency-dependent traveltime misfit function for a single event, which was previously quoted as eq. (6).

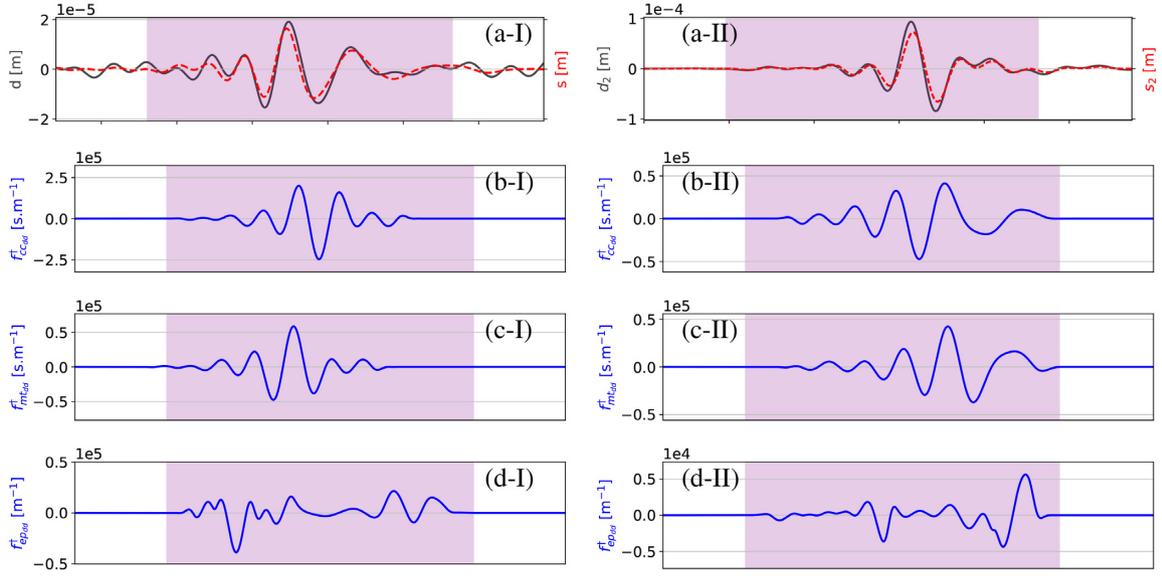


Figure A1. Observed and computed radial-component seismograms at station AFI (a-I), and station ABKT (a-II). The double-difference adjoint sources for station AFI (left-hand panel) and ABKT (right-hand panel) with CC (b), MT (c) and EP (d) measurements.

A2 Double-difference exponentiated phase measurements

The double-difference exponentiated-phase misfit function made on any pair of seismograms i and j can be defined as

$$\chi_{EP}^{dd}(\mathbf{m}) = \frac{1}{2} \sum_{i,j} \int_0^T \left\| [\tilde{s}_a^i(t) - \tilde{s}_a^j(t)] - [\tilde{d}_a^i(t) - \tilde{d}_a^j(t)] \right\|^2 dt = \frac{1}{2} \sum_{i,j} \int_0^T \left[\|\Delta\Delta R_{ij}(t)\|^2 + \|\Delta\Delta I_{ij}(t)\|^2 \right] dt, \quad (\text{A9})$$

$$\Delta\Delta R_{ij}(t) = \left(\frac{s_i(t)}{E_{s_i}(t)} - \frac{s_j(t)}{E_{s_j}(t)} \right) - \left(\frac{d_i(t)}{E_{d_i}(t)} - \frac{d_j(t)}{E_{d_j}(t)} \right), \quad (\text{A10})$$

$$\Delta\Delta I_{ij}(t) = \left(\frac{\mathcal{H}\{s_i(t)\}}{E_{s_i}(t)} - \frac{\mathcal{H}\{s_j(t)\}}{E_{s_j}(t)} \right) - \left(\frac{\mathcal{H}\{d_i(t)\}}{E_{d_i}(t)} - \frac{\mathcal{H}\{d_j(t)\}}{E_{d_j}(t)} \right). \quad (\text{A11})$$

We omit the detailed derivations, and summarize the final adjoint sources for a pair i and j as follows:

$$f_i^\dagger(\mathbf{x}, t) = \sum_j \left[\Delta\Delta R_{ij}(t) \frac{[\mathcal{H}\{s_i(t)\}]^2}{E_{s_i}^3(t)} - \Delta\Delta I_{ij}(t) \frac{s_i(t)\mathcal{H}\{s_i(t)\}}{E_{s_i}^3(t)} + \mathcal{H} \left\{ \Delta\Delta R_{ij}(t) \frac{s_i(t)\mathcal{H}\{s_i(t)\}}{E_{s_i}^3(t)} - \Delta\Delta I_{ij}(t) \frac{s_i^2(t)}{E_{s_i}^3(t)} \right\} \right] \delta(\mathbf{x} - \mathbf{x}_j), \quad (\text{A12})$$

$$f_j^\dagger(\mathbf{x}, t) = \sum_i \left[-\Delta\Delta R_{ij}(t) \frac{[\mathcal{H}\{s_j(t)\}]^2}{E_{s_j}^3(t)} + \Delta\Delta I_{ij}(t) \frac{s_j(t)\mathcal{H}\{s_j(t)\}}{E_{s_j}^3(t)} + \mathcal{H} \left\{ -\Delta\Delta R_{ij}(t) \frac{s_j(t)\mathcal{H}\{s_j(t)\}}{E_{s_j}^3(t)} + \Delta\Delta I_{ij}(t) \frac{s_j^2(t)}{E_{s_j}^3(t)} \right\} \right] \delta(\mathbf{x} - \mathbf{x}_i). \quad (\text{A13})$$

We demonstrate an example of differential time/phase measurements and their adjoint sources via the pairing of two stations at AFI and ABKT, shown in Fig. A1(a-I) and (a-II), with black lines representing observed seismograms and red lines computed seismograms. Fig. A1(b-I) and (b-II) are the adjoint sources at station AFI and ABKT, respectively, using their differential CC measurements (DD_CC); Fig. A1(c-I) and (c-II) are the adjoint sources for the differential MT (DD_MT); Fig. A1(d-I) and (d-II) are the adjoint sources for the double-difference EP (DD_EP). Similar to the absolute phase measurements, DD_CC uses a single value to characterize the difference between the time-shift of observations at AFI and ABKT and that of their synthetics, while DD_MT captures the difference of the frequency-dependent differential traveltimes measurements, and DD_EP tracks the time-continuous differential phase mismatch.

A3 The relationship between the instantaneous and the exponentiated phase

We offer two numerical examples to demonstrate the relationships that tie together the instantaneous (IP) and exponentiated (EP) phase measurements and their adjoint sources in representing phase information.

To understand whether the proposed EP method can adequately handle the discontinuities in phase which occur while measuring IP, we turn to Example I in Fig. A2, where the seismogram is from the Global Seismographic Network for the August 25, 2008 Xizang Tibet

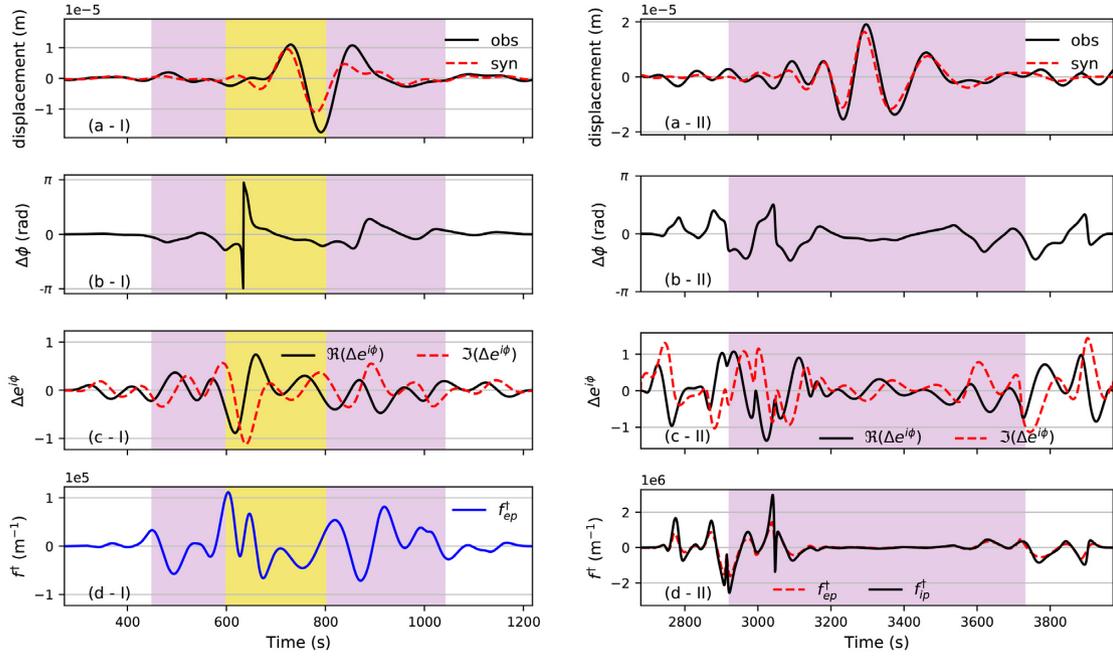


Figure A2. The relationship between instantaneous-phase (IP) and exponentiated-phase (EP) measurements. The left-hand column is Example I, a scenario when the phase difference between observations and synthetics exceeds the $\pm\pi$ limit (the 2-argument arc-tangent was used). The right-hand column shows Example II, a scenario when the phase difference remains within the $\pm\pi$ interval. (a-I) Observed and computed seismograms. (b-I) Instantaneous phase difference measured between the observed and the synthetic traces, with the phase jump highlighted by the yellow window. (c-I) Real (solid line) and imaginary (dashed line) parts of the normalized difference of the analytic observed and synthetic signals, that is the exponentiated phase difference. (d-I) The adjoint source for the EP difference. (a-II) Observed and computed seismograms. (b-II) Measured instantaneous phase difference of the observed and the synthetic traces. (c-II) Real (solid) and imaginary (dashed) components of the EP difference. (d-II) The adjoint source of the EP difference (red) compared to the adjoint source of the IP phase difference (black).

earthquake ($M_w = 6.7$, depth = 17 km, CMT 200808251322A) recorded at station TLY. We measure the real and the imaginary phase difference in Fig. A2(c-I) using eq. (15). The discontinuity present in the IP, in Fig. A2(b-I), is tuned to be continuous in the complex EP. Therefore, the EP method addresses phase discontinuities while measuring phase. Fig. A2(d-I) shows the resulting adjoint source.

To further understand the relationships between IP and EP we turn to Example II in Fig. A2, in which the phase difference lies within the $\pm\pi$ interval, and thus phase wrapping is not present in IP. Fig. A2(a-II) displays the data (black line, station AFI) and the synthetic trace (red line). We measure their phase difference using both IP in Fig. A2(b-II) and EP in Fig. A2(c-II). Although it is difficult to discern any similarities between IP and EP in terms of measurements, their adjoint sources in Fig. A2(d-II) betray their connections. When the phase difference, as in Fig. A2(b-II), is small (within $\pm\pi/4$), IP and EP adjoint sources are closely aligned. However, when the phase difference is large, EP tends to taper the big phase change, and thus smooth out the sharp rises and falls in the adjoint source.

A mathematical explanation for our observations can be found by taking the sine of the phase difference term, $\phi_d - \phi_s$ in eq. (11),

$$f^\dagger(t) = \sum_r \left[\frac{\sin[\phi_d(t) - \phi_s(t)]}{E_s^2(t)} \mathcal{H}\{s(t)\} + \mathcal{H} \left\{ \frac{\sin[\phi_d(t) - \phi_s(t)]}{E_s^2(t)} s(t) \right\} \right] \delta(\mathbf{x} - \mathbf{x}_r). \quad (\text{A14})$$

Subsequently, we rewrite eq. (18) as

$$f_{\text{EP}}^\dagger(\mathbf{x}, t) = \sum_r \left(\left[\Delta I(t) \frac{s(t)}{E_s(t)} - \Delta R(t) \frac{\mathcal{H}\{s(t)\}}{E_s(t)} \right] \frac{\mathcal{H}\{s(t)\}}{E_s^2(t)} + \mathcal{H} \left\{ \left[\Delta I(t) \frac{s(t)}{E_s(t)} - \Delta R(t) \frac{\mathcal{H}\{s(t)\}}{E_s(t)} \right] \frac{s(t)}{E_s^2(t)} \right\} \right) \delta(\mathbf{x} - \mathbf{x}_r), \quad (\text{A15})$$

where we use eq. (15) to simplify

$$\Delta I(t) \frac{s(t)}{E_s(t)} - \Delta R(t) \frac{\mathcal{H}\{s(t)\}}{E_s(t)} = \left[\frac{\mathcal{H}\{d(t)\}}{E_d(t)} - \frac{\mathcal{H}\{s(t)\}}{E_s(t)} \right] \frac{s(t)}{E_s(t)} - \left[\frac{d(t)}{E_d(t)} - \frac{s(t)}{E_s(t)} \right] \frac{\mathcal{H}\{s(t)\}}{E_s(t)} \quad (\text{A16})$$

$$= \frac{\mathcal{H}\{d(t)\}}{E_d(t)} \frac{s(t)}{E_s(t)} - \frac{d(t)}{E_d(t)} \frac{\mathcal{H}\{s(t)\}}{E_s(t)} \quad (\text{A17})$$

$$= \sin[\phi_d(t)] \cos[\phi_s(t)] - \cos[\phi_d(t)] \sin[\phi_s(t)] \quad (\text{A18})$$

$$= \sin[\phi_d(t) - \phi_s(t)] = \sin[\Delta\phi]. \quad (\text{A19})$$

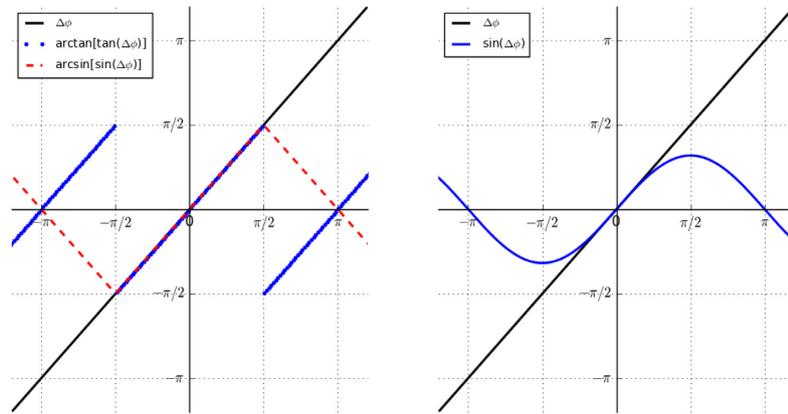


Figure A3. Left-hand panel: the behaviour of $\arctan[\tan(\Delta\phi)]$ and $\arcsin[\sin(\Delta\phi)]$ explains the phase wrapping observed in measuring phase explicitly in the IP method: the blue line demonstrates how the arc-tangent function leads to phase jump once $\Delta\phi$ exceeds the limit, and the red line shows that arc-sine avoids the phase jump but still has cycle skips. Right-hand panel: the behaviour of $\sin(\Delta\phi)$ versus $\Delta\phi$ shows the tapering effect of the sin function in the EP method.

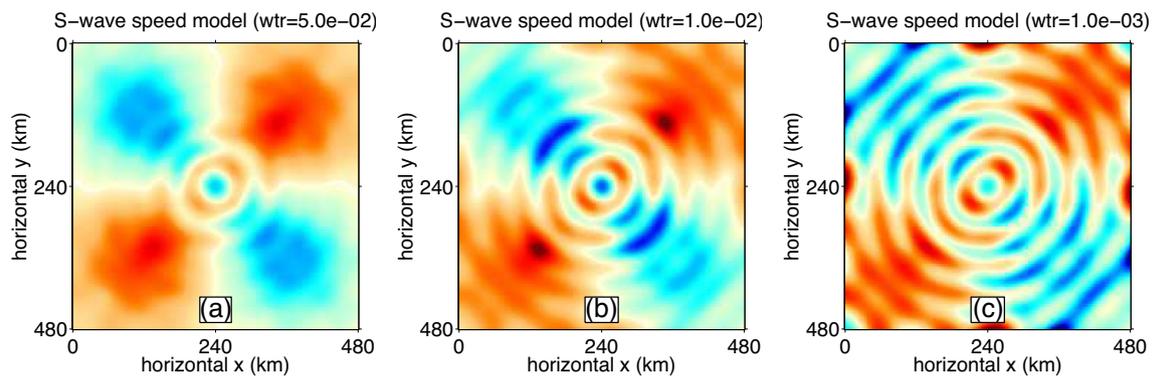


Figure A4. Inversion results using EP measurements with different water level values demonstrate the effect of water level (wtr). This experiment, as in Fig. 7, starts from a homogeneous initial model with $V_S = 3500 \text{ m s}^{-1}$.

In that case eq. (A15) becomes identical to eq. (A14), thus showing that the sine of the modified IP adjoint source and the proposed EP adjoint source in eq. (18) are each other's equivalent.

Fig. A3 demonstrates the difference and connections of the explicit IP approach discussed by Bozdağ *et al.* (2011) (left-hand panel, blue line), the modified IP method of Rickers *et al.* (2012) (left-hand panel, red line) and our proposed EP (right-hand panel, blue line). Fig. A3 (left-hand panel) shows the discontinuities in phase that arise while measuring phase explicitly. Both the arc-tangent and arc-sine are uniquely defined in the $\pm\pi/2$ interval. As opposed to the arc-tangent function, the arc-sine does not switch sign immediately upon exceeding the limit, thus avoiding rapid phase changes. Clearly, it still suffers from phase jumps, and Fig. A3 (right-hand panel) explains how the proposed EP approach turns large phase measurements into tapered continuous quantities. For small phase measurements, all three approaches are expected to give similar results.

The choice of water level will also influence the tapering effect. We conducted inversions for the experiment in Fig. 6, by using different water level values. The result in Fig. A4 shows that a larger water level would taper small-magnitude waves and result in the fine-scale structure unresolved, and a smaller water level tends to boost small signals and reveal detailed fine structure. However, in real observations with the presence of noise, we have to adjust the water level properly to intentionally downweight any data below noise level.