Geophysical Journal International

Geophys. J. Int. (2026) **245,** 1–11
Advance Access publication 2025 February 27
Research Paper

https://doi.org/10.1093/gji/ggaf503

Non-physical head wave identification in *P*-wave arrivals retrieved from seismic ambient noise at the Chémery underground gas storage (France)

Ali Riahi [1,2] Alexandre Kazantsev,[3] Eléonore Stutzmann,[1] Martin Schimmel [4] Jean-Paul Montagner,[1] Mark Noble[5] and Jean-Philippe Métaxian[1]

[1] *Université Paris Cité, Institut de Physique du Globe de Paris, CNRS,* 75005 *Paris, France. E-mail: riahi@ipgp.fr*
[2] *Department of Earth Sciences, University of Geneva, Rue des Maraîchers* 13, 1205 *Geneva, Switzerland*
[3] *Storengy, 92250 La Garenne-Colombes, France*
[4] *Geosciences Barcelona, GEO3BCN-CSIC, 08028 Barcelona, Spain*
[5] *Mines Paris, PSL University, Centre for Geosciences and Geoengineering, 77305 Fontainebleau, France*

Accepted 2025 December 5. Received 2025 December 1; in original form 2025 September 3

SUMMARY

This study aims to retrieve *P* waves from seismic ambient noise recorded by a dense local broad-band network at the Chémery underground gas storage site, where anticline deformation was previously identified through wells and seismic reflection survey. To this end, we propose a new approach for reconstructing *P* waves from ambient noise. We process the passive seismic data to reconstruct the body wave component of the empirical Green's functions. The retrieved *P*-wave arrivals were identified and analysed, revealing that in this data set, the picked arrival times likely correspond to non-physical head waves rather than direct or diving *P*-wave arrivals between virtual sources and receivers. Numerical simulations support this idea of non-unique interpretation of the passively reconstructed *P*-wave arrivals. The simulations suggest the potential for mapping lateral heterogeneities above the critical refractor as a cumulative time-delay, although for this data set the anticline-induced time-delay is likely within the measurement uncertainties. It is found that the dominance of non-physical head waves over diving waves is primarily due to the large distance from the network to ambient noise sources.

Key words: Body waves; Seismic interferometry; Seismic noise; Eikonal equation; Empirical Green's function.

1 INTRODUCTION

Seismic ambient noise holds valuable information about the subsurface. The cross-correlation of ambient noise recordings at two receivers is closely related to the medium impulse response between the two receiver locations (O.I. Lobkis & R.L. Weaver 2001; A. Derode et al. 2003; K. Wapenaar 2004; N.M. Shapiro & M. Campillo 2004; R. Snieder 2004). Surface waves often dominate the ambient noise energy (K.G. Sabra et al. 2005; N.M. Shapiro et al. 2005), as demonstrated by numerous studies (e.g. A. Köhler et al. 2015; A. Obermann et al. 2016; M. Lehujeur et al. 2018; M. Peruzzetto et al. 2018; T. Yudistira et al. 2021; J. Carvalho et al. 2022). Conversely, extracting low-energy body waves from ambient noise is typically challenging. The primary challenge includes separating body waves from the dominant surface waves and removing inherent theoretical artifact cross-terms (A. Riahi et al. 2021). Some studies have successfully extracted *P*-waves from local network recordings (e.g. P. Roux et al. 2005; J. Zhang & P. Gerstoft 2014; N. Nakata et al. 2015, 2016; A. Riahi et al. 2021, 2023), and a few have performed 3-D *P*-wave traveltime tomography (e.g. N. Nakata et al. 2015; A. Riahi et al. 2023).

Various processing strategies have been developed for the extraction of *P*-waves from ambient noise. N. Nakata et al. (2015) utilized 10 d of noise data recorded by a dense seismic array with 2500 vertical component receivers and 100 m spacing at regional scale. They obtained an average virtual *P*-wave shot from common offset binning-stacked traces, which was then used as a reference template for selective filtering to select traces with high-quality *P*-wave arrivals. An Adaptive Covariance Filter was also employed to enhance the signal-to-noise ratio of coherent *P*-wave arrivals. Subsequently, A. Riahi et al. (2023) performed 3-D P-wave traveltime tomography using approximately one year of continuous data from a dense array with around 100 stations and an interstation spacing of 2 km. Their modified methodology was based on the polarization-based approach of R. Takagi et al. (2014), which enables the separation of retrieved *P* and Rayleigh waves. The summation of ZR and RZ components of the Green's tensor can enhance *P*-waves and eliminate Rayleigh waves. A. Riahi et al. (2021, 2023) emphasized that

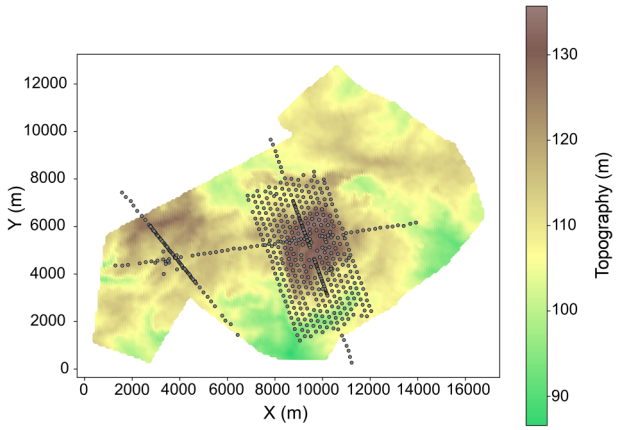


Figure 1. The location of the Chémery array (grey circles), superimposed over the topography of the area.

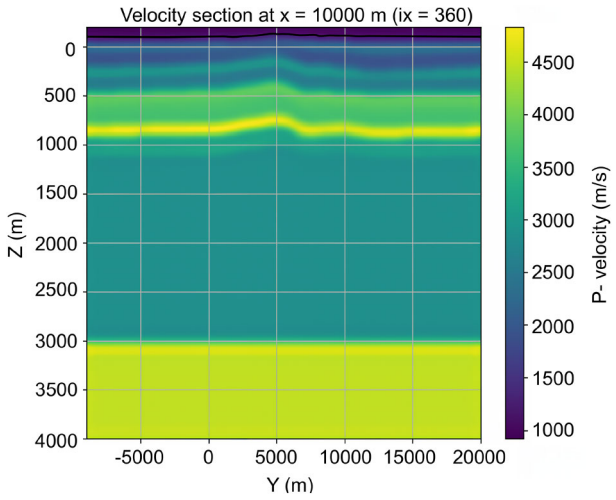


Figure 2. Cross-section of the V_P velocity model used for eikonal simulations.

only a few portions of the data contain reliable *P*-wave energy, requiring careful data selection.

The objective of this study is to extract *P*-wave traveltimes and evaluate their potential for imaging at the underground gas storage site located in Chémery, France. This study modifies previously developed methodologies to accommodate the specific features of this very local dense network deployed over several days. After introducing the data set and geological context of the site (Section 2), the paper details the *P*-wave extraction method (Section 3) and compares the results to numerical traveltime simulations and identification of non-physical head waves (Section 4), allowing for a discussion of the method's potential and limitations (Section 5).

2 DATA SET AND LOCAL GEOLOGY

2.1 Data set

The ambient noise data set used in this study was recorded by a network deployed over 16 d in November 2010 (see Fig. 1). This regular network consisted of approximately 100 three-component broad-band Trillium 20s seismometers with 400 m interstation spacing. To increase the density of recording points (for an objective unrelated to cross-correlation approaches), the network was shifted daily, resulting in each point recording ambient noise data for no more than 2–4 consecutive days. Fig. S1 (Supplementary Material) illustrates the daily evolution of the network geometry. In total, 580 recording locations are available in the data set. Also see fig. 1 of N. Riahi et al. (2013), for more details.

M. Peruzzetto et al. (2018) and C. El Khoury et al. (2024) used beamforming to characterize the composition of the ambient wavefield at the Chémery site, using the same data set. Both studies revealed the presence of *P* waves between 1 and 2 Hz, particularly strong during nighttime when anthropogenic Rayleigh waves are the weakest. The back-azimuth of *P* waves was found to be in the direction of the North Atlantic. It is important to note that the nearest coastline, and thus the closest potential *P*-wave source, is located more than 200 km away from the site.

2.2 3-D velocity model for the Chémery site

The Chémery site is located within a nearly flat sedimentary sequence in the Paris basin. The velocity model of the Chémery site is constrained by sonic logs from approximately one hundred wells and 3-D reflection seismic data, making it highly reliable. We run a 3-D finite-difference eikonal algorithm (M. Noble et al. 2014) to simulate the first arrival of the *P* wave within this model. Given that our real data set is based on low-frequency waveforms (1–2 Hz), we smooth the velocity model to reduce potential high-frequency propagation effects. A cross-section of the resulting 3-D model is shown in Fig. 2. The first kilometre consists of the well-constrained sedimentary pile. A high velocity, 100 m thick layer at a depth of 850 m corresponds to tight Bathonian carbonates, which are likely to act as a critical refractor in the simulations. We assume that the bedrock, another potential critical refractor, is located at a depth of approximately 3 km. The sediments between 1 and 3 km are modelled as a homogeneous layer with a *P*-wave velocity of 3 km s^{-1}. This model is employed to accurately determine the type of the retrieved *P* wave and to evaluate the potential for imaging the anticline structure on that basis.

3 *P*-WAVE TRAVELTIME EXTRACTION

3.1 Waveform processing workflow

3.1.1 Empirical Green's tensor construction

Prior to computing the empirical Green's function (EGF), a preprocessing procedure is applied to the raw ambient noise recordings. Initially, the data is segmented into 10-min windows with a 50 per cent overlap. Each window is then demeaned, detrended, tapered with a cosine function and decimated to 10 Hz in order to optimize memory usage and computational efficiency for subsequent steps. A sixth-order bandpass Butterworth filter is applied between 1.0 and 4.5 Hz to enhance coherent energies. The Cartesian horizontal components (E, N) are rotated into the local radial-transverse base (R, T) specific to each receiver pair. Subsequently, the 5 components of the empirical Green's tensor (ZZ, ZR, RZ, RR and TT component pairs) are estimated in the frequency domain using the cross-coherency formula proposed by N. Nakata et al. (2015), with a stabilization and regularization factor set at 0.3. This factor can take values between 0 and 1. When the parameter is set to zero, the formula corresponds to the classical cross-coherency, which whitens

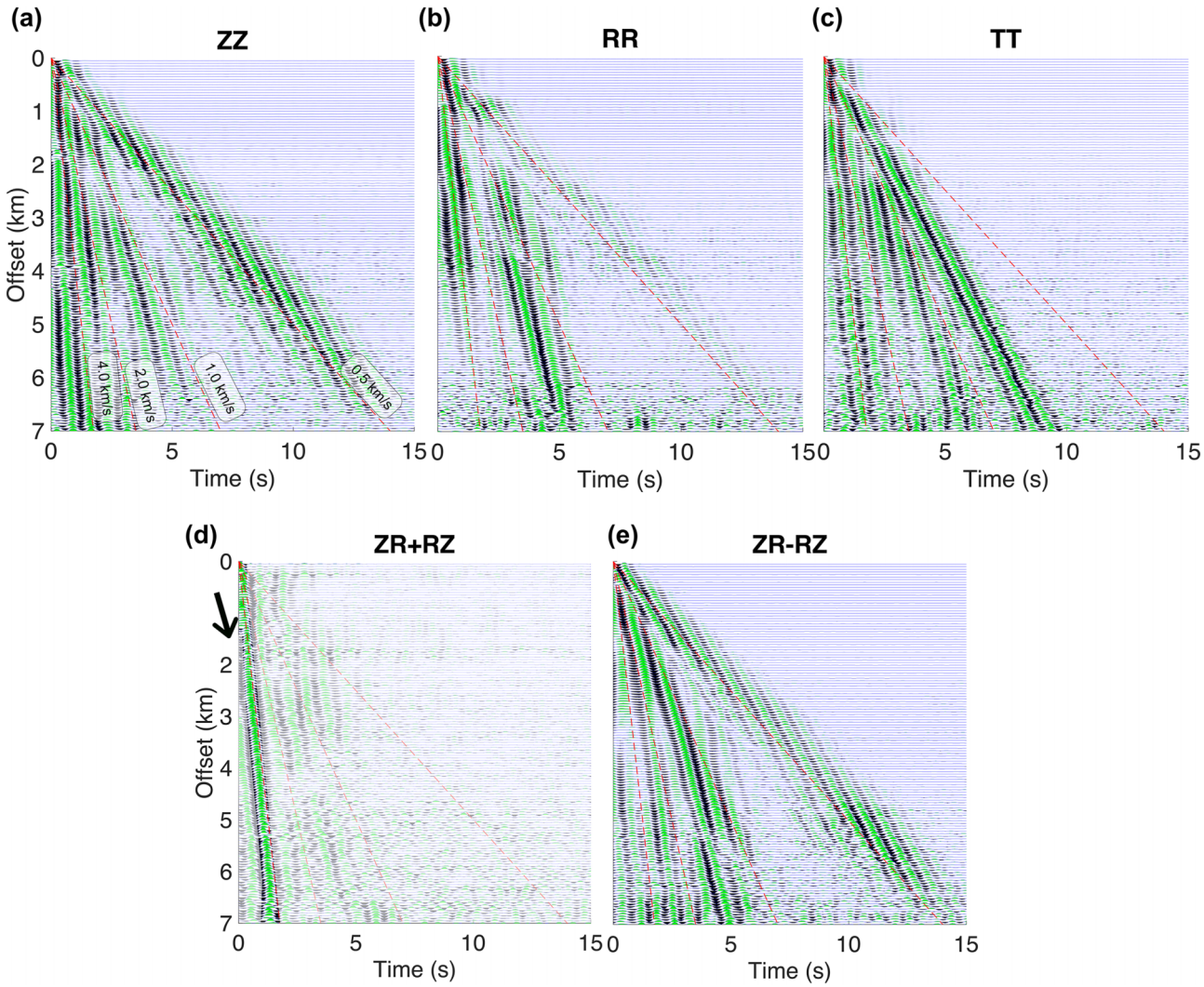


Figure 3. Common-offset bin-stack of EGFs for all virtual sources and receivers within offset bins of 50 m. The subplots relate to Green's tensor components (a) ZZ, (b) RR, (c) TT, (d) ZR+RZ and (e) ZR−RZ. Fundamental and higher modes of Rayleigh waves appear in panels (a), (b) and (e), as well as Love waves in panel (c). Enhanced *P*-waves with muted Rayleigh waves are visible in panel (d). In (d), the spurious and later arrivals are masked to highlight the retrieved *P*-wave moveout for better visualization. The highlighted *P*-phase moveout is selected following the criteria described in Section 3.2.1 and fig. 7 of A. Riahi et al. (2021). The dotted lines represent apparent velocities of 0.5, 1, 2 and 4 km s^{-1}. The arrow shows the moveout of the spurious arrival.

the amplitude spectra across all frequencies but can produce unstable results. Increasing the parameter enhances the influence of dominant frequencies in the cross-coherency output, reduces spectral whitening and improves numerical stability (N. Nakata et al. 2011).

For each Green's tensor component, a common-offset bin-stack is computed by summing EGFs for all virtual sources and receivers within offset bins of 50 m (E. Ruigrok 2014; N. Nakata et al. 2015; N. Nakata & K. Nishida 2019). Bin-stacking averages lateral variations and compresses 3-D information into 1-D. Although not suitable for 3-D imaging, bin-stacking effectively enhances the signal-to-noise ratio and amplifies low-amplitude phases. Fig. 3 illustrates the results of bin-stacking for ZZ, ZR, RZ, RR and TT components of the Green's tensor, with trace amplitudes normalized for better visualization of low-amplitude phases.

In the ZZ component (Fig. 3a), a clear fundamental mode Rayleigh wave with dispersive behaviour and an apparent group velocity of approximately 0.5 km s^{-1} is observed. This is consistent with previous array processing studies performed on this data set (M. Peruzzetto et al. 2018; C. El Khoury et al. 2024). Distinct faster arrivals, likely corresponding to higher modes of Rayleigh waves and body waves, are also visible in the ZZ component.

The RR component (Fig. 3b) shows a strong higher mode Rayleigh wave with an apparent group velocity of roughly 1.5 km s^{-1}. A distinct faster arrival at shorter offsets may correspond to body wave energy. K. Nishida et al. (2008) demonstrated the feasibility of retrieving *P* waves from the RR component.

The TT component (Fig. 3c) displays the fundamental mode of the Love wave with an apparent velocity of around 0.8 km s^{-1}, alongside several higher modes.

3.1.2 ZR+RZ component combination: Rayleigh wave muting

Dispersive Rayleigh waves represent significant energy in the ZZ and RR components, from which *P* waves need to be extracted. Following R. Takagi et al. (2014), Rayleigh waves are eliminated by separating the imaginary parts of the ZR and RZ components in the frequency domain, and then summing them in the time domain. This

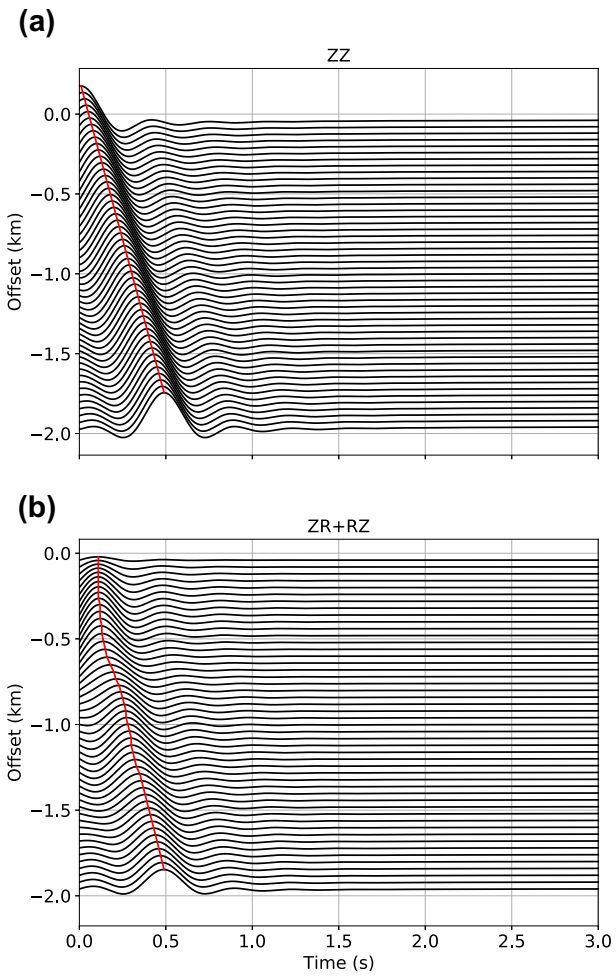


Figure 4. Synthetic test demonstrating the side-lobe effect on phase picking of (a) ZZ and (b) ZR+RZ components.

summation eliminates the Rayleigh wave due to the $\pi/2$ phase shift between the vertical and radial motions in Rayleigh wave polarization. The result (Fig. 3d) shows nearly absent surface waves and enhanced fast *P* waves, with body wave arrivals exhibiting an apparent velocity of about 4.0 km s^{-1}. Some earlier spurious arrivals are notable (black arrow in Fig. 3d), resulting from the interference of body and surface wave energies during EGF calculation (F. Forghani & R. Snieder 2010; N. Nakata & K. Nishida 2019). A. Riahi et al. (2021) studied the characteristics of these spurious arrivals using several synthetic tests and demonstrated that they may appear before and after the *P* wave and are eliminated only under an even radial distribution of noise sources. This condition is not satisfied for the Chémery data set.

Separating the real parts of the ZR and RZ components in the frequency domain and then subtracting them in the time domain enhances Rayleigh wave polarization and eliminates body waves, for which vertical and radial motions are in phase. The result of this component combination after returning into the time domain (Fig. 3e) exhibits the fundamental and higher mode Rayleigh waves, with moveouts consistent with those identified in the ZZ (Fig. 3a) and RR (Fig. 3b) components. The ZR–RZ component combination is devoid of *P*-wave energy, consistent with R. Takagi et al. (2014).

Given that the ZR+RZ component combination shows enhanced body wave energy, we focus exclusively on this component in the next sections.

3.2 Specific features of ZR+RZ component combination

In this section, we perform two numerical tests to illustrate two particular features of the ZR+RZ component combination that are crucial for traveltime picking: waveform distortion at near offset and the polarity of the *P* arrival. Both tests are conducted using the wave number integration approach of the *Computer Programs for Seismology* software (R.B. Herrmann 2013).

3.2.1 Near offset artifact

The first synthetic test assumes a homogeneous half-space medium with a *P* velocity of 4.0 km s^{-1}, a vertical source emitting a Ricker wavelet with a frequency of around 1.9 Hz, and a line of 40 receivers spaced 50 m apart. After simulating simple non-dispersive *P*-wave arrivals from this source at each receiver, we calculate the cross-correlation of the ZZ and ZR+RZ components, as shown in Fig. 4. The maximum amplitude corresponds to the theoretical arrival time, marked in red.

Fig. 4 shows that while the retrieved phase in the ZZ component exhibits an autocorrelation peak at zero time-lag, the ZR+RZ component combination is zero at zero time-lag due to its antisymmetric nature. This is consistent with the physical interpretation of the Green's tensor, whose ZR+RZ component combination is zero at the source location because of symmetry considerations. In practice, traveltime picking is often performed by following the maximum or the onset of the waveform (the first zero-crossing preceding the maximum). Fig. 4(b) illustrates the potential bias in the picked arrival time at short offsets, due to the antisymmetric nature of the ZR+RZ component combination. Fig. 3(d) exhibits a similar pattern for the real data. Based on these considerations and visual inspection of Fig. 3(d), we decide to discard all traces at offsets smaller than 2 km from traveltime picking. Assuming a velocity of 4 km s^{-1} and a dominant frequency of 1.5 Hz, this corresponds to removing traces at offsets below one wavelength, which is a standard requirement in ambient noise cross-correlation methods (N.M. Shapiro & M. Campillo 2004; N.M. Shapiro et al. 2005; G.D. Bensen et al. 2007).

3.2.2 P-wave polarity

A second synthetic test explores how the polarity of the *P*-wave arrival is influenced by the balance of causal and acausal energies. We consider a homogeneous half-space medium with a *P* velocity of 5.0 km s^{-1} and an *S* velocity of 1.8 km s^{-1}. These values were selected for illustrative purposes and are independent of the Chémery velocity model. We also define a single receiver pair, and 28 explosion sources on both sides. Synthetic seismograms contain *P*- and Rayleigh waves. We explore two scenarios of uneven noise source distribution: in scenario A, 20 out of the 28 sources are placed on the right side of the receiver pair, and the remaining 8 on the left; in scenario B, the dominant number of sources is on the left side (Fig. 5). We then compute the cross-correlation of the ZZ, RR, ZR+RZ and ZR−RZ components.

The results in Fig. 5 reveal that the dominant polarity of the P and Rayleigh waves is always positive for the ZZ and RR component pairs, both in the causal and acausal parts of the cross-correlation. However, the polarity of the *P* wave for the ZR+RZ component combination depends on the location of the dominant noise sources. While the ZZ and RR components exhibit the well-known imbalance of causal and acausal parts, the ZR+RZ component combination is antisymmetric by construction. The *P* arrival has a positive polarity

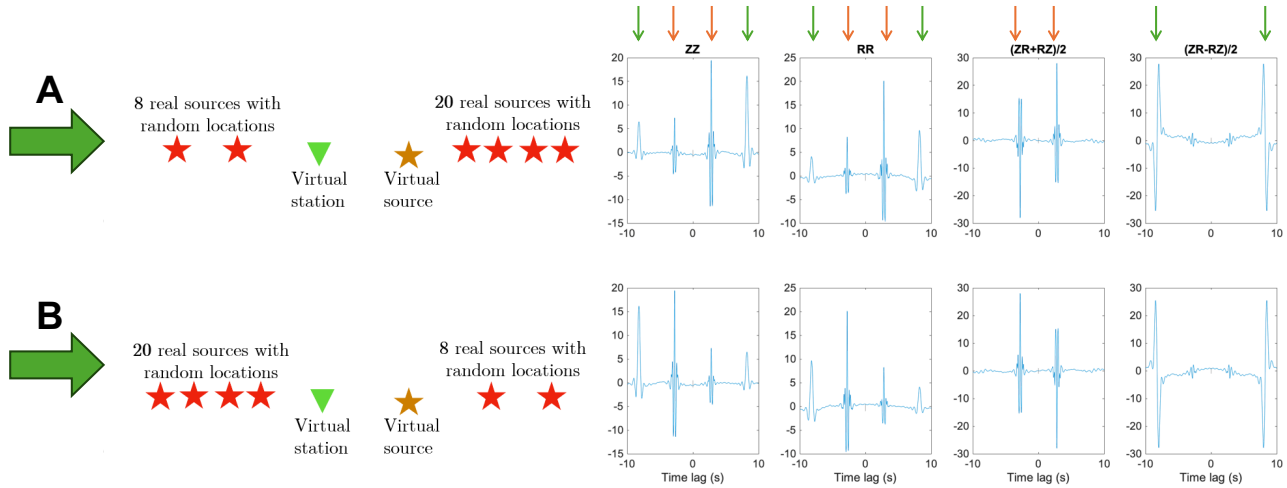


Figure 5. The effect of the uneven distribution of noise sources located in stationary zones of two receivers (labelled virtual source and virtual station) on the polarization and amplitude of different components. The noise sources emit *P* and Rayleigh waves simultaneously. (A) Dominant noise sources are located alongside the virtual source to the virtual receiver, leading to dominant amplitude in the positive lag of ZZ and RR components and positive polarization of the *P* wave in the positive lag of the ZR+RZ component. (B) When the dominant noise source is located on the opposite side, higher amplitudes at negative lags of ZZ and RR components are expected, as well as the negative amplitude of the *P* wave in the positive lag of the ZR+RZ component. The orange and green arrows refer to *P* and Rayleigh phases, respectively.

when the sources are dominantly located on the virtual source side (Fig. 5a), and a negative polarity when they are dominantly located on the virtual receiver side (Fig. 5b). Additionally, the ZR+RZ component combination mutes Rayleigh waves, while the ZR−RZ component combination mutes *P* waves.

The two implications of this second test are that for picking *P*-wave traveltime on the ZR+RZ component, we need to:

(1) determine the expected noise dominant azimuth based on detecting the lag with dominant energy in the ZZ or RR components.

(2) take the *P*-wave polarity into account in the traveltime picking algorithm.

The positive polarity observed in the common-offset bin-stacked gather in Fig. 3(d), with an arrival velocity of approximately 4.0 km s^{-1}, indicates that most of the body wave noise sources (located in the Atlantic ocean) generate energy in the causal part of the cross-correlations. This observation is valid only given the arbitrary east–west orientation convention of the station couples in the processing. It would change if we made a different orientation choice.

3.3 Traveltime measurement

While the common-offset bin-stacked gather in Fig. 3(d) exhibits a clear *P*-wave arrival, identifying this arrival for individual receiver pairs (or individual traces) is more challenging. Extracting this arrival from individual traces is the objective of this section.

3.3.1 Trace selection with semblance criterion

Estimating the EGFs between all available receiver pairs of the Chémery data set results in more than 90 000 daily individual traces (Fig. 6a). N. Nakata et al. (2015) and A. Riahi et al. (2023) have shown that only a subset of these traces usually contain reliable *P*-wave energy, while the rest should be discarded. Factors such as the lack or weakness of body wave energy, presence of noise sources in non-stationary zones, and coherent artifacts can contribute to the unreliability of individual traces.

To select the traces containing reliable *P*-wave energy, we use a template matching approach based on the common-offset bin-stacked gather in Fig. 3(d). Initially, we categorize the individual traces into two groups: *expecting positive polarity* and *expecting negative polarity*, based on whether the dominant amplitude of the ZZ component occurs at a positive or negative lag. Then, the *P*-wave arrival is isolated on the bin-stacked gather using a specific space-time taper, which is zero outside the interval defined by a moveout of 4.0 m s^{-1} and a width of 0.4 s. A 30 per cent cosine taper is used within the interval. The same taper is applied to individual traces based on the offset of the corresponding receiver pairs. Subsequently, we calculate the Correlation Coefficients (CC) between the tapered common-offset bin-stacked gather and all the tapered individual traces (N. Nakata et al. 2015). Only traces for which the CC is greater than a certain threshold are kept for arrival time picking. The process is performed separately for traces with *expecting positive* and *expecting negativeP*-wave polarity.

The threshold value is iteratively increased from 0.5 to 0.8 in steps of 0.1, with the common-offset bin-stacked gather being updated at each iteration. At the end of this procedure (4 iterations), approximately 7000 (resp. 4500) traces with positive (resp. negative) polarity are selected (Figs 6b and c).

In contrast to Figs 6(a)–(c) exhibit a clear *P*-wave arrival, illustrating the efficiency of the template matching selection procedure. Some later arrivals are also visible and may correspond to reflected or converted *P* and *S* waves.

3.3.2 Traveltime picking

P-wave arrival times are picked manually from the selected individual traces. Referring to the synthetic test illustrated in Fig. 4(b), we identify the maximum (resp. minimum) amplitudes representing the *P*-wave arrival time for traces with positive (resp. negative) polarity. As previously mentioned, we discard receiver pairs with offsets less than 2.0 km due to the near-offset artifact on the ZR+RZ component. During manual picking, we discard traces

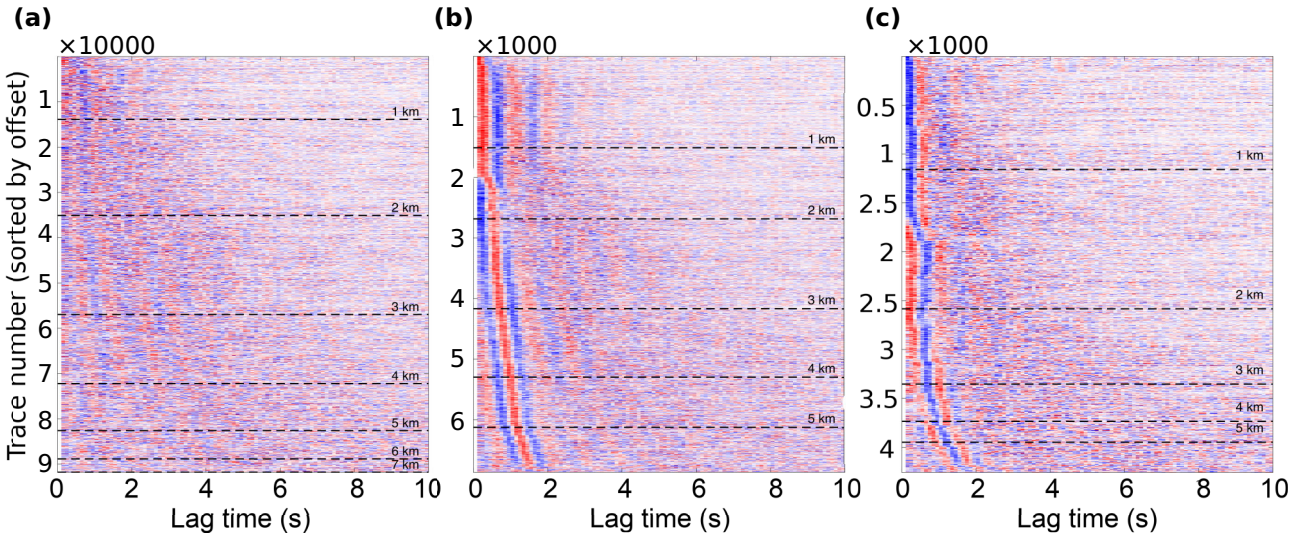


Figure 6. (a) All individual daily empirical Green's functions, (b) the selected individual traces with positive polarity of the *P*-phase dominant energy, (c) the selected individual traces with negative polarity of *P*-phase dominant energy. Note that the *Y*-axes are the trace number, sorted by distance. Dashed lines mark some distances. Red (resp. blue) colours indicate positive (resp. negative) values. The amplitudes of the individual traces have been normalized for better illustration.

with ambiguous signals where the *P* arrival cannot be clearly identified. The final picked *P*-wave arrival times of positive and negative polarities are shown in Fig. 7(a). Some examples of the picked arrival times overlaid on traces are illustrated in Fig. S2 (Supplementary Material). to provide an idea of the signal quality and coherency.

4 *P*-WAVE TRAVELTIME MODELLING

The objective of this section is to determine the nature of the first-arrival *P* phase, picked in Fig. 7(a), and assess its potential for imaging the anticline structure. To achieve this, we will use the highly reliable 3-D velocity model of the Chémery site to model the apparent first arrival times in the EGFs under two different assumptions: (1) The EGF is a good approximation of the true GF for the first arrival, and (2) the EGF measures a propagation delay of an incident head wave.

4.1 Virtual source modelling

In classical ambient noise cross-correlation theory, the EGFs are assumed to approximate the Green's function between the pair of receivers. For this case, in our eikonal simulations (M. Noble et al. 2014), the modelled traveltimes were stored at each receiver location for each virtual shot point. We will refer to this simulation configuration as the 'collocated sources configuration'.

Results are shown in Fig. 7(b). It is clear that receiver pairs with interstation distances above 3000 m are in a head wave propagation regime (black ray path in Fig. 8). The time–distance law of the simulated *P*-wave arrival times exhibits an intercept time of around 0.5 s. A non-zero intercept time related to the overburden velocity is indeed expected (J.G. Hagedoorn 1959; W.M. Telford et al. 1990) and consistent with the value obtained from numerical simulations. However, such an intercept is not observed in the real data (Fig. 7a). This mismatch motivates an alternative interpretation of the data.

4.2 Head wave time-delay modelling

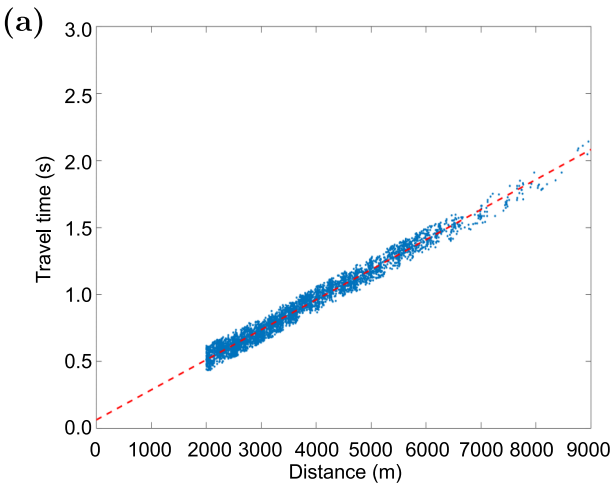
In this section, we propose to simulate the case where *P* waves extracted from ambient noise data are head waves with lateral 2-D propagation along a critical refractor. In this case, the first arrival time in the virtual shot corresponds to the time difference between two phases, which are the head wave from the same source and recorded at the two receivers. This means the EGF contains a so-called *non-physical* head wave (term used by S. Hanafy & G.T. Schuster 2017), since the corresponding virtual source location corresponds to neither of the receiver locations (ray path A'–B for a laterally homogeneous layer as shown in Fig. 8. This situation was previously studied by D. Mikesell & K. van Wijk (2011), P. Bharadwaj et al. (2012) and S. Hanafy & G.T. Schuster (2017) in the context of active refraction tomography.

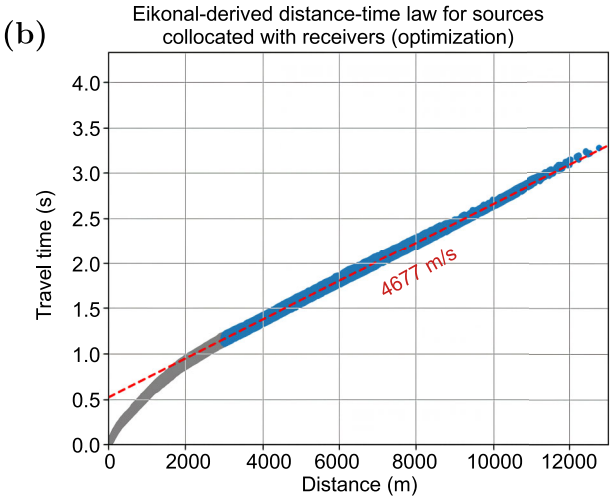
With this assumption, the standard derivations of the EGF theory for 2-D surface waves remain unchanged (see e.g. L. Boschi et al. 2013), except for a constant apparent velocity v_0 (=critical refractor velocity) replacing the dispersive surface wave velocity. Under this interpretation, the measured body wave traveltimes (Fig. 6) are simply the delay of the refracted wave front between the receivers. If we additionally assume that the cosine of the critical refraction angle is close to 1 (i.e. that the non-horizontal parts of the ray paths in Fig. 8 are subvertical), the non-physical head wave time (t_{kl} in red in Fig. 8 with $A \approx A'$) can be expressed as:

$$t_{kl} = s_0.d_{kl} + T_k - T_l, \quad (1)$$

where $s_0 = 1/v_0$ is the constant refraction wave slowness, and d_{kl} is the distance A'-B in Fig. 8 which is equal to the distance between the 2 stations k and l. $T_{k,l}$ are the traveltimes of the refracted wave along the non-horizontal part of the wave path at receivers k and l, which can be seen as site effects.

To model traveltimes under this alternative interpretation, surface sources are distributed along a truncated circle of radius 16 km around the barycentre of the receivers, with a step of 10° (see Fig. 9). Traveltimes are simulated from each source to each receiver.





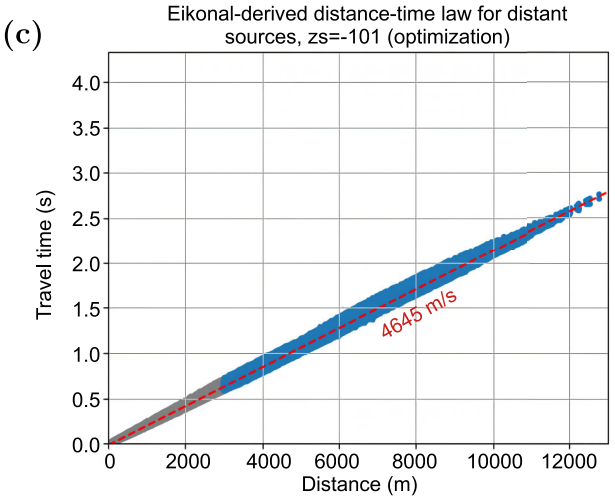


Figure 7. (a) Real first arrivals of the *P* wave, picked manually. (b, c) Synthetic traveltime curves based on the anticline velocity model shown in Fig. 2. (b) Eikonal-simulated traveltimes under the standard 3-D GF interpretation and (c) under alternative non-physical head-wave interpretation. The arrival times are shown as blue and grey dots (respectively, after and before offset selection). The dashed line represents the best-fit linear slope to the blue dots, while the grey dots are excluded from the fitting.

Traveltime estimation for the receiver pair kl is obtained as the difference of traveltimes to receivers k and l from the best-aligned source, corrected by the misalignment bias factor $\cos\phi$. As shown in Fig. 9, the RMS misalignment ϕ in this simulation setup does not exceed 2.5°, which represents only a negligible correction in practice. The resulting time–distance law of the simulated *P*-wave arrival times is shown in Fig. 7(c). It is remarkable that there is no intercept, just as in the real picked traveltimes, contrary to the standard GF interpretation prediction (Fig. 7b).

This indicates that the real data are compatible with the 'alternative' EGF interpretation as a non-physical head wave.

5 DISCUSSION

5.1 Isolating the body wavefield

In this paper, we propose a polarization-based approach to suppress the Rayleigh wave contribution from the reconstructed empirical Green's tensor. This procedure creates a side-lobe artifact at short offsets (see Fig. 4), requiring the elimination of short offsets from the data set in ZR+RZ component. However, this restriction is not more stringent than the classical 2λ limit (M. Campillo & A. Paul 2003; K. Wapenaar & J. Fokkema 2006).

Regarding body wavefield separation, an important observation concerns the polarity of the arrivals with the implementation of the ZR+RZ combination as proposed by R. Takagi et al. (2014). We demonstrated that the polarity depends on whether the original arrival is located in the causal or acausal part of the EGF. This means that in the presence of uneven source distributions, the polarity of the *P* arrival depends on the source location relative to the receiver pair (see Fig. 5). This poses a practical issue as the common-offset EGF stack will exhibit the most frequently encountered polarity among the pairs. To handle this, we treated receiver pairs with positive and negative *P*-wave polarities separately, with respect to the dominant lag of the ZZ component. An alternative approach would be to choose a station pair orientation convention in the direction of the dominant *P*-wave sources, theoretically yielding the same polarity for all receiver pairs.

Finally, for diving *P*-wave retrieval, it is assumed that we can pick the first *P*-wave arrival. However, once the body wavefield has been extracted from the empirical Green's tensor, it exhibits several coherent arrivals. Therefore, unambiguously identifying the one corresponding to the desired phase is challenging. The selection of the relevant coherent arrival is implemented through a time-domain taper and can be subjective. This issue is particularly important as we reject a significant number of traces (N. Nakata et al. 2015; A. Riahi et al. 2023) based on a semblance criterion established between the tapered common-offset stack and individual traces. The choice of the apparent velocity selected by the taper heavily influences the final set of traces kept for picking the arrival times. For the Chémery data set, we tested different apparent velocities for the taper definition and ensured that the chosen setting was a local maximum for the number of selected traces.

5.2 Accurately picking the arrival times

The precision of traveltime picking is crucial for imaging. The required level of precision is determined by the magnitude of the

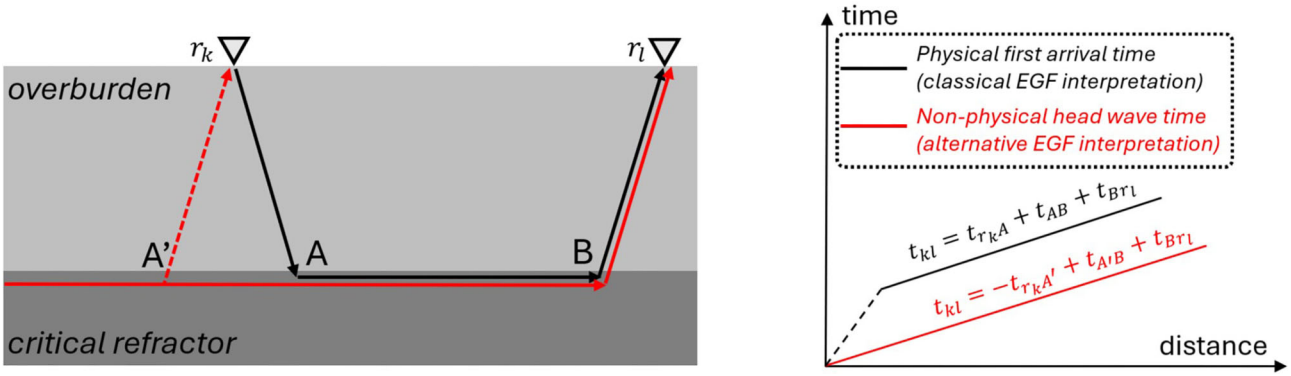


Figure 8. Conceptual illustration of ray paths under the two possible EGF interpretations and associated arrival times. Black path: physical first arrival for a virtual source in r_k and a receiver in r_l (note that the associated traveltime law at far offset exhibits a positive intercept). Red path: head waves from a distant source to the left. In a laterally homogeneous layer the non-physical ray-path is A'–B as all the other segments cancel out in the interferometric approach.

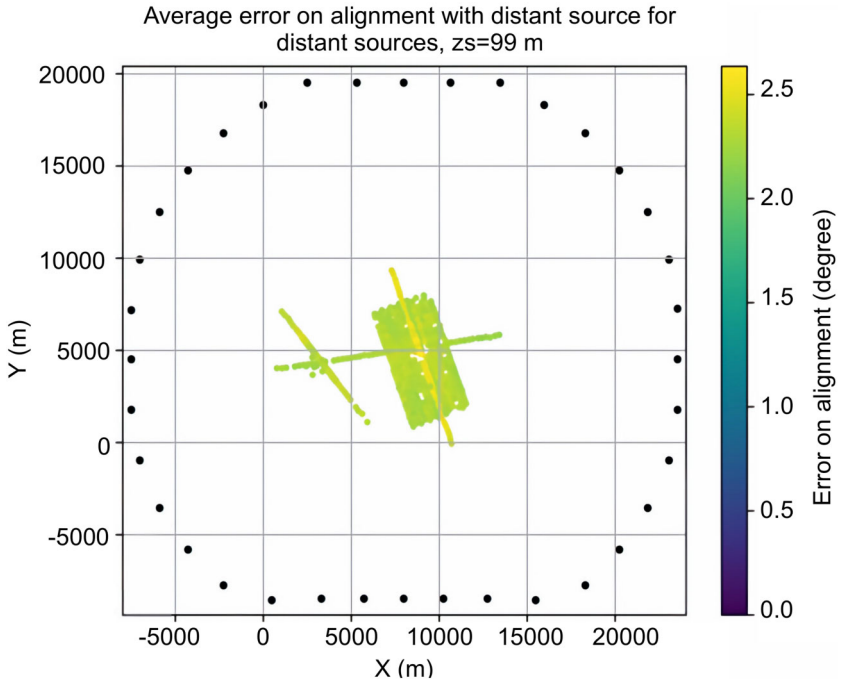


Figure 9. Eikonal simulation setup under alternative non-physical head wave interpretation. The array is surrounded by distant sources. The best-aligned source is used for each station pair. The colour bar represents the RMS misalignment for each station.

expected traveltime anomalies caused by structural heterogeneities in the medium. In the case of Chémery, the subsurface consists of relatively flat, layered strata with a smooth anticline structure exhibiting a vertical relief of approximately 100 m. Such a structure is expected to induce maximum traveltime perturbations on the order of 0.02 s (see Section 5.4 below). Achieving this level of picking accuracy is challenging for waveforms with a frequency content around 1 Hz without an interferometric approach. In this work, we used absolute traveltime picking based on waveform maximum detection, as previously done by N. Nakata et al. (2015). While this method is insufficient for imaging of the anticline, it is adequate for illustrating a more fundamental evaluation of the method, as discussed in the following paragraph.

5.3 Finding a proper forward model for the arrival times

One important point of this work concerns the modeling of the picked *P*-wave traveltimes. Previous works assumed the picked *P*-phases as the wave propagation from one receiver to another (e.g. N. Nakata et al. 2015, 2016; A. Riahi et al. 2021, 2023). However, in the Chémery data set, it is highly likely that the picked *P* wave corresponds to the propagation delay of a head wave. Its apparent velocity is close to 4–5 km s^{-1}, which could be consistent with either the fast carbonate layer at a depth of around 850 m or the bedrock (see velocity model cross-section in Fig. 2). A key element to distinguish between a true Green's function reconstruction and a non-physical head wave is to inspect the intercept of the time-offset relation. We show that a zero intercept indicates a head wave (Fig. 7c), while a positive intercept indicates a true

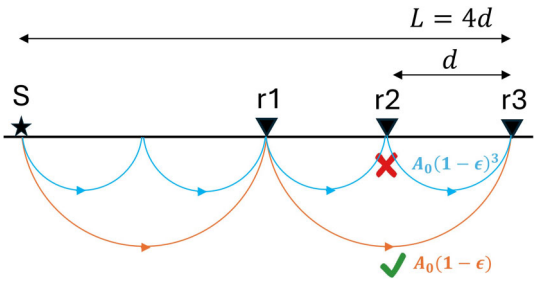


Figure 10. Simple example of diving wave reconstruction by a receiver network r_1, r_2, r_3 from ambient noise sources located in S. In this case, the amplitude ratio of the reconstructed diving wave for the long offset couple r_1–r_3 versus the short offset couple r_2–r_3 is $(1-\epsilon)^{-2}$ (as the reconstructed diving wave amplitude is proportional to $(1-\epsilon)^{L/d-1}$, with ϵ the relative amplitude loss per 'bounce').

Green's function reconstruction (Fig. 7b). The basic concept behind this check is that the apparent velocity should be higher at longer offsets, where the diving waves sample deeper and faster layers.

Beyond the practical check of whether the picked traveltimes are those desired (i.e. associated with the true Green's function), an important issue is to understand which contexts and survey designs are favourable to proper Green's function reconstruction. We believe that a key element is the ratio between the array size and the distance from the network to the dominant noise sources. The reconstruction of diving waves (i.e. the true Green's function) requires the ambient wavefield to contain waves propagating along rays bouncing from the virtual source to the receiver, as illustrated in Fig. 10. On the other hand, ambient noise sources on Earth are located exclusively at the surface (D.E. McNamara & R.P. Buland 2004). This means that the full propagation ray from the real source to the virtual source must be a succession of k identical 'bounces' (Fig. 10). Assuming a relative amplitude loss of ϵ per 'bounce' (accounting for both the reflection coefficient at the free surface and the phase conversions), the amplitude of the 'useful' part of the wavefield will decrease as $(1-\epsilon)^{k-1}$. For a given virtual source–receiver offset d, the number of 'bounces' is of the order of $k \approx L/d$, with L being the distance from the virtual source to the physical noise sources (Fig. 10). This is likely why we only detect non-physical head waves in Chémery: while our maximum offsets are around 10 km, the closest noise sources at 1 Hz are located at shallow depth in the Atlantic Ocean, which is hundreds of kilometres away. Previous studies on the Chémery site showed through beamforming that local anthropogenic sources were dominantly generating surface waves (M. Peruzzetto et al. 2018; C. El Khoury et al. 2024). In contrast, the successful application case of body wave tomography by N. Nakata et al. (2015) used an array located immediately at the ocean coast, close to physical body wave sources, or by A. Riahi et al. (2023) extracted the body wave originally produced by dams next to the array. So, generally, to have a dominant diving P-wave retrieval, small arrays with low interstation spacing, such as the Chémery network, require preferably nearby high-frequency sources (short wavelength) in certain locations. In contrast, larger arrays can make use of lower-frequency energy from more distant sources, which attenuates less and therefore has a higher chance of producing recordable diving-wave energy and masking the non-physical head waves.

Besides, the limited days of operation in the Chémery dataset decreased the probability of recording the strong diving wave. Since in the Chémery data set, the Atlantic Ocean was a strong noise source and continuously generated seismic energy, its contribution is amplified during the stacking process. To extract the diving wave, a strong enough body wave noise source is needed, preferably located in certain nearby positions relative to the array (Fig. 10). However, in the Chémery data set, no other potential body wave sources were detected in nearby locations. In the meantime, the instruments were moved during the experiment (Section 2), so most sites recorded no more than 2 d of ambient noise. This short duration reduces the chance of recording enough diving-wave energy from unknown, weak, and possibly not continuous (i.e. episodic) local sources against the Atlantic Ocean. These factors together explain the predominance of non-physical head waves in our observations.

5.4 Imaging potential with head wave time delays

The advantage of diving waves over head waves is that the areas of the subsurface sampled by the rays are more specific to the virtual source–receiver pair's geometry. For diving waves, the ray penetration depth increases with the station spacing, making it possible to resolve the velocity model at different depths, while the investigation depth is limited by the maximum offset of the network. For head waves, the maximum ray penetration depth is always the depth of the critical refractor, and the ray path above the critical refractor does not depend on the offset. This means that the velocity cannot be resolved as a function of depth. This is the sense of eq. (1): the only information that can be extracted from the arrival times are:

(1) The lateral variations in cumulated traveltimes along the (subvertical) wavepath above the critical refractor.

(2) The lateral variations in critical refractor velocity, if s_0 is made laterally variable.

We performed a synthetic test of 'tomography' based on eq. (1) applied to traveltimes modelled with the eikonal solver (M. Noble et al. 2014) as described in Section 4.2. We inverted only for the cumulated traveltimes T_i, as s_0 of the critical refractor is uniform in the velocity model. The result is displayed in Fig. 11. Qualitatively, we can identify the location of the anticline structure outlined by the negative traveltime anomaly of up to 0.02 s. However, in the real Chémery data, this type of inversion did not produce any consistent result as the noise in the traveltime picking was typically above 0.02 s, in the range of around $\pm$0.05 s. Thus, while tracking the propagation of a head wave with ambient-noise cross-correlations can still provide some 2-D information on lateral discontinuities in the subsurface (without vertical discrimination), the analyst should carefully compare the expected precision of the arrival time picking (taking the frequency band of analysis into account) to the expected effects of lateral geological heterogeneities/targets.

Therefore, the term "non-physical" head wave might be misleading and deserves clarification. Although it reveals the limitations and restricted validity of Green's function reconstruction approach, this signal reflects the traveltime difference between different stations (eq. 1). Thus, the so-called 'non-physical' signal is a physically meaningful observation that may carry medium information, provided that an appropriate corresponding tomography kernel is developed. Some studies have shown that signals falling outside the conventional Green's function interpretation can still be used to extract structural information about the medium (e.g., S. Wang & H. Tkalčić 2020a, b).

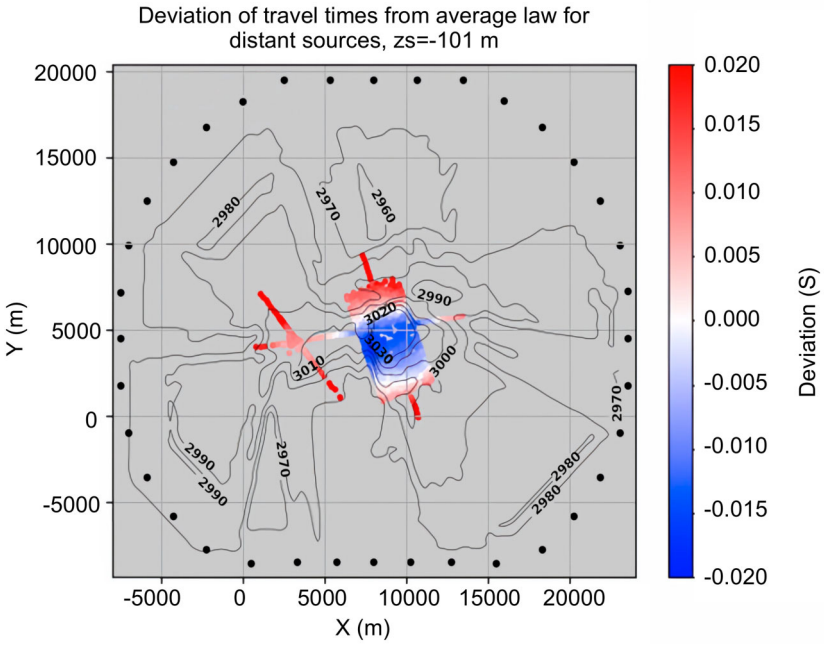


Figure 11. Site-effect map derived from inverting simulated travel-times with eq. (1).

6 CONCLUSION

In this study, we proposed a new polarization-based approach to retrieve the P-wave from ambient noise. While combining the ZR and RZ components enhances body-wave energy, we also showed that P-wave polarity depends on the causal or actual location of the body wave noise source, which is important for accurate *P*-phase retrieval. We processed continuous seismic three-component data from a dense local broad-band network of kilometric extent between 1 and 4.5 Hz to retrieve the body wave component of the EGF. The *P* arrival was identified in the stacked common-offset gather; other arrivals were suppressed with a time-domain taper, and we picked absolute arrival times on a subset of traces with clear *P* arrivals.

Analysis of the picked times indicated that they likely corresponded to a non-physical head wave from virtual sources to virtual receivers. This conclusion is supported by numerical simulations performed with a 3-D eikonal solver within a realistic velocity model of the site. A key element in drawing this conclusion is the nearly zero intercept in the time-offset law of the picked times, whereas a positive intercept would be expected for diving *P* waves traveling from virtual sources to receivers.

Although non-physical head waves can, in principle, map major lateral heterogeneities above the critical refractor as a time-delay map (without depth resolution), this was not feasible here. The medium is nearly 1-D, and the time delay from the slight anticline (100 m) is below the traveltime uncertainty.

The predominance of non-physical head waves in this data set is primarily due to the large distance from the network to the closest ambient noise sources, relative to the typical offsets used within the network. We suggest that passive reconstruction of diving body waves using cross-correlation methods requires a network aperture that is broadly comparable to the distance to the dominant physical body-wave sources at the surface. At higher frequencies, nearby sources located at favourable positions are needed, whereas at lower frequencies, more distant sources require a correspondingly larger array aperture. A dedicated numerical study would be helpful to confirm this claim.

ACKNOWLEDGMENTS

Numerical computations were partly performed on the S-CAPAD/DANTE platform, IPGP, France. The authors acknowledge the financial support of the French government and BPI France in the framework of France Relance 2020–2023 for funding the post-doctoral contract of Ali Riahi in collaboration with Storengy (a company of Engie). The authors also acknowledge Storengy for granting access to the data set.

SUPPORTING INFORMATION

Supplementary data are available at *GJIRAS* online.

Figure S1. The pattern of daily movement of the seismic instruments to the new locations, illustrated by the filled circles.
Figure S2. Examples of the coherent retrieved *P*-phase arrivals for both positive (a and b) and negative (c and d) polarization. The vertical line shows the picked phases. The *Y*-axes are the trace number, sorted by offset. The number of the trace numbers corresponds to Fig. 6.
Figure S3. Azimuthal bin-stacking for different directions: (a) 10–30°, (b) 30–50°, (c) 50–70°, (d) 70–90°, (e) 90–110°, (f) 110–130°, (g) 130–150°, and (h) 150–170° and (i) 170–190°. The number of station couples for each offset-azimuth is shown on the right side as a bar chart.
Figure S4. Daily bin-stacking.

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.

DATA AVAILABILITY

The underlying data of this paper cannot be shared publicly due to internal Storengy policies. The data will be shared on reasonable request to Storengy. The estimated cross-correlations supporting the results of this study have been deposited in the Zenodo repository and are publicly available at DOI: 10.5281/zenodo.17781397.

REFERENCES

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**(3), 1239–1260.

Bharadwaj, P., Schuster, G., Mallinson, I. & Dai, W., 2012. Theory of super-virtual refraction interferometry: theory of SVI, *Geophys. J. Int.,* **188**(1), 263–273.

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**(1), 346–358.

Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science,* **299,** 547–549.

Carvalho, J., Silveira, G., Dumont, S. & Ramalho, R., 2022. 3d-ambient noise surface wave tomography of fogo volcano, cape verde, *J. Volc. Geotherm. Res.,* **432,** 107 702.

Derode, A., Larose, E., Campillo, M. & Fink, M., 2003. How to estimate the Green's function of a heterogeneous medium between two passive sensors? Application to acoustic waves, *Appl. Phys. Lett.,* **83**(15), 3054–3056.

El Khoury, C., Kazantsev, A., Kula, D., Dartois, A. & Chauris, H., 2024. The influence of an anticline structure on ambient noise spectral anomalies at an underground gas storage, *Geophys. J. Int.,* **237**(2), 1061–1078.

Forghani, F. & Snieder, R., 2010. Underestimation of body waves and feasibility of surface-wave reconstruction by seismic interferometry, *Leading Edge,* **29**(7), 790–794.

Hagedoorn, J.G., 1959. The plus–minus method of interpreting seismic refraction sections, *Geophys. Prospect.,* **7**(2), 158–182.

Hanafy, S. & Schuster, G.T., 2017. Parsimonious refraction interferometry and tomography, *Geophys. J. Int.,* **209**(2), 695–712.

Herrmann, R.B., 2013. Computer programs in seismology: an evolving tool for instruction and research, *Seism. Res. Lett.,* **84**(6), 1081–1088.

Köhler, A., Maupin, V. & Balling, N., 2015. Surface wave tomography across the Sorgenfrei–Tornquist zone, SW Scandinavia, using ambient noise and earthquake data, *Geophys. J. Int.,* **203**(1), 284–311.

Larose, E., 2006. Mesoscopics of ultrasound and seismic waves: application to passive imaging, *Ann. Phys.,* **31**(3), 1–126.

Lehujeur, M., Vergne, J., Schmittbuhl, J., Zigone, D. & Le Chenadec, A., 2018. Reservoir imaging using ambient noise correlation from a dense seismic network, *J. Geophys. Res.: Solid Earth,* **123**(8), 6671–6686.

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**(6), 3011–3017.

McNamara, D.E. & Buland, R.P., 2004. Ambient noise levels in the continental united states, *Bull. seism. Soc. Am.,* **94**(4), 1517–1527.

Mikesell, D. & van Wijk, K., 2011. Seismic refraction interferometry with a semblance analysis on the crosscorrelation gather, *Geophysics,* **76**(5), SA77–SA82.

Nakata, N. & Nishida, K., 2019. *Body Wave Exploration,* Cambridge University Press.

Nakata, N., Snieder, R., Tsuji, T., Larner, K. & Matsuoka, T., 2011. Shear wave imaging from traffic noise using seismic interferometry by cross-coherence, *Geophysics,* **76**(6), SA97–SA106.

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.: Solid Earth,* **120**(2), 1159–1173.

Nakata, N., Boué, P., Brenguier, F., Roux, P., Ferrazzini, V. & Campillo, M., 2016. Body and surface wave reconstruction from seismic noise correlations between arrays at Piton de la Fournaise volcano, *Geophys. Res. Lett.,* **43**(3), 1047–1054.

Nishida, K., Kawakatsu, H. & Obara, K., 2008. Three-dimensional crustal S wave velocity structure in Japan using microseismic data recorded by hi-net tiltmeters, *J. Geophys. Res.: Solid Earth,* **113,** B10302

Noble, M., Gesret, A. & Belayouni, N., 2014. Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media, *Geophys. J. Int.,* **199**(3), 1572–1585.

Obermann, A., Lupi, M., Mordret, A., Jakobsdóttir, S.S. & Miller, S.A., 2016. 3D-ambient noise rayleigh wave tomography of Snæfellsjökull volcano, Iceland, *J. Volc. Geotherm. Res.,* **317,** 42–52.

Peruzzetto, M., Kazantsev, A., Luu, K., Métaxian, J.-P., Huguet, F. & Chauris, H., 2018. Broad-band ambient noise characterization by joint use of cross-correlation and music algorithm, *Geophys. J. Int.,* **215**(2), 760–779.

Riahi, A., Shomali, Z.-H., Obermann, A. & Kamayestani, A., 2021. Simultaneous retrieval of body and surface waves in the dehdasht area, iran, from the seismic ambient field and the observation of spurious artefacts, *Geophys. J. Int.,* **227**(2), 1193–1203.

Riahi, A., Shomali, Z.-H., Obermann, A. & Kamayestani, A., 2023. 3-D body-wave tomography from the seismic ambient noise recorded by a dense array in the dehdasht area, iran, *Geophys. J. Int.,* **234**(2), 1332–1341.

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.: Solid Earth,* **118**(10), 5339–5351.

Roux, P., Sabra, K.G., Gerstoft, P., Kuperman, W.A. & Fehler, M.C., 2005. P-waves from cross-correlation of seismic noise, *Geophys. Res. Lett.,* **32**(19), L19303.

Ruigrok, E., 2014. Receiver-pair seismic interferometry applied to body-wave usarray data, *Geophys. J. Int.,* **198**(2), 895–905.

Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A. & Fehler, M.C., 2005. Surface wave tomography from microseisms in southern california, *Geophys. Res. Lett.,* **32**(14), L14311.

Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.,* **31**(7), L07614.

Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science,* **307**(5715), 1615–1618.

Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase, *Phys. Rev. E,* **69**(4), 046610.

Takagi, R., Nakahara, H., Kono, T. & Okada, T., 2014. Separating body and rayleigh waves with cross terms of the cross-correlation tensor of ambient noise, *J. Geophys. Res.: Solid Earth,* **119**(3), 2005–2018.

Telford, W.M., Geldart, L.P. & Sheriff, R.E., 1990. *Applied Geophysics, 2nd edn.,* Cambridge University Press.

Wang, S. & Tkalčić, H., 2020a. Seismic event coda-correlation's formation: implications for global seismology, *Geophys. J. Int.,* **222**(2), 1283–1294.

Wang, S. & Tkalčić, H., 2020b. Seismic event coda-correlation: toward global coda-correlation tomography, *J. Geophys. Res.: Solid Earth,* **125**(4), e2019JB018848.

Wapenaar, K., 2004. Retrieving the elastodynamic Green's Function of an arbitrary inhomogeneous medium by cross correlation, *Phys. Rev. Lett.,* **93,** 254301.

Wapenaar, K. & Fokkema, J., 2006. Green's function representations for seismic interferometry, *Geophysics,* **71**(4), SI33–SI46.

Yudistira, T. *et al.*, 2021. Imaging of a magma system beneath the merapi volcano complex, indonesia, using ambient seismic noise tomography, *Geophys. J. Int.,* **226**(1), 511–523.

Zhang, J. & Gerstoft, P., 2014. Local-scale cross-correlation of seismic noise from the calico fault experiment, *Earthq. Sci.,* **27**(3), 311–318.