Geophysical Journal International

Geophys. J. Int. (2015) **201,** 207–223
GJI Seismology

doi: 10.1093/gji/ggv010

Extraction of weak *PcP* phases using the slant-stacklet transform — I: method and examples

Sergi Ventosa[1] and Barbara Romanowicz[1,2,3]

[1] *Institut de Physique du Globe de Paris, Paris, France. E-mail: ventosa@ipgp.fr*
[2] *Collège de France, Paris, France*
[3] *University of California, Berkeley Seismological Laboratory, Berkeley, CA, USA*

Accepted 2015 January 6. Received 2014 October 16; in original form 2014 June 8

SUMMARY

In order to study fine scale structure of the Earth's deep interior, it is necessary to extract generally weak body wave phases from seismograms that interact with various discontinuities and heterogeneities. The recent deployment of large-scale dense arrays providing high-quality data, in combination with efficient seismic data processing techniques, may provide important and accurate observations over large portions of the globe poorly sampled until now. Major challenges are low signal-to-noise ratios (SNR) and interference with unwanted neighbouring phases. We address these problems by introducing scale-dependent slowness filters that preserve time-space resolution. We combine complex wavelet and slant-stack transforms to obtain the slant-stacklet transform. This is a redundant high-resolution directional wavelet transform with a direction (here slowness) resolution that can be adapted to the signal requirements. To illustrate this approach, we use this expansion to design coherence-driven filters that allow us to obtain clean *PcP* observations (a weak phase often hidden in the coda of the *P* wave), for events with magnitude $M_w > 5.4$ and distances up to 80°. In this context, we then minimize a linear misfit between *P* and *PcP* waveforms to improve the quality of *PcP–P* traveltime measurements as compared to a standard cross-correlation method. This significantly increases both the quantity and the quality of *PcP–P* differential traveltime measurements available for the modelling of structure near the core–mantle boundary. The accuracy of our measurements is limited mainly by the highest frequencies of the signals used and the level of noise. We apply this methodology to two examples of high-quality data from dense arrays located in north America. While focusing here on body-wave separation, the tools we propose are more general and may contribute to enhancing seismic signal observations in global seismology in situations of low SNR and high signal interference.

Key words: Time-series analysis; Wavelet transform; Spatial analysis; Mantle processes; Body waves.

1 INTRODUCTION

Seismologists interested in the Earth's deep structure use measurements of traveltimes and amplitudes of a variety of short-period teleseismic body wave phases that interact with discontinuities in different parts of the mantle and the core. However, typically, a particular phase type can only be used within a limited distance range where it is well isolated from other interfering phases. In order to expand the range of measurements and therefore the sampling of the Earth's interior, one can devise waveform modelling approaches in which observed seismograms are compared to synthetics that already incorporate the effects of such interferences. This is currently the trend in the context of global and continental scale tomography, where progress has recently been achieved, owing to improved means of accurately computing the seismic wavefield in the presence of realistic 3-D heterogeneity using numerical methods (e.g. Fichtner & Igel 2008; Tape *et al.* 2010; Lekić & Romanowicz 2011; Zhu *et al.* 2012; French *et al.* 2013). These approaches are however computationally heavy, increasingly so at short periods, limiting their current applications to periods longer than 30–40 s, and to the modelling of phases with relatively high signal-to-noise ratio (SNR).

Alternatively, array processing techniques have been developed in order to enhance the signal of weak short-period body-wave phases and isolate them from unwanted neighbouring phases. While these techniques have often been developed in the context of communications, radio astronomy and exploration geophysics, among others, they are becoming increasingly relevant to global seismology, owing to the deployment of dense large scale arrays, such as the USArray of Earthscope or Hi-net in Japan.

Most studies use data-independent (using prior information only) array-processing techniques to improve SNR and to isolate coherent signals of weak body-wave phases, through a delay-and-sum (i.e., slant-stack) approach. These intuitive approaches are important to observe weaker signals to improve our understanding of mantle heterogeneities, topography of main discontinuities, and dynamics and evolutions of the core and inner core. For example, the use of phase-weighed stacks (Schimmel & Paulssen 1997) has provided observations on the extremely elusive PKJKP phase (valuable to model inner-core shear velocity) data from the IRIS/IDA and GEOSCOPE networks (Deuss *et al.* 2000), from the Gräfenberg array in Germany (Cao *et al.* 2005) and from Hi-net (Wookey & Helffrich 2008).

In the context of the study of fine scale structure near the core–mantle boundary (CMB), Thomas *et al.* (2002) achieve observations of the weak *PdP* phase (*P* wave reflected at the top of D″, and closer to *P* than *PcP*) with data from the Yellowknife array (19 short-period seismometers deployed in two lines in area of 20 × 20 km) using vespagrams (i.e. slant-stacks) and frequency–wavenumber (f–k) analysis. These *PdP* observations reveal a discontinuity around 241 km above CMB near the Kamchatka peninsula. Rost & Thomas (2010) observe *PdP* and *PuP* (*P* wave reflected at the top of an ultralow velocity zone, i.e. ULVZ) at closer epicentral distances (less than 32°) using the same array to infer the presence or absence of an ULVZ. They combine stacks on the theoretical slowness and backazimuth of *PcP* to form a source array, a limited kind of double-array stacking scheme (Scherbaum *et al.* 1997). Cobden & Thomas (2013) obtain amplitudes and polarities of *PdP* and SdS applying fourth-root slant stacks to small arrays to study the origin of D″ reflections in several regions. Frost *et al.* (2013) present evidence of scattered PKP waves at the eastern edge of the African LLSVP using the Yellowknife array. They use the *F*-statistic for a variety of slownesses and backazimuths to achieve higher resolution and incoherent noise rejection compared to linear slant-stack approaches.

Within the array-processing literature, a classic much better alternative to data-independent approaches are data-dependent (robust) Capon beamformers (e.g. Li *et al.* 2003; Lorenz & Boyd 2005). They define optimum sums in the least-square sense, with the constraint of keeping the signals of interest intact. Robust Capon beamformers seek various strategies to overcome uncertainties in the direction of arrival and in the elements of the array. Alternative convex optimization approaches have recently been proposed, see Gershman *et al.* (2010) for a review. Other techniques used in the geosciences are parametric methods for spectral estimation such as MUSIC (Schmidt 1986) and blind-source-separation methods such as independent components analysis (Hyvärinen & Oja 2000).

A direct application of conventional array-processing approaches is typically hindered by the low directional discrimination of the seismometers (e.g. as opposed to telescopes) and the relatively high interstation distance (in wavelengths) of dense large scale seismic arrays. When array deployments are dense and close to regular, such as the ones used in exploration geophysics to obtain high resolution profiles, tools developed with image processing applications in mind have proven useful in data-independent approaches.

Radon-based methods project signals to sparser transformed domains where they are enhanced and better separated, and if needed, they are back-transformed to the original domain. Traditionally used in exploration for filtering and migration (Yilmaz 2001), in seismology they help, for example for mapping upper-mantle discontinuities (Gu & Sacchi 2009). A variety of multiscale transformations with spatial, directional and frequency selectivity originating from image processing (e.g. Jacques *et al.* 2011) are used in similar applications to help with feature extraction by efficiently concentrating the information of local plane waves in a few coefficients. For example, curvelets, introduced to better capture edge information in images, have been used for denoising (e.g. Ma & Plonka 2010) among other applications.

The short time duration of body waves and the uncertainties in their instantaneous slowness, especially along their codas, hinders optimal data-dependent approaches. Since our objective is to create simple and flexible designs with a reduced beamformer complexity, we here opt for a data-independent strategy built around the slant-stacklet transform that we introduce in the next section. Originally designed for close to regular 1-D arrays in the context of exploration geophysics (Ventosa *et al.* 2011), we here extend this approach to more irregular large-scale 2-D seismic arrays.

Next, we illustrate the proposed method in an application to the case of teleseismic *P* and *PcP* waves as observed on the USArray. Here *PcP* is a weak phase often hidden in the coda of the *P* wave, and generally difficult to isolate. We also introduce a method to estimate traveltime differences based on the minimization of a linear misfit between the *P* and *PcP* waveforms in order to improve the quality of *PcP–P* differential traveltime measurements compared to the widely used cross-correlation method. We finally apply these tools to obtain clean *PcP–P* observations on two events that are representative of the main practical difficulties, with a special focus on the analysis of the main sources of bias.

Although we focus on body-wave separation, and in particular illustrate our approach for the case of *P* and *PcP*, as observed on USAarray, with the ultimate goal of studying CMB topography, the tools proposed here are generic, and they may prove useful in applications where signal can be isolated or enhanced using high-resolution slowness filters.

2 THE SLANT-STACKLET TRANSFORM

Our purpose is to overcome the challenge in the observation of seismic waves with low SNR and in the presence of interference from other signals. Here noise is a random process while interference (i.e. other signals or coherent-noise) is not. We conventionally deal with a low SNR through stacking, in which case the SNR improves by $\sqrt{N}$ if the signals are identical across all stations and their noises are independent, Gaussian and of equal power. Observations in low signal-to-interference ratio (SIR) environments are more challenging. In the absence of noise, N fairly well-located stations suffice to separate N uncorrelated signals coming from N different directions. However, their often finite SNR, close direction of arrival, and high cross-correlation hinder their separation and increase the number of stations required to minimize their cross-interference.

Body waves are finite-duration broad-band signals with distinct polarization, traveltime, slowness and waveform. We exploit the fact that teleseismic waves can be locally approximated by plane waves, in order to separate them according to their instantaneous slowness. The slant-stacklet transform approach we introduce here is closer in spirit to other multiscale directional frames developed in image processing in the context of wavelet analysis (e.g. Jacques *et al.* 2011). Our main distinctive design goal is to promote flexibility in the choice of optimal-resolution compromises in analysis operators that expand wavefields in slowness. Scale diversity helps in controlling slowness resolution in different frequency bands. Compared to non-redundant approaches, redundant expansions provide a much greater flexibility, in return for a higher computational cost, and often lead to algorithms able to retrieve signals with a lower SNR

and SIR. This redundancy enables the slant-stacklet transform to achieve extremely-adaptable resolution compromises, and reduces the complexity of filters with a much higher slowness resolution. In particular, we distinguish two adaptive-filtering schemes depending on the cross-interference among signals in the transformed domain. In Section 2.2, cross-interferences can be considered negligible and a direct selection thus suffices; while in Section 2.3 they are not, and more involved synthesis operations are required. The design of the actual filters is often a tedious task due to the non-stationarity of the seismic waves and the high dimensionality of the transformed domain. In Section 2.4, we overcome this practical drawback by introducing signal-adapted filters based on instantaneous slowness measurements.

2.1 Analysis

The slant-stacklet transform is conceptually a combination of local slant-stack transform (LSST), for example Ventosa *et al.* (2012b), and the continuous wavelet transform (CWT). We write the local slant-stack decomposition of a wavefield $u(t, x)$ as a weighted sum around a location x_c along a set of wave fronts with slowness p:

$$v_{p,x_c}(\tau) = \int_{-\infty}^{\infty} a(x - x_c) u[\tau + p^T(x - x_c), x]\, \mathrm{d}x. \tag{1}$$

Here p^T denote the transpose p vector. The time-space trajectory of the wave front with a slowness p is $t = \tau + p^T(x - x_c)$, where τ is delay, $x - x_c$ distance between locations x and x_c, and x, x_c, $p \in \mathbb{R}^N$. The spatial-weighting function $a(x)$ is smooth and has unit area.

The CWT of $u(t, x)$ along the time axis and is defined as:

$$Wu_x(\tau, \lambda) = \int_{-\infty}^{\infty} u(t, x) \frac{1}{\sqrt{\lambda}} \psi^* \left(\frac{t - \tau}{\lambda} \right) \mathrm{d}t, \tag{2}$$

where λ is scale and $\psi(t)$ is the mother wavelet, conventionally, a zero-mean unit-energy function. $\psi^*(t)$ denotes the complex conjugate of $\psi(t)$. Then the slowness expansion of $Wu_x(\tau, \lambda)$ equivalent to eq. (1) is:

$$Wv_{p,x_c}(\tau, \lambda) = \int_{-\infty}^{\infty} a_\lambda(x - x_c) Wu_x[\tau + p^T(x - x_c), \lambda]\, \mathrm{d}x, \tag{3}$$

where $a(x)$ may be function of λ and x_c.

CWT is translation invariant [i.e. a translation on $u(t)$ results in a translation on $Wu(\tau, \lambda)$]. This property enables us to include the delay component that varies with distance, $p^T(x - x_c)$ in eq. (2), in the mother wavelet as an extra delay term,

$$Wv_{p,x_c}(\tau, \lambda) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(t, x) \phi_{p,\lambda}(t - \tau, x - x_c)\, \mathrm{d}t\, \mathrm{d}x \tag{4}$$

with

$$\phi_{p,\lambda}(t, x) = a_\lambda(x) \frac{1}{\sqrt{\lambda}} \psi^* \left(\frac{t - p^T x}{\lambda} \right). \tag{5}$$

The family of functions $\phi_{p,\lambda}(t, x)$ controls the resolution of the slant-stacklet expansion in the time, scale, space and slowness axes (Appendix A). The directional mother wavelet $\phi_p(t, x)$ resembles a localized planar wavefield with a slowness p, instead of the 1-D function employed in eq. (2). This wavelet oscillates in the time dimension and remains smooth along the wave front with slowness p.

In summary, the slant-stacklet transform expands a time-space wavefield into a time-scale space-slowness domain, where space and slowness are real vectors with no more than three components.

2.2 Filtering and synthesis: separable signals

A synthesis operator derived from the frame inequality eq. (B1) is the pseudo-inverse or dual frame. The analysis frame and its dual frame are equal when the frame is tight. When these operators are used, filtering operations, for example denoising, are conventionally perform using threshold functions (Mallat 2009). Our interest is separating signals that can be modelled locally as plane waves, from other signals and noise. Instead of following the pseudo-inverse approach, we opt to exploit previous knowledge on these signals to be able to define synthesis operators and high resolution slowness filters specifically adapted for them. In particular, we define punctual synthesis operators that do not use information from neighbouring stations, in contrast to the dual frame.

Let u denote R overlapping plane waves of slowness q_r,

$$u(t, x) = \sum_{r=1}^{R} u_r \left(t - q_r^T x \right). \tag{6}$$

and its slant-stacklet transform $Wv_{p,x_c}(\tau, \lambda)$. The most simple inverse operation we can write when these signals are well separated in slowness is probably the 'lazy inverse'. In the previous section, we introduce the slant-stacklet transform, eq. (4), as a combination of the LSST, eq. (1) and the CWT eq. (2). Therefore, when signals are clearly separated in slowness in the transform domain (equivalently, with the LSST alone), we can select their components at each slowness and space locations (e.g. $p = q_r$ and $x_c = x$ to estimate u_r), and apply the inverse wavelet transform to the result. This is:

$$u_r(t, x) = \frac{1}{C_\psi} \int_0^{\infty} \int_{-\infty}^{\infty} Wu_{q_r,x}(\tau, \lambda) \frac{1}{\sqrt{\lambda}} \psi \left(\frac{t - \tau}{\lambda} \right) \mathrm{d}\tau \frac{\mathrm{d}\lambda}{\lambda^2} \tag{7}$$

where

$$C_\psi = \int_0^{\infty} |\hat{\psi}(\omega)|^2 \frac{\mathrm{d}\omega}{\omega} < \infty \tag{8}$$

is called the wavelet admissibility condition. Although sub-optimal, this constitutes a fast and trivial solution that suffices in applications where spatial-resolution constraints allow the use of an array aperture large enough to clearly separate each signal in the transformed domain.

The filter above applied to the R overlapping plane waves of eq. (6) writes in frequency:

$$\hat{v}_p(\omega, x_c) = \sum_{r=1}^{R} \hat{h}(\omega, q_r - p) \hat{u}_r(\omega) \mathrm{e}^{-j x_c q_r \omega}, \tag{9}$$

where the transfer function of the filter is:

$$\hat{h}(\omega, q_r - p) = \frac{1}{C_\psi} \int_0^{\infty} \hat{a}_\lambda[(q_r - p)^T \omega] |\hat{\psi}(\lambda\omega)|^2 \frac{\mathrm{d}\lambda}{\lambda}, \tag{10}$$

where v_p is the estimated signal. This filter preserves the spectrum of signals with slowness q_r because we define $\hat{a}(0) = 1$ and thus $\hat{h}(\omega, q_r - p)|_{p=q_r} = 1$. However, signals with close slowness are not fully rejected because $\hat{a}$ is not impulsive, that is the aperture of the array is finite and hence also its slowness resolution.

The main parameter controlling slowness resolution is the scaling function of a with λ, $f(\lambda)$ (Appendix A). Two particular cases are of interest, no scaling and scaling proportional to λ.

2.2.1 No spatial-weighting scaling

When no scaling is used, $a_\lambda(x) = a(x)$, $\hat{h}(\omega, q_r - p) = \hat{a}[(q_r - p)^T \omega]$ equals the transfer function of the *p–f* transform (Forbriger 2003). This is equivalent to the LSST (Ventosa *et al.*

2012b), where $\hat{a}[(q_r - p)^T\omega]|_{p=q_r} = 1$, and for any other slowness $\hat{a}_\lambda[(q_r - p)^T\omega]|_{p\neq q_r}$ is a low-pass filter with a cut-off frequency proportional to $1/(q_r - p)$.

2.2.2 Spatial-weighting scaling proportional to λ

The transfer function of the filter is independent of frequency when the scaling of the spatial-weighting function is proportional to scale, $a_\lambda(x) = \lambda^{-1}a(\lambda^{-1}x)$. In this case, eq. (10) writes,

$$\hat{h}(\omega, q_r - p) = \frac{1}{C_\psi}\int_0^\infty \hat{a}[(q_r - p)^T\lambda\omega]|\hat{\psi}(\lambda\omega)|^2\frac{\mathrm{d}\lambda}{\lambda}. \tag{11}$$

Applying the change of variables $b = \lambda\omega$, $\mathrm{d}b = \omega\,\mathrm{d}\lambda$, the frequency dependence of the transfer function vanishes:

$$h(q_r - p) = \frac{1}{C_\psi}\int_0^\infty \hat{\phi}^*_{q_r-p}(b)\hat{\psi}(b)\frac{\mathrm{d}b}{b} = \frac{C_{\phi_{q_r-p},\psi}}{C_\psi}, \tag{12}$$

where we define $\hat{\phi}^*_{q_r-p}(\omega) = \hat{a}[(q_r - p)^T\omega]\hat{\psi}^*(\omega)$ to clearly distinguish the analysis and synthesis operators, ϕ and ψ respectively. Note that $C_{\phi_{q_r-p},\psi}$ equals the wavelet admissibility coefficient of a wavelet transform using different analysis and synthesis mother wavelets. This reduces to C_ψ, eq. (8), when $p = q_r$.

2.3 Filtering and synthesis: non-separable signals

Apart from requirements of optimal spatial-resolution, slowness and waveform variations along signal wave fronts are major constraints on the maximum aperture of the weighting function due to the plane wave assumption. Limited apertures often impede defining a slant-stacklet expansion capable of fully separating signals in the transformed domain, and consequently hinders the design of filters to isolate them. We can still separate close signals in these scenarios, despite their cross-interference, as long as we are able to distinguish their individual maxima.

We improve the slowness selectivity of filters built with the lazy inverse significantly, keeping the framework of the slant-stacklet expansion, by modelling the cross-interference of plane waves. We see in eq. (12) that if $a_\lambda(x)$ scales proportionally with λ, $f(\lambda) = \lambda$, a wavefield composed by R overlapping plane wave of slowness q_r expands in slowness as:

$$v_s(t, x_c) = \sum_{r=1}^{R} u_r(t - q_r x_c)h(q_r - p_s). \tag{13}$$

where p_s is discretized slowness and $s \in \mathbb{Z}$ the slowness index (Appendix B). This allows us to improve the slowness resolution of filters further from the slant-stacklet transform limitations, through a deconvolution operation.

In general, cross-interference in slowness changes with scale when $f(\lambda) \neq \lambda$ and, to a lower extent, with delay. We consider in the following the delay dependency negligible in practice, to focus on the scale dependency. As a result we approximate the wavefield expansion in slowness and scale as:

$$W v_{s,x_c}(\tau, \lambda) \simeq \sum_{r=1}^{R} W u_r(\tau - q_r x_c, \lambda)h(\lambda, q_r - p_s), \tag{14}$$

modelling the cross terms (including the synthesis operation) in slowness at each scale as:

$$h(\lambda, q_r - p_s) = \frac{1}{\lambda^2 C_\psi}\int_{-\infty}^{\infty} \phi^*_\lambda(\tau, q_r - p_s)\psi_\lambda(\tau)\,\mathrm{d}\tau. \tag{15}$$

Similar to (eq. 12), ϕ and ψ are the analysis and synthesis operators, respectively. However here the scaling of ϕ is not proportional to λ and thus eq. (15) does not simplifies.

Filters based on this cross-interference estimation can be written as weighted sums in slowness,

$$W y_{x_c}(\tau, \lambda) = \sum_s f(\tau, \lambda, p_s) W v_{s,x_c}(\tau, \lambda), \tag{16}$$

where f denotes the filter to design, and $W y_{x_c}(\tau, \lambda)$ the filtered signal in the time-space-scale domain. We then use the inverse wavelet transform to obtain $y(t, x_c)$ back in the time-space domain.

Let us assume the slownesses of R signals are different and known, and let $g_r(\tau, \lambda)$ be the imposed gain at each slowness, which may vary according to τ, λ and x_c. Then any filter f satisfying eq. (16) has to satisfy the following R equations,

$$\sum_{s=0}^{S-1} f(\tau, \lambda, p_s)h(\lambda, q_r - p_s) = g_r(\tau, \lambda) \quad \forall r \in [1, R]. \tag{17}$$

Equivalently, in vector notation,

$$\mathbf{H}(\lambda)\mathbf{f}(\tau, \lambda) = \mathbf{g}(\tau, \lambda) \tag{18}$$

where

$$\mathbf{H} = \begin{bmatrix} h_\lambda(q_1 - p_1) & \cdots & h_\lambda(q_1 - p_s) & \cdots & h_\lambda(q_1 - p_S) \\ \vdots & & \vdots & & \vdots \\ h_\lambda(q_r - p_1) & \cdots & h_\lambda(q_r - p_s) & \cdots & h_\lambda(q_r - p_S) \\ \vdots & & \vdots & & \vdots \\ h_\lambda(q_R - p_1) & \cdots & h_\lambda(q_R - p_s) & \cdots & h_\lambda(q_R - p_S) \end{bmatrix} \tag{19}$$

is a rectangular matrix that varies with λ, $\mathbf{f}$ a vector of filter coefficients that are functions of τ and λ, and $\mathbf{g}$ a vector of gain constraints in slowness.

The number of constraints R is in practice much smaller than the number of slowness components S. We thus have an underdetermined system of equations with an infinite number of solutions. The stability of these solutions for a given $h_\lambda(p)$ is directly determined by the slowness differences of contrasting gain conditions on $\mathbf{g}$. As these slowness differences get smaller, $\mathbf{H}$ becomes closer to singular, and hence smaller inaccuracies on $\mathbf{H}$ lead to larger errors on the filter solution. The ability to separate two close signals is limited ultimately by the approximation made in modelling the cross terms in eq. (15). Next, we focus on two of these solutions, the minimum noise and the minimum interference solutions.

2.3.1 Minimum noise

When the slowness of all signals is well-known, the least-square solution leads to a minimum noise level. The general solution $\mathbf{f}(\tau, \lambda)$ with minimum energy under the R constraints of eq. (18),

$$\min_{\mathbf{f}(\tau,\lambda)} \mathbf{f}^T(\tau, \lambda)\mathbf{f}(\tau, \lambda) \quad \text{with} \quad \mathbf{H}(\lambda)\mathbf{f}(\tau, \lambda) = \mathbf{g}(\tau, \lambda), \tag{20}$$

is a function of the noise distribution (see, e.g. Kay 1993). The particular solution for white Gaussian noise is the pseudo-inverse:

$$\mathbf{f}(\tau, \lambda) = \left[\mathbf{H}^T(\lambda)\mathbf{H}(\lambda)\right]^{-1}\mathbf{H}^T(\lambda)\mathbf{g}(\tau, \lambda). \tag{21}$$

As long as the system of equations defined in eq. (20) is not close to singular, this solution leads to smooth filters that employ all the slowness components available.

2.3.2 *Minimum interference*

In the previous solution, we implicitly classified all signals within a seismic wavefield into two clearly separate groups: coherent signals and non-coherent signals or noise. We have prior information on the number and instantaneous slowness of the first ones, and on the distribution of the second ones. In scenarios where this separation is clear, we measure signals with an optimal SNR; however, seismic wavefields are usually very complex, and this separation is frequently not very clear.

Among relatively strong coherent signals and background noise, there exist a large number of much weaker and less coherent signals, for example those due to scattering. This set of signals consists of many waves with very complex travel paths, traditionally referred to as coherent noise; but also includes waves with long travel paths and/or high attenuation. We consequently interpret coherent noise as a group of coherent signals for which the exact number and individual slownesses are unknown.

A simple method to reduce the effects of coherent noise is to focus exclusively on the slowness components where coherent signals are expected to be stronger. With this constraint, the system of equations defined in eq. (17) reduces to a full system of equations with, typically, a single solution,

$$\mathbf{f}(\tau,\lambda) = \mathbf{H_R}^{-1}(\lambda)\mathbf{g}(\tau,\lambda), \tag{22}$$

where $\mathbf{H_R}$ is formed by the subset of columns of $\mathbf{H}$ satisfying $p_s = q_r$ with $r \in [1, R]$. This is

$$\mathbf{H_R} = \begin{bmatrix} 1 & \cdots & h_\lambda(q_1-q_r) & \cdots & h_\lambda(q_1-q_R) \\ \vdots & \ddots & \vdots & & \vdots \\ h_\lambda(q_r-q_1) & \cdots & 1 & \cdots & h_\lambda(q_r-q_R) \\ \vdots & & \vdots & \ddots & \vdots \\ h_\lambda(q_R-q_1) & \cdots & h_\lambda(q_R-q_r) & \cdots & 1 \end{bmatrix}, \tag{23}$$

where $h_\lambda(0) = 1$. Note that this is equivalent to the lazy inverse shown in Section 2.2 for negligible off-diagonal elements.

2.4 Signal adaptation

To design minimum noise or interference filters, previous knowledge about the slowness of the signals is required to set $\mathbf{g}(\tau,\lambda)$ appropriately. When prior information is not accurate, we can estimate slowness using any method, for example Bear & Pavlis (1997). A trivial approach is to search for energy maxima of the slant-stacklet expansion close to the expected slowness. This approach often poses problems because of the wide dynamic range of the energy of seismic signals and of variations of the noise level that hinder the definition of thresholds able to distinguish signals from noise. Alternatively, we can seek solutions more closely related to SNR or to waveform similarity along the wave front, while reducing the dynamic range of the observations. Coherence estimators are useful in this context. They measure similarity of two or more signals, with values within the range of [0, 1] or [−1, 1], typically.

In particular, to assess slowness, we search for coherency maxima of plane waves in the transformed domain. We basically replace the spatial-weighting function $a(x)$ used in the slant-stacklet expansion, eq. (4), with a coherence estimator. Then, we search for coherency maxima close to the expected theoretical values. With this approach, we achieve slowness maxima defined directly by the quality or clearness of the signal, and not by its local power, much more polluted by noise.

To measure instantaneous slowness, we opt for coherence estimators for analytic signals because they are independent of the instantaneous phase, see appendix of Taner *et al.* (1979). Equivalent coherence estimators for real signals require smoothing, and thus sacrifice time-resolution, to avoid zero-crossing uncertainties. Coherence estimators for analytic signals do not have this problem, even when no smoothing is applied, as long as the envelope is not close to zero.

In summary, our master pieces to assess slowness are: (1) the analytic wavelet transform, (2) coherence estimators for analytic signals and (3) algorithms to search and track slowness maxima.

A wavelet transform is called analytic if the negative frequencies of the mother wavelet are zero, $\hat{\psi}_a(\omega) = 0 \quad \forall\omega < 0$, or equivalently, $\mathrm{Real}(\psi_a(t)) = \psi(t)$. The wavelet transform of a wavefield $u(t, x)$ is then:

$$Wu_{p,x_c}(\tau,\lambda,x) = \int_{-\infty}^{\infty} u(t,x)\frac{1}{\sqrt{\lambda}}\psi_a^*\left[\frac{t-\tau-p^T(x-x_c)}{\lambda}\right]\mathrm{d}t, \tag{24}$$

where $\psi_a^*(t - p^T x)$ is the analytic mother wavelet. As in eq. (3), the additional factor $p^T(x - x_c)$ plays the role of an independent delay term that accounts for the relative traveltime difference between x and x_c of a wavefield with slowness p.

A diversity of coherence estimators for analytic signals are available in the literature (e.g. Taner *et al.* 1979; Schimmel & Paulssen 1997). The main differences among them originate in the particular definitions of what is similar and what is not, and the properties of the signal domain chosen. Nonetheless, many of them are intimately related. We define similarity as a measure of how close the waveforms of two or more signals are, independently of their amplitude. We accordingly employ coherence estimators based on normalized cross-correlations or phase stack, instead of energy ratios such as semblance. We specifically apply coherence estimators along x around x_c, with an optional smoothing along τ in low SNR settings.

Let us define $s_i[k] = Wu_{p,x_c}(\tau + k, \lambda, x_c + x_i)$. A natural coherence estimator based on cross-correlations calculates the mean of geometrically normalized cross-correlations between all the pairs at zero delay (this is τ for $s_i[k]$):

$$C_{\mathrm{GNCC}} = \frac{2}{M(M-1)}\sum_{i=1}^{M}\sum_{j>i}^{M}\frac{r_{ij}}{\sqrt{r_{ii}r_{jj}}}, \tag{25}$$

where M is the number of stations, i and j signal indices, r_{ij} cross-correlation between s_i and s_j at zero delay, and r_{ii} and r_{jj} their auto-correlations. $C_{\mathrm{GNCC}} \in [-1/(M-1), 1]$, fully coherent signals attain the maximum value, while totally incoherent signals give zero. The phase-stack coherence estimator (Schimmel & Paulssen 1997) is a much faster alternative that can be expressed as:

$$C_{\mathrm{PS}}^{v} = \sum_k \left|\frac{1}{M}\sum_{i=1}^{M}\frac{s_i[k]}{|s_i[k]|}\right|^v, \tag{26}$$

where M is the number of stations used, $\sum_k$ smoothing along the time dimension, and v a power factor accounting for different means. $C_{\mathrm{PS}} \in [0, 1]$, similarly to C_{GNCC}, $C_{\mathrm{PS}}^{v} = 1$ for fully coherent signals, and $C_{\mathrm{PS}}^{v} = M^{-v}$ for the totally incoherent ones. Both estimators are independent of the relative amplitude of the signals, the key difference between them is in the weighting of each signal in time. C_{GNCC} promotes signals along time according to their envelope, while C_{PS} is totally independent of their instantaneous amplitude.

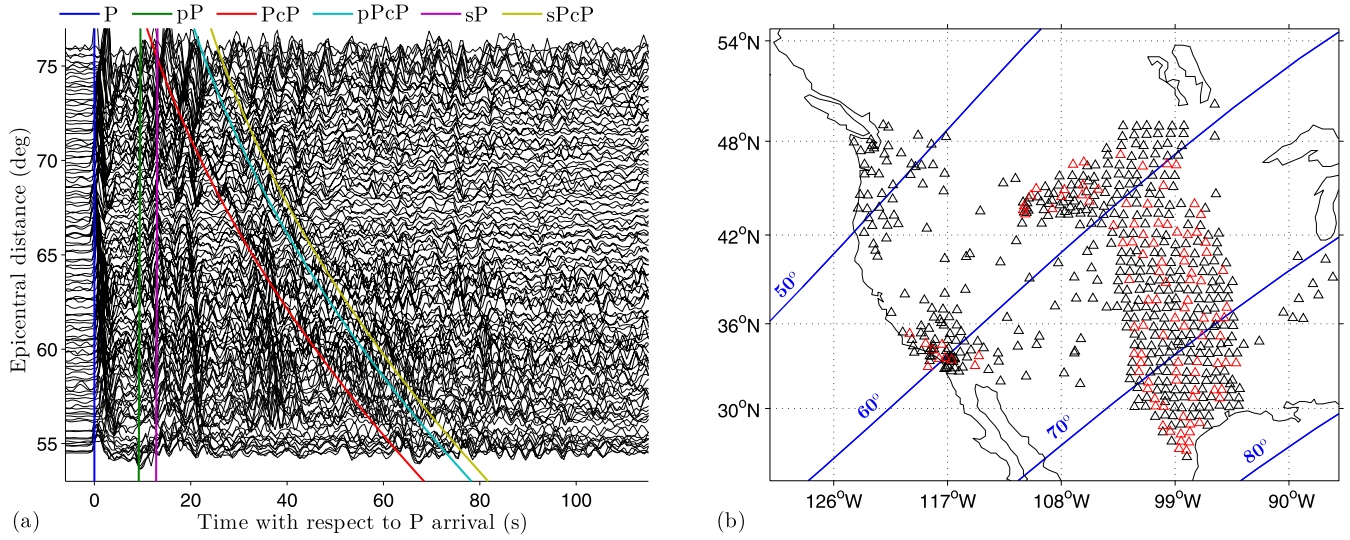

Figure 1. (a) Velocity data from an earthquake off the east coasts of Kamchatka (2010/07/30 3:56:19.2 UTC, $M_w = 6.3$, depth = 30.3 km). Instrument response is removed and data are bandpass filtered from 1 to 60 s. (b) Epicentral distances (blue lines) and location of the stations with good quality signal shown in (a) (red triangles) and not shown (black triangles).

To locate slowness maxima, we first roughly define stop and pass bands in the transformed domain based on previous information on the signals. Then, within each of these bands, we search for the n highest maxima in slowness at each time-scale-space location above a predefined threshold. For stability reasons in the synthesis operation, we impose a minimum slowness separation between maxima of opposite bands, according to the maximum slowness resolution we can achieve after the inverse operation.

The above coherence estimators are more effective with well-separated signals in slowness. As signals get closer in slowness, strong signals hinder the detection of overlapping weaker signals (even when their separation would still be possible), because they behave as a powerful noise that masks weaker signals. In these scenarios, we keep using coherence estimators to locate most of the signals, but we impose gain constraints in $\mathbf{g}(\tau, \lambda)$ according to prior information on the instantaneous slowness of the weakest signals.

3 APPLICATION TO *P* AND *PcP* ISOLATION

Fig. 1 shows raw velocity time-series with the instrument response removed, for a shallow earthquake off the east coast of Kamchatka ($M_w = 6.3$, depth = 30.3 km) observed at a subgroup of 119 stations (BHZ channel) located in north America (mostly from USArray), from a total of 640 stations with good quality signal selected from the IRIS database sampling the CMB under the Gulf of Alaska. For more information on the scale structure in this area (see e.g. Vidale & Benz 1992; Castle & van der Hilst 2000; Rost & Revenaugh 2004). We distinguish the direct P wave a few seconds after its theoretical arrival time. The core-reflected PcP can be distinguished only in distance windows where other signals are weaker, for example from 65° to 70° and less clearly from 55° to 60°. We can also observe a rich group of strong P post-cursors and even some weaker PcP post-cursors. The first main signal after P is probably pP, and that after PcP is probably $pPcP$. Indeed, the radiation pattern for P waves towards USArray is strong for P, PcP and their p depth phases, but very weak for sP and $sPcP$. Later signals may be due to reflections and P-to-s conversions at the Moho and the 410 and 660 km discontinuities in the upper mantle, in view of a time of arrival of about 20 s after P of the first one, and their similarity with signals arriving about 35 s after P and later.

The high energy of all these P post-cursors is the main source of signals that hinder the observation of PcP in this relatively strong event. The waveform of P is short enough to not overlap with PcP until epicentral distances closer to 80°, outside the epicentral distance range sampled. This is an important problem in stronger and shallower events with, frequently, a much longer P coda. Background noise becomes a major problem only for weak events, where PcP is difficult to appreciate directly, but it may still be possible to extract PcP waveforms thanks to their high coherence along their wave fronts.

Two main parameters constrain the minimum aperture needed to separate overlapping signals: (1) their slowness difference and (2) the inter-station distance. The slowness difference between P and PcP decreases as epicentral distance increases, while the station distribution across the array limits the actual minimum and maximum aperture we can achieve.

An event with depth of 30.3 km has a predicted slowness difference for P and PcP, according to AK135, of 2.860 s deg^{-1} at 60°, 1.912 s deg^{-1} at 70° and 1.467 s deg^{-1} at 75°. We thus need a minimum slowness resolution of 3 s deg^{-1} to distinguish two maxima in slowness at 75° with the non-separable signals assumption (e.g. Section 2.3). This is about half the slowness resolution needed when neglecting the cross terms (e.g. Section 2.2), where a full separation of the signals in the transformed domain is required. Although we could separate signals using coarse resolution with accurate slowness information on the signals, we constrain the slowness resolution to twice the slowness difference between P and PcP, to promote the smoothness of the filters in the slowness axis. Smoothness is of importance for stability reasons, or somehow equivalently, to keep the gain of other signals under control at unconstrained slowness.

The minimum aperture is limited by the interstation distance. The station separation for the USArray is approximately 0.6°; however, a constant slowness resolution with period of 3 s deg^{-1} leads to apertures as low as 0.3° at 1 s using uniform station weighting. To

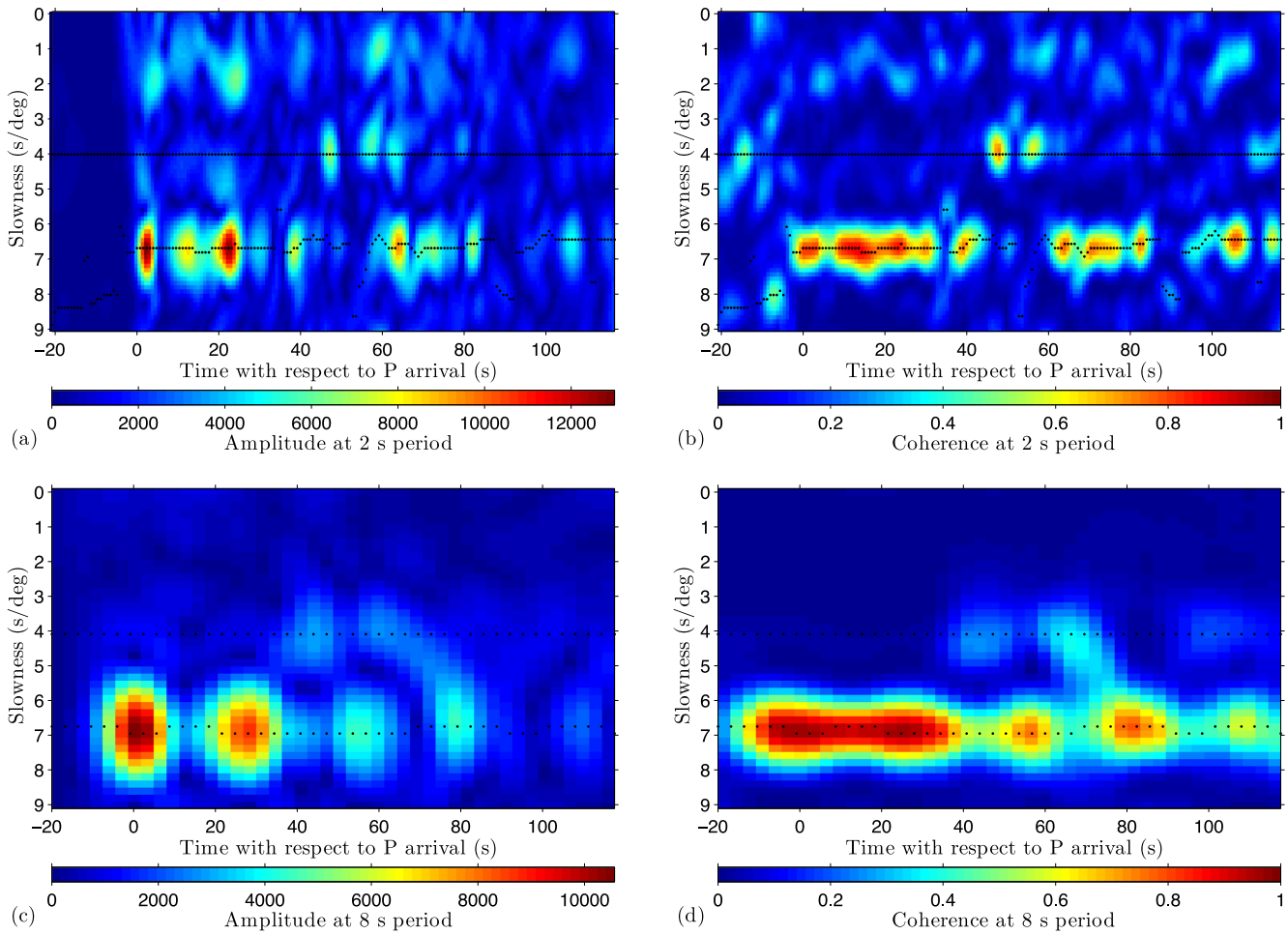

Figure 2. Time-slowness sections at a distance of 60.24° (TA.H27A) for the data shown in Fig. 1. (a) Amplitude and (b) coherence at a period of 2 s. (c) Amplitude and (d) coherence at 8 s. Black dots mark the slowness components where we impose gain constraints on the synthesis operation.

avoid these anomalous apertures, we impose a minimum value of 2° (aperture at 6 s period), wide enough to enclose 8 stations, and a maximum aperture of 6.7° for practical purposes.

Aliasing in slowness may become problematic at short periods. The extreme case occurs when the interstation separation equals their separation in epicentral distance, in the case of USArray this is about 0.6°, which would produce aliasing at $\pm nT/0.6$, where T is period and $n \in \mathbb{N}$. This is at ±3.33° at a period of 2 s. This aliasing is reduced in practice for a variety of reasons, for example the signals have a short duration, arrays are not regularly spaced, and stations have different epicentral distances. However, they become the main source of biases at very short periods. We therefore opt for not using periods shorter than 1 s at this stage.

Fig. 2 shows time-slowness sections of a record section centred around an epicentral distance of 60.24° (Fig. 1) expanded in the time-scale-slowness domain. We choose the complex Morlet wavelet for analysis, because it has optimal time and frequency resolution and it is approximately analytic. Its mother wavelet has the form:

$$\psi(t) = \pi^{-1/4} \mathrm{e}^{-\mathrm{i}\omega_0 t} \mathrm{e}^{-t^2/2}, \tag{27}$$

where t is time and ω_0 is central frequency. A standard choice of $\omega_0 = \pi\sqrt{2/\ln 2}$ makes the amplitude of the side lobes equal to half of the main lobe.

We generate a frame of time-space wavelets in scale and slowness according to Appendix B. The subsampled wavelet frame, eq. (B4), uses four voices, six octaves and a sampling period of $b_0 = 1$. The central frequency of the wavelet with the lowest scale is 1 Hz. The periods analysed then range from 1 s to about 54 s. The sampling in slowness is about 10 times the actual slowness resolution at each scale, with a minimum value of 3 smpl (s deg^{-1})$^{-1}$. We think this high level of redundancy achieves a balance between implementation efficiency and filter flexibility. This balance allows for a relatively fast implementation (still far slower than non-redundant alternatives) and, in return, it simplifies the design of the filter, improves signal adaptation, and makes the configuration extremely flexible. Figs 2(a) and (c) display amplitude measurements at periods of 2 and 8 s, while Figs 2(b) and (d) show the corresponding measurements of coherence using phase-stack, eq. (26), with $v = 2$ and a square window of five samples. This is a length of 4 s at a period of 2 s and 16 s at a period of 8 s. Note that the finite time resolution at individual scales of the symmetric wavelets used produces relatively wide maxima in the time-scale domain, which may extend before the time of arrival of signals at low frequencies. This does not limit the actual time resolution, which is determined by the low scale components, and helps in the filtering.

We use coherence to measure the instantaneous slowness of *P* and its post-cursors. *PcP*, far weaker than *P*, is still usually strong

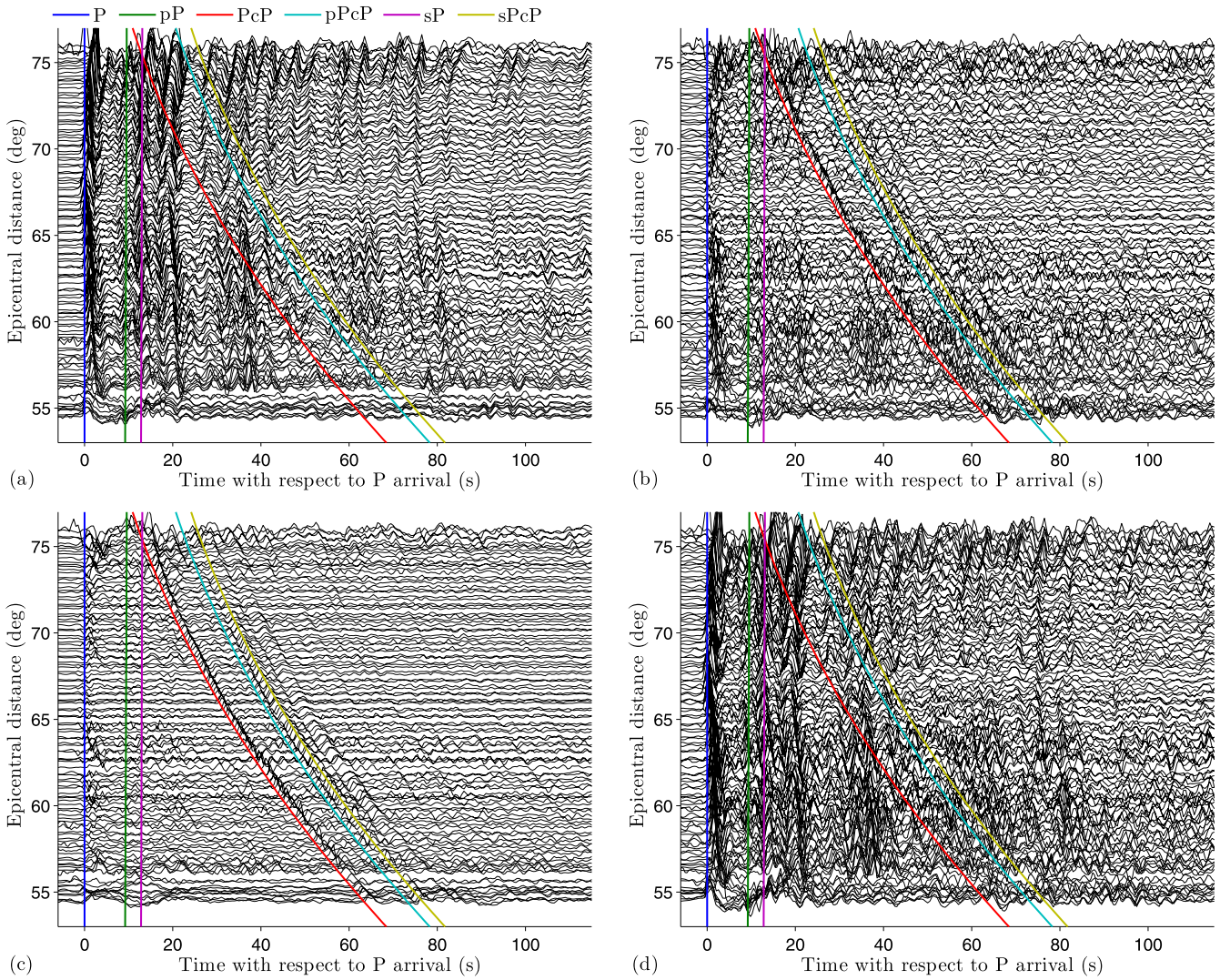

Figure 3. *P* and *PcP* extractions from data shown in Fig. 1. (a) *P* extracted. (b) Data minus the *P* extracted. (c) *PcP* extracted. (d) Data minus the *PcP* extracted. Signals in (a) and (c) are clean but spatially smoothed, while in (b) and (d), they are not smoothed but noiser. All figures use the amplitude normalization used for the raw data.

enough to obtain a robust measurement of its slowness at these periods. We can identify *PcP* in all figures with the first main lobe at slowness of about 4 s deg^{-1} at around 50 s arrival time. However, the use of 1-D model predictions usually provides more accurate information, especially at large epicentral distances where the slownesses of *P* and *PcP* are closer, and *PcP* coherence maxima are masked by the *P* coda. We can additionally appreciate that the two main lobes of Fig. 2(a) have weak side lobes with slowness maxima about 1.5 s deg^{-1} (at the top edge of the figure), produced by aliasing effects.

Fig. 3 presents results after the filtering and synthesis operations using the minimum noise solution, eq. (21). Fig. 3(a) shows *P* and its post-cursors extracted with a gain constraint of 1 at the slowness of *P* and a gain of 0 at the slowness of *PcP*. Fig. 3(b) shows the result after subtracting this result from the raw data (Figs 1–3a). Similarly, Fig. 3(c) shows *PcP* and their post-cursors using the same slowness constraints but with their gains exchanged, and Fig. 3(d) its subtraction from the raw data.

Figs 3(a) and (c) show clean and smooth observations of *P*, *pP*, *PcP*, *pPcP* and other signals. Comparing Fig. 3(c) with the raw data (Fig. 1), we see a dramatic attenuation of *P* and all the post-cursors as well as of the noise level. We can only distinguish a very small remainder of *P* caused mainly by waveform differences and misalignments. In contrast to the raw data, now it is possible to clearly see *PcP* at all epicentral distances, even at distance larger than 70° where *pP* and *P* post-cursors were very strong. Differential observations from Figs 3(b) and (d) are valuable to validate more accurate extracted observations from Figs 3(a) and (c).

An inherent shortcoming that may arise in some applications is due to the smoothing produced by the local stacks. In the *PcP–P* application, the resolution lost at the CMB is not significant until reaching long periods, as scaling is proportional to period for the array aperture and to its square root for the Fresnel zone. To be more precise, the Fresnel zone at 2 s is about 3° (diameter) for a zero epicentral distance. And it reduces to 8° on epicentral direction and to 3.5° orthogonal to it for epicentral distances of 60° (incidence angle at the CMB of 64.4°). Then, for an array aperture equal to 1/3 deg s^{-1} at all periods, the resolution at the CMB on orthogonal epicentral direction equals the resolution of the array at a period of 130 s. We use a minimum aperture of 2° (resolution of 1.83° at the CMB) in order to enclose a minimum number of stations. This resolution is sufficient in the period range we consider.

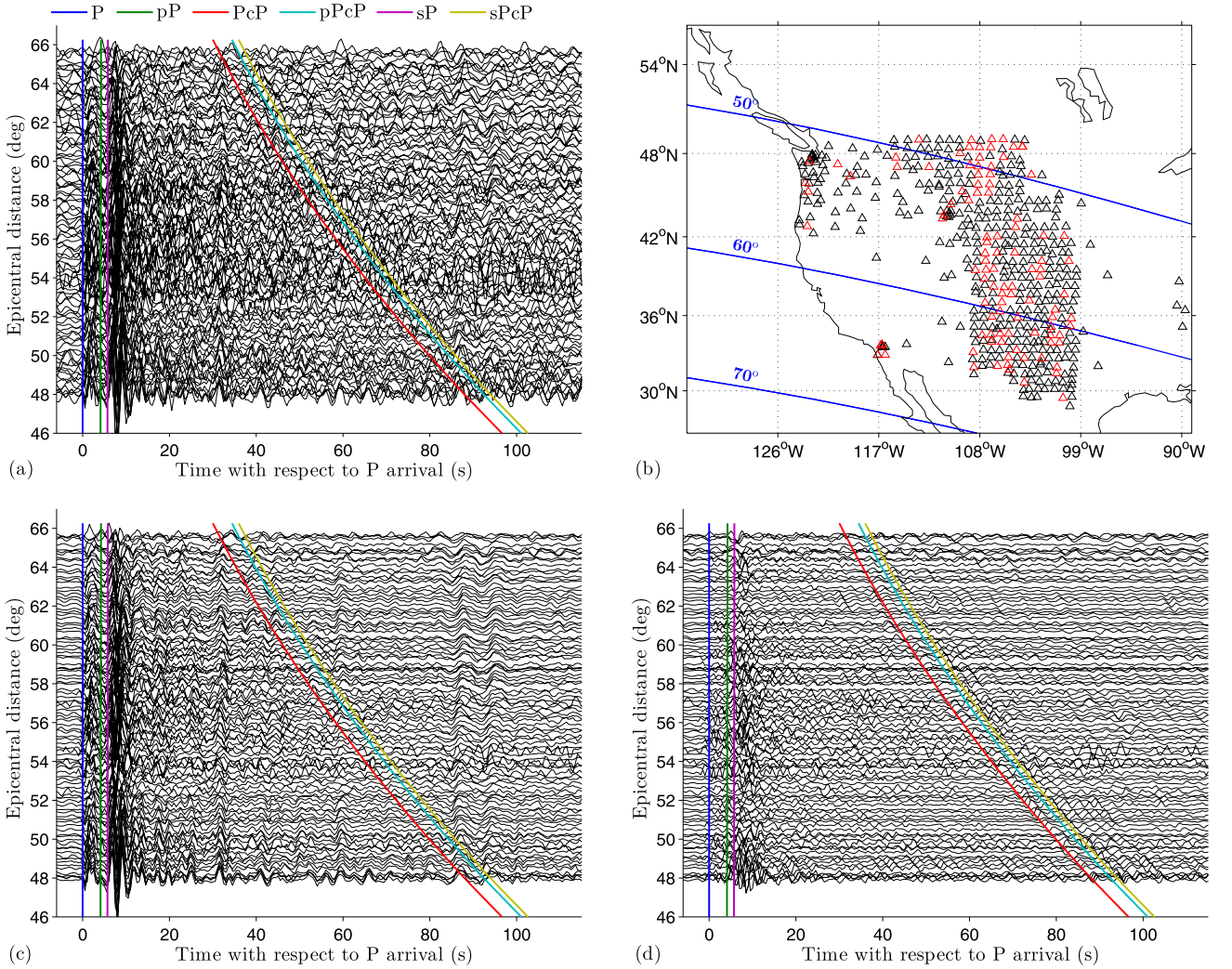

Figure 4. (a) Velocity data from an earthquake in the Svalbard region (2009/06/22 18:15:40.09 UTC, $M_w = 5.4$, depth = 13.1 km). Instrument response is removed and data are bandpass filtered from 1 to 60 s. (b) Epicentral distances (blue lines) and location of the stations with good quality signal shown in (a) (red triangles) and not shown (black triangles). (c) *P* extracted and (d) *PcP* extracted from the data shown in (a). All figures use the same amplitude normalization.

When smoothing is not admissible, we can use the differences (Figs 3b and d) instead of the extractions. These observations are noisier than the previous ones since random noise is not attenuated. However, we are able to obtain relatively clean observations of *PcP* without any smoothing because we remove most of the *P* signal and its post-cursors.

Fig. 4(a) shows velocity time-series from a shallow earthquake of magnitude $M_w = 5.4$ observed in north America (Fig. 4b). Due to its low magnitude, low SNR is the main problem to be addressed in this example in order to observe *PcP*. At these azimuths *P* post-cursors caused by Moho and upper-mantle discontinuities appear much weaker than in Fig. 1. Figs 4(c) and (d) present results after the filtering and synthesis operations using the configurations applied in the previous example. The noise attenuation is high enough to observe clearly several *P* post-cursors on Fig. 4(c) and *PcP* on Fig. 4(d). SNR can be further improved by increasing the array aperture in exchange for reducing spatial resolution.

In summary, the extracted observations from raw data improve SNR and SIR but are smoothed, while differential observations improve SIR and do not suffer from smoothing, but result in lower SNR.

4 *PcP–P* TRAVELTIME DIFFERENCE ESTIMATION WITH TWO SLIDING WINDOWS

The traveltime difference or delay between two seismic phases is obtained as an optimal solution of a given objective function. These functions conventionally define a distance between two signals based on their similarity or coherence. Valid notions of distance are any norm and some non-linear functions. The definition of similarity is also non-unique. It can be based on the actual signals, on their waveforms (neglecting amplitude differences), on their instantaneous phase, among others. For example, let x_1 and x_2 be two similar uniformly continuous signals separated by an unknown delay t_d and let us write the objective function most frequently used as:

$$\min \|x_1(t) - x_2(t - t_d)\|^2 . \tag{28}$$

Then it can be proven that the optimal solution is the delay corresponding to the absolute maximum of cross-correlation between x_1 and x_2, if similarity involves waveform and amplitude. If similarity involves only waveform, it is the delay with the maximum of 'geometrically normalized' cross-correlation.

Main difficulties conventionally arise from random noise and additional correlated signals. Broadly, noise reduces accuracy, while signals introduce bias. The reference seismic phase, signal x_1 in eq. (28), is chosen to be a clear signal, but its inherently limited frequency content and short duration may produce suboptimal solutions, due to, for example cycle-skipping effects. Additional signals arriving close in time may introduce additional suboptimal solutions. Noise has a direct impact on the accuracy of all solutions and deteriorates their amplitudes. When the effects of random noise are combined with uncertainties in the time of arrival of the reference signal and with lack of previous knowledge on t_d, the selection of the solution of choice among all optimal and suboptimal solutions becomes non-trivial.

The quality of *PcP–P* traveltime difference observations is exposed to noise and other correlated signals. Here, *P* is the reference signal x_1 and *PcP* x_2, leading to positive delays. *PcP* is a relatively weak signal with low SNR, which limits accuracy. Of more importance are the *P* and *PcP* depth phases. These signals may introduce additional solutions, as a result of their high similarity with the direct phases that arise from their close travel paths. The high similarity and particularly their close arrival times may introduce significant bias. In shallow events, less than 100 km depth, *pP* and *pPcP* arrive about 0.25 s km^{-1} after *P* and *PcP,* respectively per kilometre of the hypocentral depth, and *sP* and *sPcP* about 0.36 s km^{-1}. This is not a problem for very shallow events where differences among the observed *PcP–P*, *pPcP–pP* and *sPcP–sP* compared to predictions obtained with 1-D models are negligible. In fact, extra pairs of signals close to *P* and *PcP* and with very similar traveltime differences (below the actual resolution) often contribute to enhance accuracy. It is also not a problem for much deeper events where *P* and *PcP* depth phases arrive later than *P* and *PcP*. In this scenario, depth phases actually provide genuine information derived from significant differences in their travel paths, and hence in the CMB sampling. Depth phases are problematic at intermediate depths due to their close time of arrival but still significant traveltime difference. This leads to substantial bias when *pPcP–pP* or *sPcP–sP* are misinterpreted as *PcP–P*.

Other coherent signals may be present but their lower amplitude and/or different time of arrival make them less problematic. This is the case of *PdP*, a seismic phase with a travel path analogous to *PcP* but with a reflection at the top of D″ instead of CMB. This phase arrives a few seconds before *PcP* and has a slightly higher slowness. *PdP* observations provide precious information on the topography of D″. The similar travel paths of *PdP* and *PcP* but an often lower reflection coefficient at the top of D″ than at CMB hinders *PdP* observations. The presence of this phase may lead to a solution for *PdP–P* traveltime difference clearly separated from the most likely *PcP–P* traveltime difference, which should not lead to misinterpretations.

4.1 The two sliding window approach

A blind selection of an optimal portion of the seismogram as a reference signal in the sense of maximizing SNR and SIR based on predictions of 1-D models is in general problematic. Uncertainties on arrival time of *P* are higher than on the *PcP–P* traveltime difference, strong depth phases are not always present, and signal duration has strong variations. A long window guarantees that a major portion of the reference signal is covered, but leads to low SNR, and depth phases may not be excluded. Additionally, late portions of the *P* coda in strong events may suffer from non-negligible attenuation effects, in the sense of *t** operator, and introduce additional bias. A short window enables the maximization of SNR while rejecting depth phases, but cycle-skipping problems may become more severe, and uncertainties on the time of arrival of the reference signal hinders the optimal location window.

To overcome this dilemma we opt for short windows, but instead of using a fixed window on *P* and a sliding window on *PcP*, we let both move. Given the objective function shown in eq. (28), this is equivalent to adding a delay term in x_1 and x_2,

$$\min \|x_1(t - t_a) - x_2(t - t_d - t_a)\|^2 . \tag{29}$$

In summary, the 1-D search along t_d in eq. (28) turns into a 2-D search along the t_d and t_a variables in eq. (29). We could search for the global optimal solution, but it is far more powerful to relax constraints on t_a and t_d and search for optimal and suboptimal solutions. With this approach, we can potentially observe variations of *PcP–P* traveltimes along their waveforms, as well as those of *pPcP–P* and *sPcP–sP*, and other coherent signals such as *PdP–P*, among other signals that could be due to noise and cycle-skipping. In this scenario, the main complexity is to develop algorithms smart enough to identify all these signals using the prior information available.

The definition of similarity in objective functions for *PcP–P* traveltime difference estimation are based on the waveform. Amplitude differences are neglected due to the additional attenuation that *PcP* suffers compared to *P* as a result of its longer travel path. In fact, amplitude differences provide valuable observations on their own to distinguish velocity perturbations from topography. However, fully ignoring amplitudes may lead to optimal solutions with unrealistic attenuation values. Similarly, another issue that derives from using many reference signals, one for every t_a value used, eventually may arise when strong minima appear even when $x_1(t - t_a)$ only contains weak signals from the coda of the *P* wave, or just noise. To overcome these problems, we define a slightly different objective function to search for the signal x_2 that obtains the maximum energy reduction using the reference signal. We also add extra constraints on the amplitude difference to avoid unrealistic values. This writes

$$\max\ f(t_d, t_a), \tag{30}$$

with

$$f(t_d, t_a) = \|x_2(t - t_d - t_a)\|^2 - \|a x_1(t - t_a) - x_2(t - t_d - t_a)\|^2 , \tag{31}$$

where a is a real attenuation scalar.

The energy reduction varies according to t_d, t_a and a. The objective function $f(t_d, t_a)$ is maximized with respect to t_d when the second term is minimized. The optimal solution for attenuation at a given t_a and t_d that minimizes this term is:

$$a(t_d, t_a) = \frac{\langle x_1(t - t_a), x_2(t - t_d - t_a)\rangle}{\|x_1(t - t_a)\|^2}, \tag{32}$$

where $\langle x_1(t), x_2(t)\rangle$ is the inner product between x_1 and x_2. This is the optimum estimator of the *PcP/P* amplitude ratio using their actual signals in the least-square sense.

Introducing the optimal solution for attenuation in eq. (31), and simplifying we obtain

$$f(t_d, t_a) = \frac{|\langle x_1(t - t_a), x_2(t - t_d - t_a)\rangle|^2}{\|x_1(t - t_a)\|^2}. \tag{33}$$

Using eq. (32), the objective function defined in eq. (33) can be interpreted as

$$f(t_d, t_a) = a(t_d, t_a)\, \langle x_1(t - t_a), x_2(t - t_d - t_a)\rangle . \tag{34}$$

This equation facilitates the interpretation of thresholds in both the *PcP/P* amplitude ratio and the inner product between *P* and *PcP* signals.

We impose maximum and minimum thresholds as additional constraints on the amplitude ratio a. The first threshold contributes to the rejection of unlikely low-energy reference signals, since we expect *PcP* to have lower energy than *P*, while the second rejects x_2 signals with an unlikely low energy that leads to anomalously low *PcP/P* amplitude ratios.

4.2 Bias due to intrinsic attenuation

The accuracy of observations of both traveltime differences and amplitude ratios varies strongly with the signal quality but also with their similarity. Intrinsic attenuation has an effect on amplitude but also introduces frequency-dependent dispersion that may eventually cause a significant waveform distortion, leading to important bias. To a first order approximation (e.g. Shearer 2011) we can model attenuation as a function of travelled distance x, frequency ω and time t as

$$A(x,\omega,t) = A_0 \mathrm{e}^{-\omega t_0^*/2} \mathrm{e}^{-\mathrm{i}\omega(t-x/c_0)} \mathrm{e}^{-\mathrm{i}\omega t_0^* \ln(\omega/\omega_0)/\pi}, \tag{35}$$

where $t_0^* = x/c_0 Q$ ('tstar'), c_0 is the velocity at the reference frequency ω_0 and Q the quality factor, inversely proportional to attenuation. Note that the Fourier convention is opposite here: $\hat{u}(\omega) = \int_{-\infty}^{\infty} u(t)\mathrm{e}^{+\mathrm{i}\omega t}\mathrm{d}t$. The last exponential term in eq. (35) models the traveltime delay that intrinsic attenuation introduces at frequencies that deviate from ω_0. An option to reduce the effects of waveform distortion in differential traveltime measurements is to compensate for the integral of $t_0^* \ln(\omega/\omega_0)/\pi$ along the travel paths considered, according to a reference model of velocity and attenuation. Intrinsic attenuation attains higher values in the upper mantle than in the lower mantle, Q_μ is as low as 80 for the asthenosphere (depths from 80 to 220 km) compared to 312 in the lower mantle in PREM (Dziewonski & Anderson 1981). To evaluate waveform distortion, we use the value given by PREM in the lower mantle, $Q_\beta = Q_\mu = 312$, along the full travel path, considering that the *P* and *PcP* paths diverge mostly in the lower mantle, where they spend most of their time, especially at large epicentral distances. This translates into a $Q_\alpha = (9/4)Q_\mu = 702$ assuming infinite Q_κ, and writes:

$$(PcP\text{–}P)_{t_0^*} \simeq (t_{PcP} - t_P)\frac{\ln(\omega/\omega_0)}{Q_\alpha \pi}. \tag{36}$$

Dispersion due to intrinsic attenuation is thus approximately proportional to *PcP–P* traveltime difference. Here it may become significant only at low frequencies and large epicentral distances because Q_α is relatively high. Thus, we can safely ignore it in a first approximation because the high frequencies components of the signals, where dispersion is lower, provide accuracy to *PcP–P* traveltime observations. However, since corrections using eq. (36) in the frequency domain are simple, we do opt to correct for waveform distortion to maximize *P* and *PcP* similarity. These corrections may become mandatory in other applications such as *ScS–S*, due to their higher differential traveltimes, lower Q, and lower frequencies.

5 APPLICATION TO *PcP–P* TRAVELTIME DIFFERENCES

The resolution of the objective function in t_d is proportional to the cross-correlation length of x_1 and x_2, which is roughly inversely proportional to the maximum frequency components. However the resolution is far lower in t_a because the signal waveforms and the windows have finite width in time. This in practice produces relatively long and smooth ridges of suboptimal solutions along t_a with a relatively constant t_d value. To evaluate every solution, we first group them in separate ridges and then we estimate their significance based on the impact on energy reduction of the reference signal, $f(t_d, t_a)$. We group individual suboptimal solutions along increasing t_a in a single ridge as long as variations on t_d and gaps are small, and $f(t_d, t_a)$ does not reduce significantly. Ridges far shorter than the window length are discarded. To evaluate the rest, we calculate smooth histograms along t_d weighed by $f(t_d, t_a)$. More specifically, every solution along the ridge adds a narrow raised cosine function (Hann window) to a 1-D function with amplitude equal to $f(t_d, t_a)$ and centred at t_d. Once every solution is considered, the amplitude of the highest peak gives an estimate of the quality of the ridge and its position the traveltime difference. We finally identify the ridge that best represents *PcP–P*, or any other signal, based on its quality and its offset with respect to the prediction made using the prior model.

Figs 5 and 6 illustrate the objective function in two scenarios, (1) strong direct and depth phases and (2) moderate cycle-skipping ambiguity. Figs 5(a) and 6(a) present velocity seismograms with the instrument response removed and filtered with a bandpass filter from 1 to 60 s. The next two figures show the estimation of *P* (Figs 5b and 6b) and *PcP* (Figs 5c and 6c) using the slant-stacklet transform introduced in the previous section. Then, Figs 5(d) and 6(d) show the objective functions together with their main optimal solutions grouped in ridges in white and the values of *PcP–P* and *pPcP–pP* predicted by the global 1-D model AK135 (corrected for ellipticity) in magenta. We use a square window of 12 s to evaluate the objective function, and for the predicted solutions we consider that all signals are delta functions located at the predicted time of arrival.

In the first scenario (Fig. 5d), we can clearly identify *PcP–P* with the first strong ridge from about −2 to 9 s, and *pPcP–pP* with the secondary maximum from 14 to 22 s. Both traveltime differences agree with predictions. The traveltime differences along the main part of the *PcP–P* maximum ridge are very stable. We only have deviations at the extremes were maxima reduce severely, thus they do not perturb the observation, that is the location of the main peak of the amplitude-weighted histogram (Fig. 5e) is not affected by the extrema of the ridge due to their fluctuations and much lower maxima. The waveforms of *pP* and especially *pPcP* are weaker, and more affected by high frequency noise. This widens the secondary maxima and generates a far more unstable ridge.

In the second scenario (Fig. 6d), depth phases are extremely weak due to a low radiation in their direction of propagation; consequently, we can identify *PcP–P* as the main ridge from −4 to 6 s and, far weaker, *pPcP–pP* as the ridge between 7 and 18 s. The traveltime difference values are higher (positive) than predicted. Comparing the waveform of *P* and *PcP* is difficult to distinguish if *P* is faster than predicted, *PcP* is slower or both. In the evaluation of the objective function (Fig. 6d), we obtain a very stable ridge that leads to a *PcP–P* traveltime difference of about 0.5 s. The secondary ridge with traveltime difference of about −4 s is probably due to moderate cycle-skipping effects. They are produced by the combination of two factors, differences in *P* and *PcP* waveforms and the presence of secondary lobes. Waveform differences between *P* and *PcP* may reduce the amplitude of this solution below the values of others produced by weaker signals but that better resemble the reference signal. From 46 to about 49 s (Fig. 6c) we can observe a weak lobe which has a similar waveform to *P*. This matching produces the

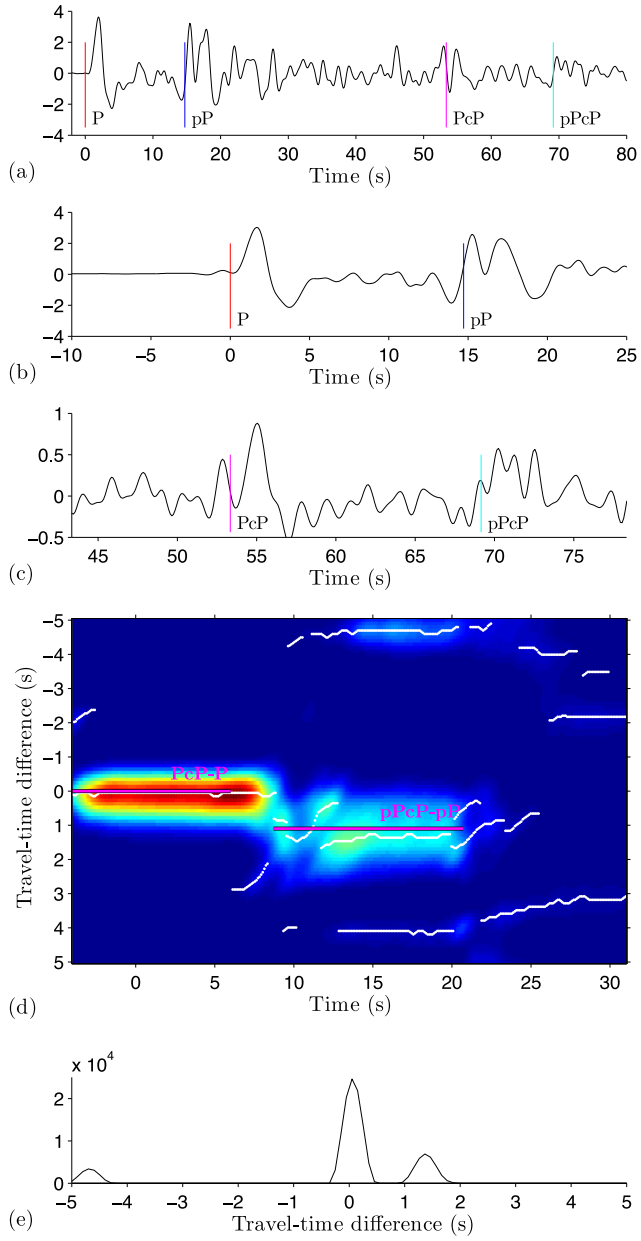

Figure 5. *PcP–P* traveltime estimation from an earthquake near the east coast of Kamchatka (2006/08/24 21:50:36.7 UTC, $M_w = 6.5$, depth = 54.2 km) with strong *pP* and *pPcP* signals towards the USArray. (a) Original signal at a distance of 57.2° (TA.N10A). (b) Observed *P* wave, x_1. (c) Observed *PcP* wave, x_2. (d) Objective function, eq. (34), with ridges in white and predictions from the AK135 model (corrected for ellipticity) in magenta. Vertical axis is traveltime difference variation with respect to predicted *PcP–P* with AK135, t_d, and horizontal axis is the time at the centre of the reference window, t_a. (e) Smooth histogram of traveltime differences of all ridges starting from −6 to 6 s having a minimum length of 4 s.

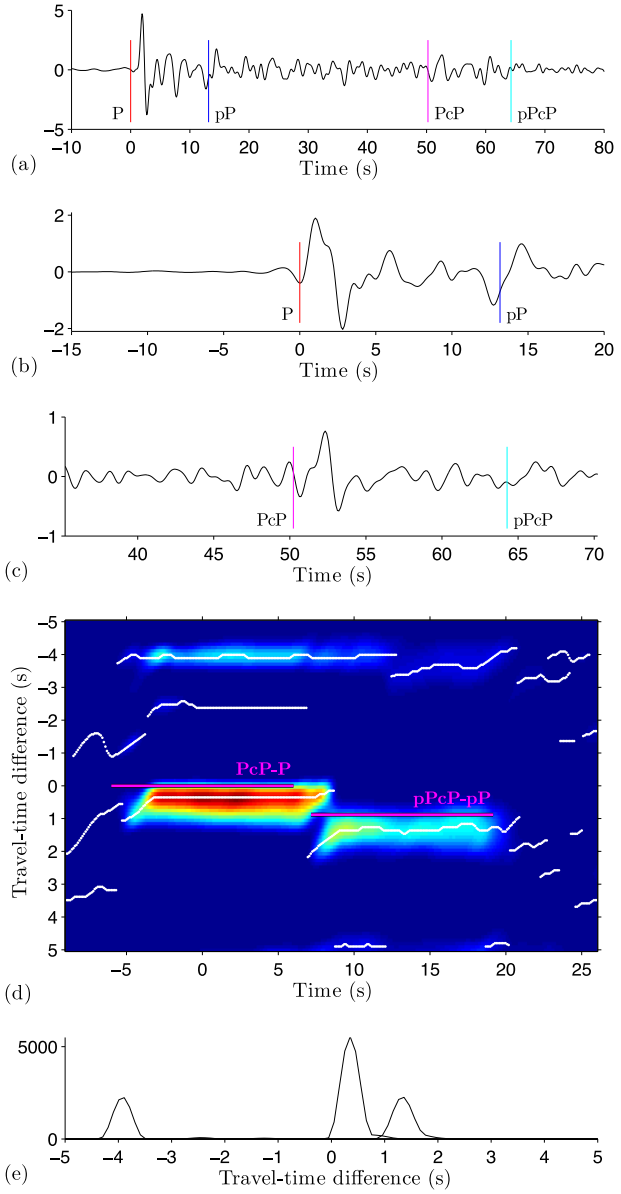

Figure 6. *PcP–P* traveltime estimation from an earthquake near the coast of northern Chile (2010/05/06 2:42:47.9 UTC, $M_w = 6.2$, depth = 46.8 km) with a minor cycle skipping problem. (a) Original signal at a distance of 58.72° (TA.X34A). (b) Observed *P* wave. (c) Observed *PcP* wave. (d) Objective function with ridges indicated. (e) Smooth histogram of traveltime differences of all ridges.

ridge in Fig. 6(d) with a traveltime difference of about −4 s. With the actual SNR it is difficult to distinguish if this signal has a real origin, for example *PuP* (caused by a drop in compressional wave speed in an ULVZ), or it is just coherent-noise.

We test robustness of *PcP–P* traveltime observations to strong interferences (i.e. *P* post-cursors and scatters), building synthetic seismograms using DSM (Kawai *et al.* 2006) with the isotropic PREM model (Dziewonski & Anderson 1981). In Fig. 7(a), we model the seismograms from the event shown in Fig. 1 using the global CMT solution for the moment tensor (Dziewonski *et al.* 1981; Ekström *et al.* 2012) with a maximum frequency of 2 Hz, but with an event depth of 100 km. We extract *P* (Fig. 7c) and *PcP* (Fig. 7d) from this data set using the configuration of the slant-stacklet transform used in Figs 1 and 4. Figs 7(e) and (f) plot *PcP–P* differential traveltime observations, assessed using the two sliding window approach, according to theoretical values for the isotropic PREM model calculated with the TauP Toolkit (Crowell *et al.* 1999) with the same ellipticity corrections used in DSM. On the raw data results (Fig. 7e), we see that the *PcP–P* observations are contaminated by the strong energy of the *P* coda, *pP* and *sP*; while on the extracted signals (Fig. 7f), *PcP–P* observations are close to zero (mean of −0.12 s and standard deviation of 0.08 s),

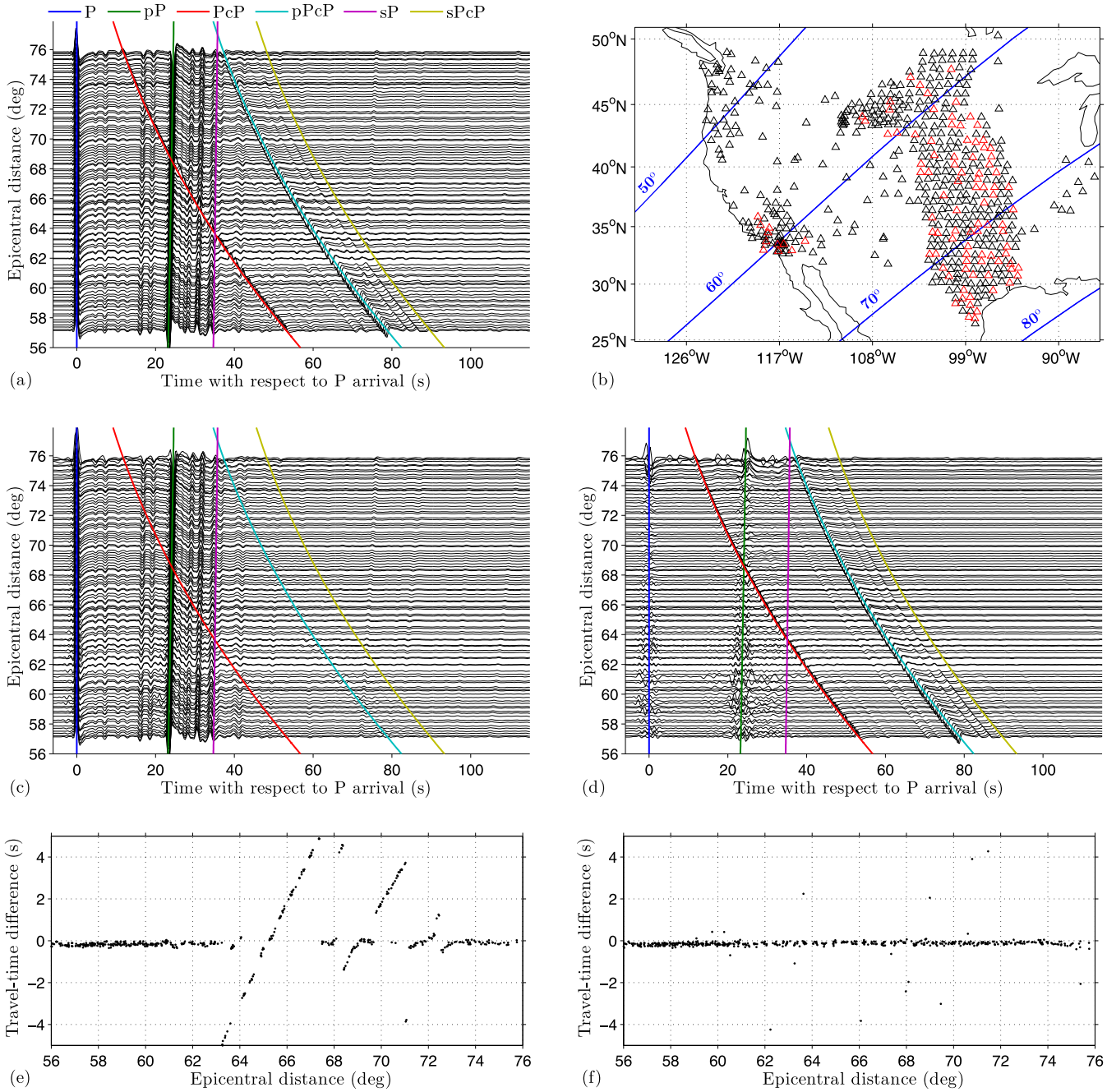

Figure 7. *PcP–P* traveltime differences on synthetics (DSM) compared to expected observations (TauP) using the isotropic PREM model. (a) Synthetics are for the geometry shown in Fig. 1 but with an event depth of 100 km. (b) Epicentral distances (blue lines) and location of the stations shown in (a) (red triangles) and not shown (black triangles). Extracted *P* (c) and *PcP* (d) using the slant-stacklet transform. *PcP–P* traveltime difference variations using the method of two sliding windows on (e) the raw synthetics and (f) on the extracted *P* and *PcP*.

as it is expected from using the same 1-D model in DSM and TauP. Remaining variations are mainly due to the non-ideal separation of *P* and *PcP* and their limited high frequency contents.

5.1 Illustration on real data

Fig. 8 shows observed traveltime difference variations according to the AK135 1-D model corrected for ellipticity for the event shown in Fig. 1 along epicentral distance. Seeing the distribution of the observations in Fig. 8, we consider outliers as any observation outside the interval of ±2 s. Fig. 8(a) shows observations taken using the raw data, Fig. 8(b) using the extracted *P* and *PcP* signals (Figs 3a and c), and Fig. 8(c) using their differences (Figs 3b and d).

PcP–P differential traveltime observations are based on their waveform similarity. However, as *P*, *pP* and later signals of similar slowness are highly cross-correlated, it is hard to distinguish maxima of the objective function caused by *PcP* or other signals. When we measure *PcP–P* on the raw data (Fig. 8a), these signals become the major interference. Their often higher amplitude and similarity compared to *PcP* produce higher maxima which may lead to systematic errors such as those observed in Fig. 8(a), where

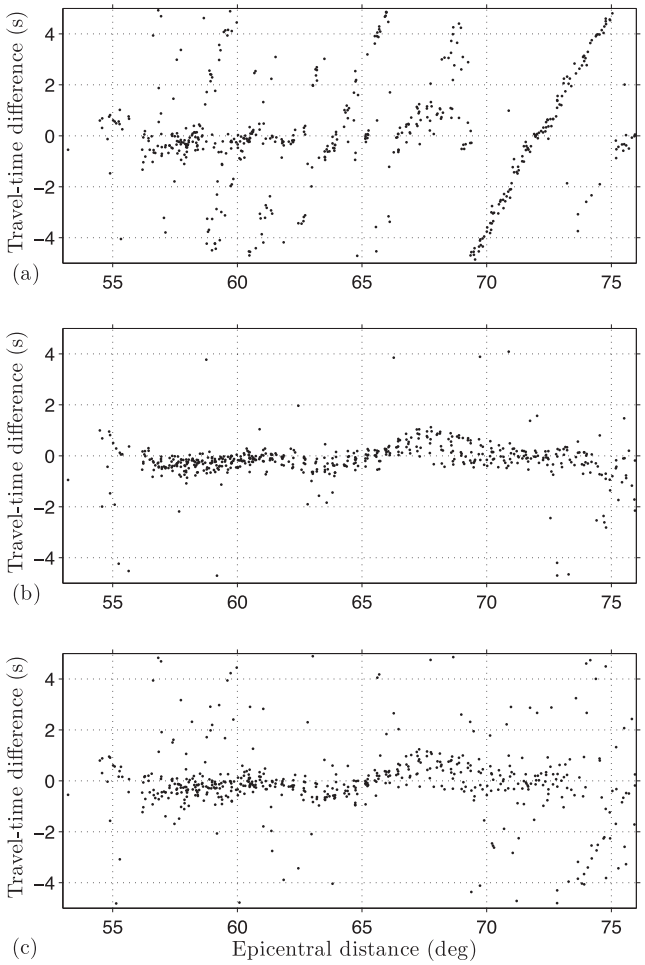

Figure 8. Observed *PcP–P* traveltime difference variations compared to AK135 1-D model, $(PcP{-}P)_{\text{data}} - (PcP{-}P)_{\text{model}}$, using (a) the raw data of Fig. 1, (b) the *P* and *PcP* extractions (Figs 3a and c), (c) the data minus the extracted *P* and the data minus the extracted *PcP* (Figs 3b and d).

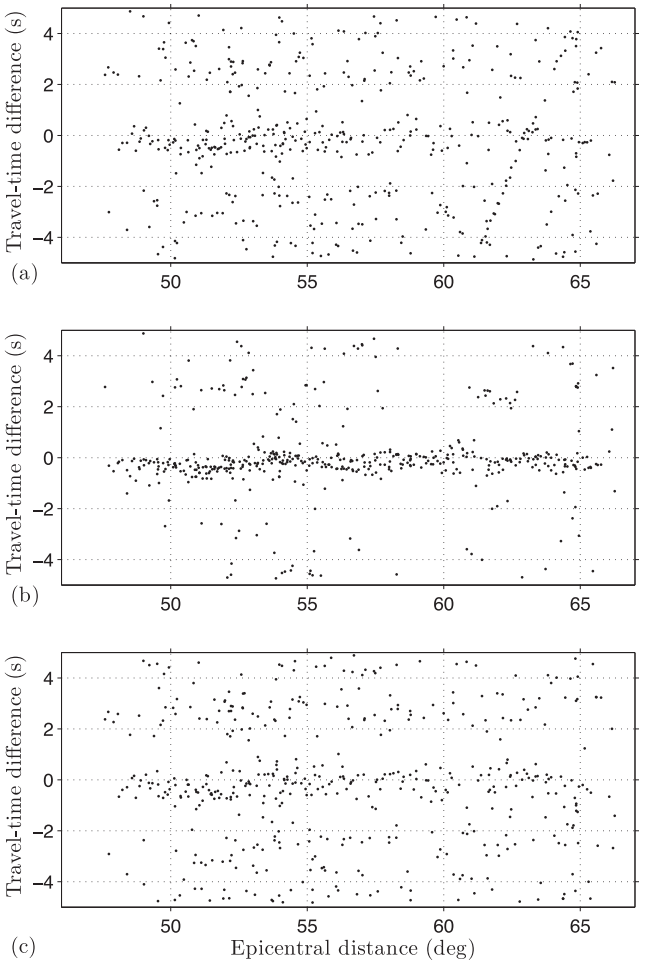

Figure 9. Observed *PcP–P* traveltime difference variations compared to AK135 1-D model corrected for ellipticity, $(PcP{-}P)_{\text{data}} - (PcP{-}P)_{\text{model}}$, using (a) the raw data of Fig. 4(a), (b) the *P* and *PcP* extractions (Figs 4c and d), (c) the data minus the extracted *P* and the data minus the extracted *PcP*.

most observations are aligned along anomalous diagonal lines. This problem is fully solved in Fig. 8(b) by using the extracted *P* and *PcP* signals, and even when we use the difference measurements in Fig. 8(c), no systematic error appears. In the latter case, we can observe a higher number of random outliers, and a slightly higher dispersion on the observations, which is expected considering that random noise is not attenuated in this case and that *P* and *PcP* do not suffer any spatial smoothing. Interpretation of these results in terms of structure in the vicinity of the CMB requires an additional step to correct for mantle structure. In this example, correcting for mantle structure reduces the variance of the observations significantly, in particular, from 65° to 70°. This will be addressed in a companion paper (Ventosa & Romanowicz 2015).

Equivalently to Fig. 8, Fig. 9 shows observed *PcP–P* traveltime difference variation for the event shown in Fig. 4. *PcP–P* observations are not possible in the raw data (Fig. 9a) due to the low SNR of *PcP*; however, in Fig. 9(b) we observe relatively clean observations using the extracted *P* and *PcP* signals (Figs 4c and d). In this figure, most observations concentrate at about zero. Small variations around this value are due to local variations and, to a smaller extent, to measurement uncertainties. SNR can be further enhanced using wider array apertures at the cost of reducing spatial resolution. We consider spatially close observations far from this main trend as outliers. In this scenario the observations on the differences (Fig. 9c) perform closer to the raw data because noise is not attenuated and other correlated signals are relatively small.

6 CONCLUSIONS

We have shown that we can obtain *PcP–P* traveltime difference measurements of unprecedented quality using array data in the presence of interfering signals. The key ingredients which make these observations possible are accurate signal extraction and local traveltime difference measurements.

The slant-stacklet transform achieves a high slowness selectivity without sacrificing flexibility on the resolution compromises. This feature has proven useful in *P* and *PcP* signal extraction. Now it is possible to obtain clean observations of *PcP* even when it is hidden in the coda of *P*. The high slowness resolution allows the separation of *P* and *PcP* up to 80° of epicentral distance, where their slowness difference and time difference are very small, and probably more importantly, it allows a high level of rejection of interfering signals, such as *pP* and *sP*, even when their accurate slowness and time of arrival are not precisely known in advance. Our present limit in signal extraction is the level of noise. Working at larger epicentral distances is possible with larger apertures; however, *P* and *PcP* kernels converge (at 85° the bottoming point of *P* is

slightly deeper than 2500 km) and may partially overlap at relatively short periods, making them less attractive. When it is not possible to attain the slowness resolution required in the signal extraction due to spatial resolution constraints, alternative approaches based on the prediction of interfering signals followed by their adaptive subtraction with no slowness and spatial resolution limitations (e.g. Ventosa *et al.* 2012a; Pham *et al.* 2014) may be of interest.

The main problem in the estimation of the traveltime difference is the high cross-correlation with other close and usually stronger signals. Although the slant-stacklet transform removes the strongest correlated signals that mask *PcP*, a local traveltime difference estimation is nearly as important to reduce bias and outliers. In particular, the objective function we introduce maximizes the energy reduction on the objective signal (*PcP*). We conclude that its solution is less sensitive to noise than the standard cross-correlation solution and naturally allows the introduction of additional constraints on the amplitude ratio of the two signals. Furthermore, its 2-D representation (time – differential time), in addition to providing a less ambiguous *PcP–P* identification, allows a more involved analysis of many often weaker signals with precious complementary information such as *pPcP–pP*, *sPcP–sP*, *PdP–P*, among others.

At least as important as the accuracy of *PcP–P* observations is an exhaustive analysis of all sources of bias. We conclude that although epicentre mislocation and intrinsic attenuation biases are present, they are negligible compared to the accuracy of our observations. As we will show in our companion paper, they are much smaller than the bias caused by mantle heterogeneities.

Although we illustrate these tools with *PcP* data, they address general problems of teleseismic signal processing. Hence, they should be useful for a wider range of applications. Apart from applications under similar conditions using other phases constraining CMB and D′′ or the inner-core boundary, the slant-stacklet transform may prove useful in applications that benefit from removing random noise and interfering signals in slowness without compromising resolution, such as in the imaging of the upper-mantle discontinuities using SS precursors (e.g. Zheng *et al.* 2015), or in the observation of weak signals from the core.

ACKNOWLEDGEMENTS

The data used in this work were obtained from the IRIS Data Management Center. The slant-stacklet transform has its origin in the PhD thesis of Sergi Ventosa at the UTM-CSIC supervised by Carine Simon, Martin Schimmel and Juanjo Dañobeitia. We thank to anonymous reviewers for helping improve the clarity of the manuscript. We also acknowledge the support from the European Research Council under the European Community's Seventh Framework Programme (FP7-IDEAS-ERC)/ERC Advanced Grant WAVETOMO.

REFERENCES

Bear, L.K. & Pavlis, G.L., 1997. Estimation of slowness vectors and their uncertainties using multi-wavelet seismic array processing, *Bull. seism. Soc. Am.,* **87**(3), 755–769.

Cao, A., Romanowicz, B. & Takeuchi, N., 2005. An observation of PKJKP: inferences on inner core shear properties, *Science,* **308**(5727), 1453–1455.

Castle, J.C. & van der Hilst, R.D., 2000. The core-mantle boundary under the Gulf of Alaska: no ULVZ for shear waves, *Earth planet. Sci. Lett.,* **176**(3), 311–321.

Cobden, L. & Thomas, C., 2013. The origin of D′′ reflections: a systematic study of seismic array data sets, *Geophys. J. Int.,* **194**(2), 1091–1118.

Crowell, H.P., Owens, T.J. & Ritsema, J., 1999. The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, *Seism. Res. Lett.,* **70,** 154–160.

Deuss, A., Woodhouse, J.H., Paulssen, H. & Trampert, J., 2000. The observation of inner core shear waves, *Geophys. J. Int.,* **142**(1), 67–73.

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

Dziewonski, A.M., Chou, T.-A. & Woodhouse, J.H., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. geophys. Res.,* **86,** 2825–2852.

Ekström, G., Nettles, M. & Dziewonski, A.M., 2012. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes, *Phys. Earth planet. Inter.,* **200–201,** 1–9.

Fichtner, A. & Igel, H., 2008. Efficient numerical surface wave propagation through the optimization of discrete crustal models—a technique based on non-linear dispersion curve matching (DCM), *Geophys. J. Int.,* **173**(2), 519–533.

Forbriger, T., 2003. Inversion of shallow-seismic wavefields – I: wavefield transformation, *Geophys. J. Int.,* **153**(3), 719–734.

French, S., Lekić, V. & Romanowicz, B., 2013. Waveform tomography reveals channeled flow at the base of the oceanic asthenosphere, *Science,* **342**(6155), 227–230.

Frost, D.A., Rost, S., Selby, N.D. & Stuart, G.W., 2013. Detection of a tall ridge at the core-mantle boundary from scattered PKP energy, *Geophys. J. Int.,* **195**(1), 558–574.

Gershman, A.B., Sidiropoulos, N.D., Shahbazpanahi, S., Bengtsson, M. & Ottersten, B., 2010. Convex optimization-based beamforming: from receive to transmit and network designs, *IEEE Signal Process. Mag.,* **27**(3), 62–75.

Gu, Y.J. & Sacchi, M., 2009. Radon transform methods and their applications in mapping mantle reflectivity structure, *Surv. Geophys.,* **30**(4-5), 327–354.

Hyvärinen, A. & Oja, E., 2000. Independent component analysis: algorithms and applications, *Neural Networks,* **13**(4), 411–430.

Jacques, L., Duval, L., Chaux, C. & Peyré, G., 2011. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity, *Signal Process.,* **91**(12), 2699–2730.

Kawai, K., Takeuchi, N. & Geller, R.J., 2006. Complete synthetic seismograms up to 2 Hz for transversely isotropic spherically symmetric media, *Geophys. J. Int.,* **164,** 411–424.

Kay, S.M., 1993. *Fundamentals of Statistical Signal Processing,* Prentice-Hall.

Lekić, V. & Romanowicz, B., 2011. Inferring upper-mantle structure by full waveform tomography with the spectral element method, *Geophys. J. Int.,* **185**(2), 799–831.

Li, J., Stoica, P. & Wang, Z., 2003. On robust Capon beamforming and diagonal loading, *IEEE Trans. Signal Process.,* **51**(7), 1702–1715.

Lorenz, R. & Boyd, S., 2005. Robust minimum variance beamforming, *IEEE Trans. Signal Process.,* **53**(5), 1684–1696.

Ma, J. & Plonka, G., 2010. The Curvelet transform—a review of recent applications, *IEEE Signal Process. Mag.,* **27**(2), 118–133.

Mallat, S., 2009. *A Wavelet Tour of Signal Processing: The Sparse Way,* 3rd edn, Academic Press.

Pham, M.Q., Duval, L., Chaux, C. & Pesquet, J.-C., 2014. A primal-dual proximal algorithm for sparse template-based adaptive filtering: application to seismic multiple removal, *IEEE Trans. Signal Process.,* **62**(16), 4256–4269.

Rost, S. & Revenaugh, J., 2004. Small-scale changes of core-mantle boundary reflectivity studied using core reflected PcP, *Phys. Earth planet. Inter.,* **145**(1), 19–36.

Rost, S. & Thomas, C., 2010. High resolution CMB imaging from migration of short-period core reflected phases, *Phys. Earth planet. Inter.,* **183**(1–2), 143–150.

Scherbaum, F., Krüger, F. & Weber, M., 1997. Double beam imaging: mapping lower mantle heterogeneities using combinations of source and receiver arrays, *J. geophys. Res.: Solid Earth,* **102**(B1), 507–522.

Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.,* **130**(2), 497–505.
Schmidt, R., 1986. Multiple emitter location and signal parameter estimation, *IEEE. Antenn. Propag.,* **34**(3), 276–280.
Shearer, P.M., 2011. *Introduction to Seismology,* 2nd edn, Cambridge Univ. Press.
Taner, M.T., Koehler, F. & Sheriff, R.E., 1979. Complex seismic trace analysis, *Geophysics,* **44**(6), 1041–1063.
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.
Thomas, C., Heesom, T. & Kendall, J.M., 2002. Investigating the heterogeneity of the D″ region beneath the northern Pacific using a seismic array, *J. geophys. Res.,* **107**(B11), ESE 3–1–ESE 3–7.
Ventosa, S. & Romanowicz, B., 2015. Extraction of weak PcP phases using the slant-stacklet transform – II: constraints on lateral variations of structures near the core-mantle boundary, *Geophys. J. Int.,* in press.
Ventosa, S., Rabeson, H. & Duval, L., 2011. Coherent noise removal in seismic data with redundant multiscale directional filters, in *Proceedings of the 19th European Signal Processing Conference,* pp. 1150–1154, Barcelona, Spain.
Ventosa, S., Roy, S.L., Huard, I., Pica, A., Rabeson, H., Ricarte, P. & Duval, L., 2012a. Adaptive multiple subtraction with wavelet-based complex unary wiener filters, *Geophysics,* **77**(6), V183–V192.
Ventosa, S., Simon, C. & Schimmel, M., 2012b. Window length selection for optimum slowness resolution of the local-slant-stack transform, *Geophysics,* **77**(2), V31–V40.
Vidale, J.E. & Benz, H.M., 1992. A sharp and flat section of the core-mantle boundary, *Nature,* **359,** 627–629.
Wookey, J. & Helffrich, G., 2008. Inner-core shear-wave anisotropy and texture from an observation of PKJKP waves, *Nature,* **454**(7206), 873–876.
Yilmaz, Ö., 2001. *Seismic Data Analysis,* Vol. 1, Society of Exploration Geophysicists.
Zheng, Z., Ventosa, S. & Romanowicz, B., 2015. High resolution upper mantle discontinuity images across the Pacific Ocean from SS precursors using local slant stack filters, *Geophys. J. Int.,* in press.
Zhu, H., Bozdağ, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nat. Geosci.,* **5**(7), 493–498.

APPENDIX A: RESOLUTION ANALYSIS

The slant-stacklet expansion shown in eq. (4) can be rewritten using the convolution operator over t and x:

$$W v_{p,x_c}(\tau,\lambda) = u \star \phi_{p,\lambda}(-\tau,-x_c). \tag{A1}$$

The Fourier transform of $\phi_{\lambda,p}(t,x)$ is

$$\hat{\phi}_{\lambda,p}(\omega,k) = \sqrt{\lambda}\hat{\psi}^*(-\lambda\omega)\hat{a}_\lambda(k+p\omega), \tag{A2}$$

where ω is frequency in rad s^{-1} and k wavenumber in rad deg^{-1}. $\hat{\psi}(\omega)$ denotes the Fourier transform of $\psi(\omega)$.

This equation involves the product of the Fourier transforms of the wavelet family of functions, $\sqrt{\lambda}\hat{\psi}^*(-\lambda\omega)$, and a set of slowness bandpass filters, $\hat{a}_\lambda(k+p\omega)$. The first term is the transfer function of a set of bandpass filters of bandwidth proportional to ω, that is of constant quality factor, and central frequency η/λ, where $\eta = \frac{1}{2\pi}\int_0^\infty \omega|\hat{\psi}(\omega)|^2 \mathrm{d}\omega$. The second term decomposes the Fourier transformed domain into a set of diagonal bands with slope of $-1/p$ centred at the origin. This is illustrated in Fig. A1.

The time and frequency resolutions are controlled by $\psi_{\tau,\lambda}(t)$, while the spatial and wavenumber resolutions by $a_\lambda(x)$. The

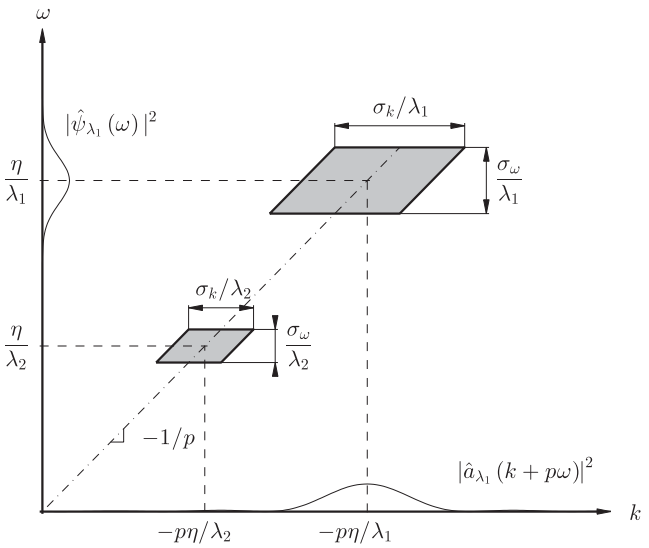

Figure A1. Frequency wavenumber resolution of two slant-stacklet atoms with $a_\lambda(x) = \lambda^{-1}a(\lambda^{-1}x)$ and x, $p \in \mathbb{R}$.

resolution in time is $\lambda\sigma_t$, in frequency σ_ω/λ, and the area of the Heisenberg box is $\sigma_t\sigma_\omega$. Similarly, the energy spread in the space dimension is:

$$\frac{1}{\|a_\lambda(x)\|^2}\int_{-\infty}^{\infty} x^2\,|a_\lambda(x)|^2\,\mathrm{d}x = f^2(\lambda)\sigma_x^2 \tag{A3}$$

and in wavenumber

$$\frac{1}{2\pi\,\|a_\lambda(x)\|^2}\int_{-\infty}^{\infty} k^2\,|\hat{a}_\lambda(k)|^2\,\mathrm{d}k = \frac{\sigma_k^2}{f^2(\lambda)}. \tag{A4}$$

The space and slowness resolutions are thus controlled by $f(\lambda)$. The resolution in space is $f(\lambda)\sigma_x$ and in wavenumber $\sigma_k/f(\lambda)$. The space wavenumber area remains equal to $\sigma_x\sigma_k$.

The slowness resolution at the scale λ is the ratio between the wavenumber resolution and the central frequency of the wavelet at this scale, $\sigma_p = \sigma_k\lambda/\eta f(\lambda)$. The function $f(\lambda)$ is in general smooth. When this function is constant, $f(\lambda) = b$, this leads to the same frequency wavenumber resolution as the LSST (Ventosa *et al.* 2012b), $\sigma_p = \sigma_k\lambda/b\eta$; and when it is proportional to scale, $f(\lambda) = b\lambda$, to slant-stacklet atoms whose shape and slowness resolution remain constant with scale, $\sigma_p = \sigma_k/b\eta$ (see Fig. A1).

APPENDIX B: DISCRETIZATION

The actual resolution of the slant-stacklet expansion is not regular across the transformed space. A wavefield u can be recovered from its inner products with a finite family of functions generated from $\phi_{\tau,\lambda,p,x_c}(t,x)$ if this family of functions satisfies the frame inequality (i.e. constitutes a frames)

$$A \le \sum_{n\in\Gamma}|\hat{\phi}_n(\omega,k)|^2 \le B, \tag{B1}$$

where $B \ge A > 0$ are two constants for almost all ω and k in $\mathbb{R}^N$, n is a shorthand for the indices τ, λ, p, x_c and Γ for their domains. When $A = B$ the frame is called tight and if additionally $\{\phi_n\}_{n\in\Gamma}$ are linearly independent, the frame is not redundant and is called an orthonormal basis.

We search for a close to tight frame, $A \simeq B$, because the balance it provides between simplicity—having a nearly perfect

reconstruction to reduce estimation errors on the signals extracted—and maximum freedom when setting the expansion. In addition, it is convenient to have freedom in the choice of the mother wavelet, and of the slowness bandpass filter and its scaling law to maximize the performance of filtering operations. For these reasons, we opt for complex-wavelet transforms based on frames of wavelets with several voices per octave, and a sampling of the slowness dimensions proportional to the actual resolution. We preserve at this stage the sampling across the spatial dimensions of the wavefield, frequently sparse and always limited.

We thus define the family of functions as

$$\phi_{r,j,v,s,x_c}(t,x) = a_{j,v}(x - x_c)\psi^*_{r,j,v}\left(t - p_s^T(x - x_c)\right) \tag{B2}$$

with

$$a_{j,v}(x) = \frac{1}{f(2^{j+v/V})}a\left(\frac{x}{f(2^{j+v/V})}\right) \tag{B3}$$

and

$$\psi^*_{r,j,v}(t) = \frac{1}{\sqrt{2^{j+v/V}}}\psi^*\left(\frac{t - rb_0 2^j}{2^{j+v/V}}\right), \tag{B4}$$

where scale is $\lambda = 2^{j+v/V}$ and delay $\tau = rb_0 2^j$, where $j \in \mathbb{Z}$ is an octave, $v \in [0, V-1]$, a voice, V the number of voices, $r \in \mathbb{Z}$ delay and b_0 sampling period. We discretize slowness as $p_s = p_0 + s\lambda c_0/f(\lambda)$, considering that slowness resolution is inversely proportional to frequency and controlled by $f(\lambda)$, where $s \in \mathbb{Z}$ slowness index and c_0 slowness sampling period.