

OPEN

# Virtual-source imaging and repeatability for complex near surface

Yang Zhao<sup>1\*</sup>, Tao Liu<sup>2</sup>, Genyang Tang<sup>1</sup>, Houzhu Zhang<sup>3</sup> & Madhumita Sengupta<sup>3</sup>

Based on seismic interferometry, the virtual source (VS) method is able to produce virtual gathers at buried receiver locations by crosscorrelating the direct-downgoing waves with corresponding reflected-upgoing waves from surface-source gathers. Theoretically, the VS records can improve seismic quality with less negative impact from overburdened complexities. However, shallow complex structures and weathering layers at near surface not only severely distort the wavepaths, but also introduce multiples, surface waves, scattering noise, and interference among different wave modes. These additional seismic responses contaminate both direct-downgoing and reflected-upgoing wavefields. As a result, the VS gathers experience spurious events and unbalanced illuminations associated with distorted radiation patterns. Conventional stacking operator can produce significant artifacts for sources associated with ineffective-wavepath cancellation. We review three publications and summarize a comprehensive workflow to address these issues using data-driven offset stacking, wavelet-crosscorrelation filtering, and radiation-pattern correction. A data-driven offset stacking theme, with each individual source contribution is weighted by certain quality measures, is applied for available offsets. The wavelet crosscorrelation transforms time-offset data into local time-frequency and local time-frequency-wavenumber domains. Filters are designed for the power-spectrum in each domain. The radiation-pattern correction spatially alters the contaminated direct-wavefields using a zero-phase matched filter, such that the filtered wavefield is consistent with the model-based direct P-wavefields observed at buried receiver locations. Our proposed workflow produces significant improvement as demonstrated in the 13 time-lapse field surveys that included substantial repeatability problems across a 17-month survey gap.

Thirteen repeated 2D surveys were acquired over the course of 19 months above an onshore field in the Middle East. The first six surveys (S1-S6) were collected within a period of three-months, after a 17-month break, an additional seven surveys (S7-S13) were acquired over a period of a week. All surveys were acquired with Mertz 26 vibrators with most of the shot locations repeated with better than 1 m accuracy. A 2D line of 80 receiver stations were installed with geophones cemented in individual vertical boreholes with 30 m spacing. The sensors are at depths of 50 m below surface. Dense 3D areal shooting (7.5 m inline and 7.5 m crossline) was performed for linear noise removal, and above the output VS location for optimum illumination. The corresponding offsets range between 0–2400 m. The geology is composed of a number of layers with large velocity contrasts overlaying a target reflector at approximately 2000 m depth. The thickness of superficial sand-layer varies from a few meters to more than 50 meters. The near-surface that is covered by thick sand is considered to be an area which produces sub-optimal seismic data quality. Below the near-surface, there is a simple layer-cake geology and the associated reflector dip is less than 5 degrees. According to Bakulin, *et al.*<sup>1</sup>, the image quality and repeatability of both post-stack and pre-stack data suggests the majority of the changes are associated with extremely shallow near-surface variations. The use of a buried system provides an opportunity to redatum the surface sources to the receiver level using the virtual-source methods. This is expected to greatly improve the image quality, coupling changes, and diurnal/seasonal temperature variations in the time-lapse processing.

<sup>1</sup>State Key Laboratory of Petroleum Resource and Prospecting, and Unconventional Petroleum Research Institute, China University of Petroleum, 18 Fuxue Road, Changping District, Beijing, 102249, China. <sup>2</sup>Sinopec Petroleum exploration & Production Research Institute, 100083, Beijing, China. <sup>3</sup>Houston Research Center, Aramco Services Company, Houston, TX, 77074, USA. \*email: [zhaoyangprof1@qq.com](mailto:zhaoyangprof1@qq.com)

The virtual-source (VS) method, based on seismic interferometry, originates from global seismology. Shapiro and Campillo<sup>2</sup> and Lobkis and Weaver<sup>3</sup> extracted Green's functions by crosscorrelating long sequences of ambient seismic noise, normally discarded in traditional processing, which contains information about the structure of the shallow and middle crust. In exploration seismology, the VS may help produce images from underneath a complex overburden with without knowledge of the overburden velocities and near-surface changes. This workflow crosscorrelates direct-downgoing waves with upgoing (reflected) seismic waves to retrieve the reflection response, and then redatums the surface-source records to buried receiver locations<sup>4,5</sup>. The power of actual buried receivers is that they allow for the direct measurement of wavefields propagating through the intricate overburden structure, and offers much more accurate measurements than model-based approaches<sup>6,7</sup>. Wapenaar<sup>8</sup> and van der Neut, *et al.*<sup>9</sup> derived interferometric representations to retrieve exact VS Green's function in elastic media. More importantly, redatuming is able to improve time-lapse survey repeatability by correcting time-lapse noise, such as that induced by near-surface diurnal and seasonal weather cycles as well as small changes in acquisition geometry and shot coupling<sup>5</sup>. In summary, VS should simplify a recorded seismic wavefield and correct the distortions associated with heterogeneities located between the seismic sources and the receivers.

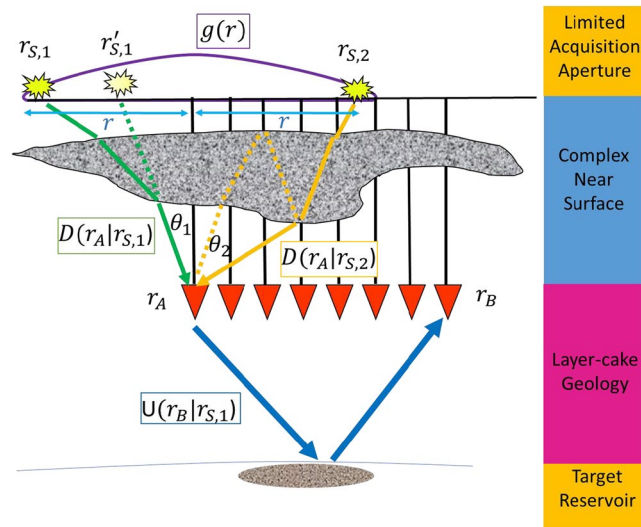
The VS theory<sup>10–12</sup> states that a redatumed VS response can be fully recovered by integration over sources located on a closed surface surrounding the receiver. In the stationary-phase zone, contributions to the VS are summed constructively and the common ray paths can be effectively cancelled. However, this is not true for sources located outside of the stationary zone<sup>13–15</sup>. The current practice of finding source contributions within the stationary phase zone is based on a simple distance-weighting scheme<sup>5</sup>. Near offsets (with respect to the VS) are usually considered to be stationary, leading to a simple linear weight function and decays exponentially with offsets. Nevertheless, due to the limited offsets and insufficient sampling of land acquisition data, repeatability and image-quality issues still exist because the VS assumptions failed<sup>8,16</sup>. As a result, onshore seismic data processing requires specialized denoising steps to restore the coverage so that redatuming can work effectively. For example, wavefield separation (dual sensors) and direct-arrival windowing have been shown to be the most effective approaches to isolate the direct-downgoing P-wave<sup>5,17,18</sup>. However, the complex near surface can severely distort the ray paths so that even near-offset sources may lie outside of the stationary area, whereas far-offset sources may provide constructive contributions to VS. Additionally, near-surface reverberations, surface multiples, and other mode-converted waves may leak into the windowed early arrivals and further corrupt the direct wavefields<sup>19,20</sup>. As a result, early arrivals often contain near-surface reverberations and multiples which interfere with upgoing reflections. Wavefield contamination results in poor suppression of multiples, scattered energy, and surface waves. The amplitude radiation-pattern of the VS distorts and creates non-uniform illumination for target reservoirs. This leads to degraded image quality where the VS stack has a signal-to-noise ratio (S/N) lower than that obtained by non-VS processing.

We experienced the abovementioned near-surface challenges in this time-lapse monitoring feasibility dataset where the spurious events remain as key issues in the VS processing. To address these challenges, we propose a new VS redatuming workflow to improve the processing quality in the steps of traditional processing: offset stacking, crosscorrelation, and unbalanced illumination. During offset stacking, we developed a data-driven method taking into account the non-stationary behavior via a selective weighting scheme over varying offsets. We adaptively adjust the stacking weights to honor stationary-source contributions and suppressing the non-stationary effects. In terms of crosscorrelation, we perform VS in the wavelet-wavenumber domain to exploit the non-stationary characteristics of severe noise effects. More specifically, this step maps data from the time-offset (TX) domain into the time-frequency (TF) or time-frequency-wavenumber (TFK) domain in which the signal and noise can be better separated. The original signal phase is preserved, but the amplitudes are filtered to suppress VS spurious events. This method allows effective noise suppression and high-resolution separation of scattered energy and surface waves using TF or TFK filtering of the wavelet-correlation coefficients. We then construct and apply a matched filter iteratively to correct and recover the 3D amplitude radiation pattern for each buried receiver. Specifically, this matched filter is designed to involve the 3D FK amplitude spectrum of the direct arrivals for each VS, estimating the ideal spectrum of the direct P-wave by approximating the near surface with a homogenous model, and iteratively solving for the filter coefficients to minimize the misfit function of the radiation pattern between the computed and the ideal spectrum. Next, the P-wave velocity for the local homogeneous model is updated. After the matched filter, the output VS, is expected to possess isotropic radiation patterns and provide balanced illuminations. These renovated steps compensate for near-surface complexity and reduce time-lapse noise, and therefore produces redatumed data with fewer artifacts. The output VS records are then stacked to produce final image.

The proposed VS workflow consists of the following steps: (a.) estimate and update a 3D matched filter; (b.) applying the 3D matched filter to the direct-arrival wavefields; (c.) store the corrected direct arrivals if misfit criteria is satisfied; (d.) transform direct arrivals and reflections into the wavelet domain; (e.) crosscorrelate the wavelet coefficients; (f.) denoise spurious events and suppress surface waves via TFK domain filtering; (g.) invert the wavelet-transform filtered data back to the TX domain; (h.) estimate the predicted upgoing wavefields and compute the cross-coherence weight with the original upgoing wavefields; (i.) data-driven offset stack the cross-correlation gathers over the available offsets with the obtained weight. The output VS records are then stacked to produce a final image. A diagram summarizing the abovementioned workflows is given at the end. However, not all of the overburden related effects can be removed from the estimated reflectivity using the proposed workflow in which case deconvolution of full downgoing wavefields with sufficient wavefield separations<sup>9</sup> is required.

## Review of the Conventional VS and the Challenges

To investigate the reasons of the degradation of the VS responses in the presence of complex near-surface, we briefly review the VS methodology here. It essentially involves crosscorrelating the seismic responses observed at different receivers to the same source and summing over the contributing sources, which can be written as<sup>21</sup>



**Figure 1.** An illustration of virtual-source redatuming in the presence of near-surface complexity. Two surface sources, with the same distance (blue solid lines) with respect to the VS, generate two direct arrivals recorded by the buried receivers. The straight ray path is broken into a set of piecewise rays (yellow solid lines) and results in inequivalent angle coverage. The surface multiples, near-surface reverberations (yellow dashed lines) are also introduced to contaminate direct arrivals (yellow solid lines). The conventional stacking operator (Gaussian function  $g(r)$ ) is shown as a purple curve on the top of surface. The surface sources  $r_S$  generates direct arrivals recorded by the buried receivers  $r_A$  to  $r_B$ , which make a positive contribution to VS. The near-surface complexity breaks the raypath into a set of piecewise rays (green solid lines) with potentially complex raypaths. The incorrectly projecting source  $r_{S'}$  (green dashed lines) does not share a common raypath, and therefore makes a nonstationary contribution to the VS. Conventional weighting  $g(r)$  assigns inappropriate weights to surface sources and degrades VS responses.

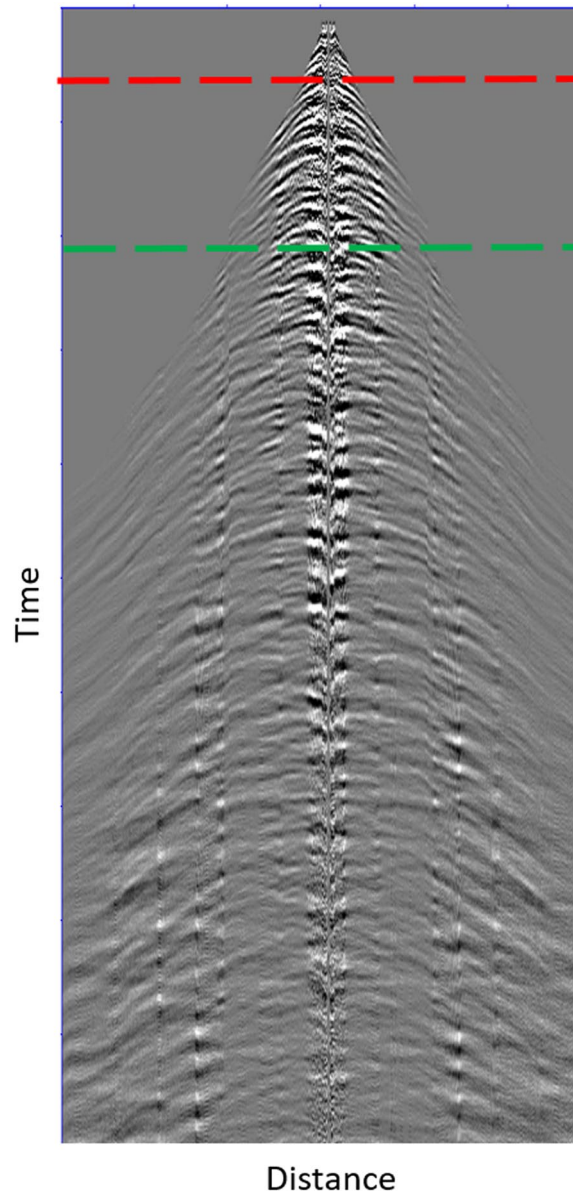
$$V(r_B|r_A; t) = \sum_{r_S} g(r) D(r_A|r_S; t) \otimes U(r_B|r_S; t) \quad (1)$$

where  $\otimes$  denotes crosscorrelation.  $r_A$ ,  $r_B$  and  $r_S$  denote the spatial locations of the two receivers at A and B, and at the source location S, respectively.  $V(r_B|r_A; t)$  is the resulting interferometric data with the time lag  $t$  at receiver  $r_B$  when  $r_A$  is treated as a virtual source.  $D(r_A|r_S; t)$  represent the direct-downgoing wavefields recorded at buried receivers  $r_A$ , while  $U(r_B|r_S; t)$  represents the reflected-upgoing wavefields at  $r_B$ .  $g(r)$  is the weight operator which is a simple Gaussian function centered at the VS location:  $g(r) = e^{-r^2/2R^2} / \sqrt{2\pi}R$ , where  $R$  is the maximum offset,  $r$  is the offset between the source  $r_S$  and the buried receiver  $r_A$ .  $g(r)$  ranges from 1 at the receiver location to 0 at the maximum offset.

Figure 1 summarizes underlying principle that the near-surface complexity degrades the VS response. Direct arrivals  $D(r_A|r_{S,1})$  and  $D(r_A|r_{S,2})$  excite from surface sources  $r_{S,1}$  and  $r_{S,2}$ , and are recorded by a buried receiver  $r_A$ .  $r_{S,1}$  and  $r_{S,2}$  that have the same offset  $r$  (light blue solid lines) with respect to  $r_A$ . In the absence of the wavefield separation capability, we treat the direct arrivals as the downgoing wavefields and the late-arriving reflections as the upgoing wavefields, respectively. Figure 2 shows a typical record. We consider the unprocessed near-offset data as the downgoing wave, usually at offsets less than 30 m and within 100 ms (the time section above the red dash lines). These values were estimated from our experiences that the early arrivals are relatively unpoluted and is still dominated by the direct P-wave. The later-arriving reflections ( $>300$  ms, the time section below the green dashed lines) at all offsets are used as approximations to the upgoing waves. The horizontal distance is 2400 m and the vertical time axis is from 0 s to 2 s ( $\sim 3300$  m).

There are mainly three challenges that need to be addressed:

- (1) Nonstationary offset stacking. Figure 1 illustrates a scenario in which  $D(r_A|r_{S,1})$  and  $U(r_B|r_{S,1})$  share a common raypath from shot  $r_{S,1}$ , and will have the correct compensation for the phase associated with the near-surface propagation by crosscorrelating with  $U(r_B|r_{S,1})$ , and making a stationary contribution to  $V(r_B|r_A)$ . Therefore, it should be added into  $V(r_B|r_A)$  with maximum weight. However, since the weight  $g(r)$  (purple bell curve) is not data driven, it provides a minor weight during offset stacking. In contrast to  $r_{S,1}$ , the near-offset shot  $r_{S,1'}$ , linear mapped from the buried VS to the surface location, does not share the common raypath (green dash lines) with  $r_A$  and therefore has deconstructive contributions. From the above analysis, we observe that weights  $g(r)$  based on the surface location (purple bell curve) could assign inappropriate weights and lead to non-stationary contributions. Other shots such as  $r_{S,2}$  may also qualify for contribution to  $V(r_B|r_A)$  constructively, as long as its source raypath to  $r_A$  can cancel with the raypath to  $r_B$  effectively.



**Figure 2.** A geophone shot record (with source location located in the center of the array) after noise removal in the common-receiver domain. The red dashed line indicates the time window of direct arrivals, and the green dashed line is for the reflections.

- (2) Multi-wave mode contamination. The intricate overburden introduces near-surface reverberations (yellow dashed lines), and S-wave modes which contaminate direct arrivals (yellow solid lines) regardless of careful time-windowing and spatial energy tapering. Considering this, each of the up- and downgoing-wavefields can be written as a summation of different modes:

$$\begin{aligned} D(r_A|r_S; t) &= D_{LP}(r_A|r_S; t) + D_{LM}(r_A|r_S; t) + D_{LS}(r_A|r_S; t) + \dots \\ U(r_B|r_S; t) &= U_{LP}(r_B|r_S; t) + U_{LM}(r_B|r_S; t) + U_{LS}(r_B|r_S; t) + \dots, \end{aligned} \quad (2)$$

where subscripts  $P, M, S$  represent direct  $P$ -arrivals, near-surface multiples, and S-waves modes, respectively. We add subscripts  $L$  (land data) in front of  $P, M, S$  to avoid subscript conflicts in other sections.  $(D_{LP}, U_{LP}), (D_{LM}, U_{LM}), (D_{LS}$  and  $U_{LS})$  are the associated up- and downgoing components, respectively. Only the crosscorrelation of the downgoing  $D_{LP}(r_A|r_S; t)$  with the upgoing  $U_{LP}(r_B|r_S; t)$  in Eq. (2) would produce a kinematically correct event. Other terms,  $D_i(r_A|r_S; t) * U_j(r_B|r_S; t)$  where  $i$  is not the same as  $j$ , would typically generate spurious events.

- (3) Unbalanced illuminations. The straight ray path breaks into a set of piecewise rays (yellow and green solid lines). This generates an inequivalent angle coverages ( $\theta_1$  of  $D(r_A|r_{S,1}) \neq \theta_2$  of  $D(r_A|r_{S,2})$ ), although  $r_A$  is covered by the same offset therefore the VS illumination is heavily biased beneath the overburden.

## Method: A Comprehensive VS Workflow to Tackle Complex Near-Surface

**Step 0: Review of Continuous wavelet transform.** At the beginning of the methods section, we briefly review the fundamentals of continuous wavelet transformation (CWT) as it underpins the following sections heavily. The CWT is used to decompose a signal into wavelets. Wavelets are small oscillations that are highly localized in time. While the Fourier Transformation decomposes a signal into infinite length sines and cosines, it effectively loses all time-localization information. The CWT's basic functions are scaled and shifted versions of the time-localized mother wavelet. The CWT is used to construct the TF representation of a signal and offers good time and frequency localization<sup>22–28</sup>. The CWT is an excellent tool for mapping changing properties of non-stationary seismic signals.

For a given 1D seismic trace  $D(t)$ , the forward CWT expands it from 1D to a 2D signal:

$$\Psi^\varphi(f_\alpha, \tau) = \frac{1}{\sqrt{f_\alpha}} \int_{-\infty}^{\infty} D(t) \varphi\left(\frac{t - \tau}{f_\alpha}\right) dt, \quad (3)$$

where  $\Psi^\varphi(f_\alpha, \tau)$  is a 2D wavelet coefficient, and  $\varphi(t)$  denotes the Ricker source wavelet. The symbols  $f_\alpha$  and  $\tau$  are Ricker wavelet frequency and local time delay, respectively. The inverse wavelet transformation brings the 2D signal back to the time domain:

$$D(t) = \frac{1}{C_\varphi} \int_0^\infty \int_{-\infty}^{\infty} \Psi^\varphi(f_\alpha, \tau) \frac{1}{\sqrt{f_\alpha}} \varphi\left(\frac{t - \tau}{f_\alpha}\right) d\tau \frac{df_\alpha}{f_\alpha^2}, \quad (4)$$

where  $C_\varphi$  is a frequency scaling factor. A similar process can be applied to 2D seismic gather  $D(r, t)$ , in which multi-dimension transformation is given by

$$\Psi^\varphi(f_\alpha, \tau, k_r) = \frac{1}{\sqrt{f_\alpha}} \int_{-\infty}^{\infty} D(r, t) \varphi\left(\frac{t - \tau}{f_\alpha}\right) dt \int_0^R e^{-ik_r r} dr \quad (5)$$

where  $k_r$  and  $r$  represent wavenumber and spatial distance, respectively, and  $\Psi^\varphi(f_\alpha, \tau, k_r)$  denotes the 3D wavelet coefficients including times  $\tau$ , frequencies  $f_\alpha$ , and wavenumbers  $k_r$ .

**Step 1: Data-driven weight offset stacking.** The design of stacking weights for the VS have rarely been fully studied in the existing literature (Zhao, *et al.*<sup>21</sup>). In contrast to previous works, this study proposes a fully data-driven weighting approach and wavelet domain implementation. The proposed new approach is formulated as

$$V_{w(r_s)}(r_B|r_A; t) = \sum_{r_s} w(r_s) \{D(r_A|r_S; t) \otimes U(r_B|r_S; t)\}. \quad (6)$$

Our goal is to adaptively choose weighting coefficients  $w(r_s)$  for all sources within the available offsets such that  $V_{w(r_s)}(r_B|r_A; t)$  can be computed with less contribution from non-stationary raypaths. The coefficients  $w_{r_s}$  are entirely data-driven and directly computed from downgoing and upgoing wavefields. In order to evaluate the quality of individual  $r_s$  contributions to the retrieved  $V_{w(r_s)}(r_B|r_A; t)$ , we use  $V(r_B|r_A; t)$  (Eq. (1)) to predict upgoing wavefields

$$U_{pred}(r_B|r_S; t) = D(r_A|r_S; t) * V(r_B|r_A; t), \quad (7)$$

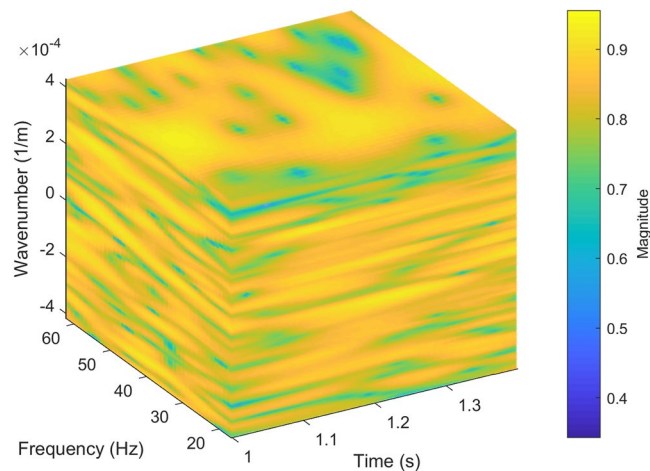
where  $*$  denotes convolution. Then we use the quality of the upgoing prediction  $U_{pred}(r_B|r_S; t)$  to obtain weight coefficients. For time-lapse monitoring purposes, we customize Eq. (7) to be a target-oriented function focusing on the target reservoir. This function is designed to include a variety of seismic attributes such as  $\tau, f_\alpha, k_r$  (TFK) of the target pre-stack events. Therefore we first perform transformation (3) for each upgoing trace  $\Psi^\varphi(f_\alpha, \tau)$  to obtain the 2D wavelet coefficients. To improve the robustness of the obtained weights, Eq. (5) is adapted with  $U_{pred}(r_B|r_S; t)$  along the space dimension:

$$\Psi_{U_{pred}}^\varphi(f_\alpha, \tau, k_r) = \frac{1}{\sqrt{f_\alpha}} \int_{-\infty}^{\infty} U_{pred}(r_B|r_S; t) \varphi\left(\frac{t - \tau}{f_\alpha}\right) dt \int_0^R e^{-ik_r r} dr \quad (8)$$

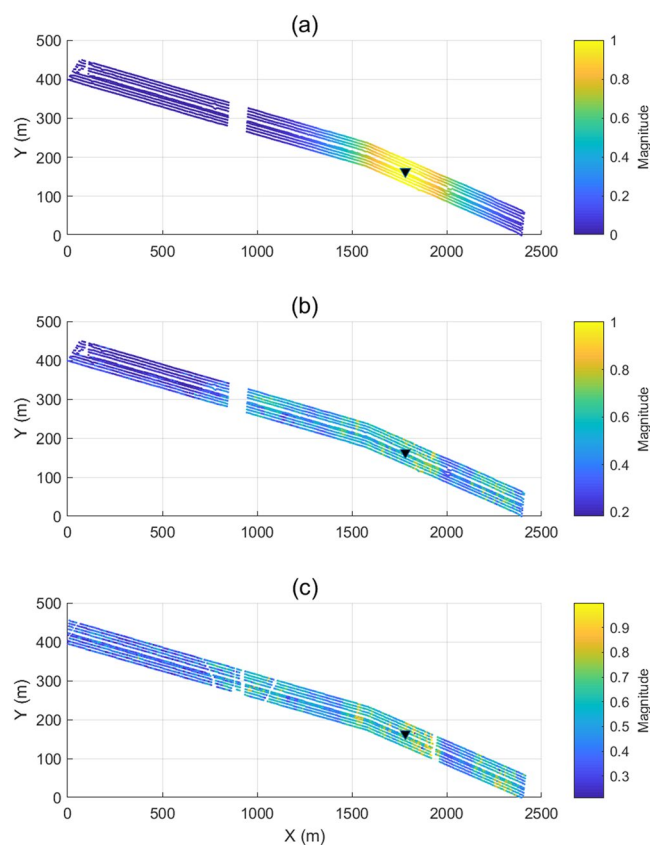
The initial weighting coefficients is computed as follows

$$w_{r_s} \equiv \sum_{k_r} \sum_{\tau} \sum_{f_\alpha} w_{r_s}(f_\alpha, \tau, k_r) = \sum_{k_r} \sum_{\tau} \sum_{f_\alpha} \frac{|\Psi_{U_{pred}}^\varphi(f_\alpha, \tau, k_r) \Psi_{U_{pred}}^\varphi(f_\alpha, \tau, k_r)|}{|\Psi_{U_{pred}}^\varphi(f_\alpha, \tau, k_r)| |\Psi_{U_{pred}}^\varphi(f_\alpha, \tau, k_r)|}. \quad (9)$$

Equation (9) represents the normalized cross-spectrum  $w_{r_s}$  between the predicted  $U_{pred}(r_B|r_S)$  and the recorded  $U(r_B|r_S)$  in the TFK domain Figure 3 illustrates a cross-spectrum TFK cube from a selected event. We may restrict the range of  $\Psi_{U_{pred}}^\varphi(f_\alpha, \tau, k_r)$  and  $\Psi_{U_{pred}}^\varphi(f_\alpha, \tau, k_r)$  to retain the target features, and mask other unassociated events. As discussed in the introduction section, the target horizon (2000 m) is at about 1 s–1.4 s, and has an effective bandwidth of about 20 to 60 Hz<sup>1</sup>. Therefore we limit the ranges of  $\tau$  and  $f_\alpha$ , and sum over the corresponding frequencies, time lags, and wavenumbers to produce  $w_{r_s}$ . Equation (9) is then applied to these time-lapse datasets.

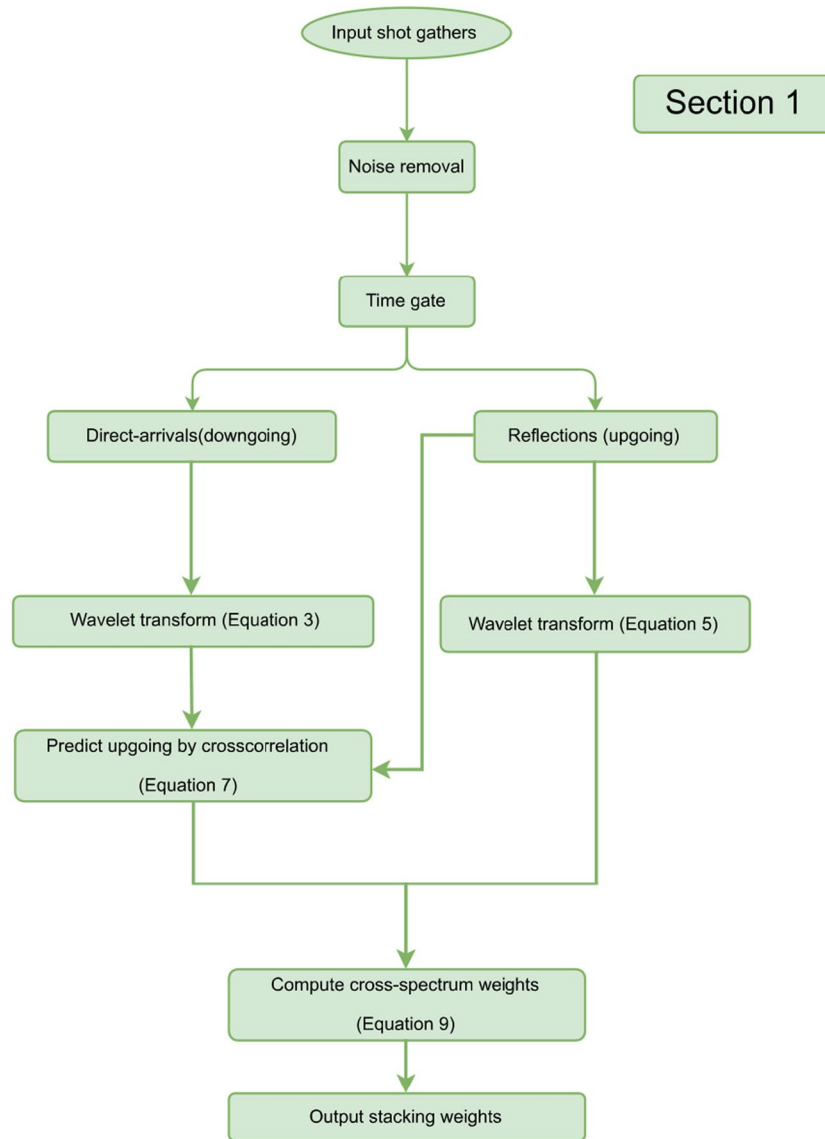


**Figure 3.** The 3D cross-coherence spectrum (display in log scale) of the TFK cube of the target reflections, computed via Eq. (5).



**Figure 4.** Stacking weights from 9 shot lines of surface sources that contribute to VS (#60) including those from (a) a conventional stacking operator of survey S1 and S11, (b) from the offset stack of survey S1, and (c) from the offset stack of survey S11.

The resulting stacking weights for the conventional and the new stacking are compared in Fig. 4. The offset stacking of survey S1 (Fig. 4b) and survey S11 (Fig. 4c) offers data-driven, and survey dependent weights, therefore, respond adaptively to the changes between times in the near-surface at a given VS. In contrast, the conventional spatial taper (Fig. 4a) only provides a static weight distribution for all surveys. Generally, the offset weights decrease with offset and tend to be smaller than the equivalent weights from the conventional taper. Among all surface sources to each of the selected VS locations, the offset weights show similar contribution patterns for different surveys, but also adapt to time-lapse changes shown in the minor pattern dissimilarity. It also can be observed that the near-offset shots have larger contributions. Due to the data-driven nature of the weighting



**Figure 5.** The summarized workflow is colored in green for the section of step 1. The input data is from the buried geophones, and the output comprises adaptive weights for VS offset stacking.

coefficients, they may be influenced by noise and experience large lateral variations. A smoothing function can be applied to stabilize the data-driven stacking process. We summarize this section of step 1 in a workflow, colored as green, in Fig. 5.

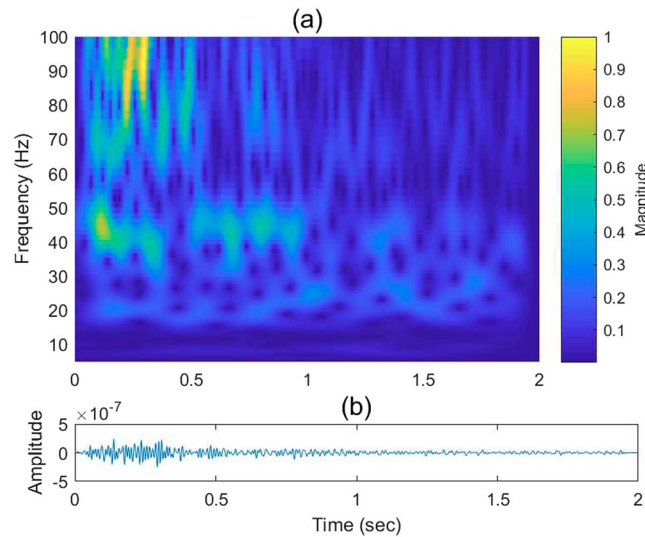
**Step 2: Wavelet-crosscorrelation filtering for multi-wave mode contamination.** Crosscorrelation can be used to determine the relative time delay between two seismic signals. The wavelet coefficients provide a local time and frequency distribution of the seismic traces. Equation (1) can be written in the form of the wavelet crosscorrelation<sup>29</sup>:

$$WV_{V(r_B|r_A)}(f_\alpha, \tau) = \sum_{r_s} g(r) \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \Psi_{D(r_A|r_s)}^\varphi(f_\alpha, t) \Psi_{U(r_B|r_s)}^\varphi(f_\alpha, t + \tau) dt \right\}, \quad (10)$$

where  $\Psi_D^\varphi(f_\alpha, t)$  and  $\Psi_U^\varphi(f_\alpha, t)$  are the wavelet coefficients of direct-downgoing  $D(r_A|r_s)$  and reflection-upgoing waves  $U(r_B|r_s)$ , respectively. The wavelet crosscorrelation function  $WV_{V(r_B|r_A)}(f_\alpha, \tau)$  allows the detection of nonstationary coherence structures and the potential time-lag between two seismic traces.

It can be shown that  $WV(f_\alpha, \tau)$ <sup>26</sup> can be related to the classical crosscorrelation  $V(r_B|r_A)$  in the Fourier domain:

$$WV(f_\alpha, \omega) = f_\alpha V(\omega) |\varphi(f_\alpha, \omega)|^2, \quad (11)$$



**Figure 6.** Field-data example showing (a) the wavelet coefficients of a VS trace in the TF domain via wavelet crosscorrelation and (b) the corresponding VS trace in the time domain.

$$|WV(f_\alpha, \omega)| = f_\alpha |V(\omega)| |\varphi(f_\alpha \omega)|^2, \text{ and,} \quad (12)$$

$$\angle WV(f_\alpha, \omega) = \angle V(\omega), \quad (13)$$

where  $WV(f_\alpha, \omega)$  is the wavelet crosscorrelation coefficient represented in the wavelet frequency and angular-frequency domain.  $V(\omega)$  and  $\varphi(f_\alpha \omega)$  denotes the cross-spectrum and mother wavelet in the Fourier domain, respectively. Equation (11) suggests that  $WV(f_\alpha, \omega)$  has the same phase spectrum as  $V(\omega)$ , and  $\angle$  represents the phase spectrum. In Eq. (12), the amplitude spectrum  $|WV(f_\alpha, \omega)|$  is the cross-spectrum  $|V(\omega)|$  weighted by the factor  $f_\alpha |\varphi(f_\alpha \omega)|^2$ . Note that  $WV(f_\alpha, \omega)$  is a complex-valued function. Subsequent filtering is applied to the amplitude spectrum only. The original phase of the recorded data is retained to honor the kinematics of the Green's function extracted from the VS process.

The 2D TF domain filter, denoted as  $H(f_\alpha, \tau)$ , is applied to the obtained wavelet crosscorrelation  $WV_i(f_\alpha, \tau)$ :

$$WV_o(f_\alpha, \tau) = H(f_\alpha, \tau) *_{\alpha, \tau} WV_i(f_\alpha, \tau), \quad (14)$$

Equation (14) denotes the 2D convolution operator, and  $WV_o(f_\alpha, \tau)$  is the output wavelet crosscorrelation in the TF domain.  $H(f_\alpha, \tau)$  may take various 2D filter forms. For the discussion here, we use the first derivative of a 2D Gaussian filter as an example:

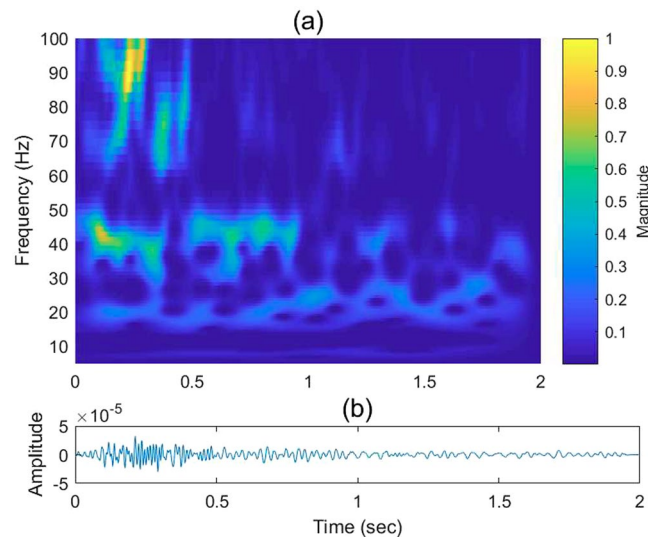
$$\begin{aligned} H(f_\alpha, \tau) &= \frac{\partial G_{f_\alpha}}{\partial f_\alpha} \frac{\partial G_\tau}{\partial \tau} \\ &= \left\{ \frac{f_\alpha}{\sigma_{f_\alpha}^2} \exp \left[ -\frac{1}{2} \left[ \left( \frac{f_\alpha}{\sigma_{f_\alpha}} \right)^2 + \left( \frac{\tau}{\sigma_\tau} \right)^2 \right] \right] \right\} \left\{ \frac{\tau}{\sigma_\tau^2} \exp \left[ -\frac{1}{2} \left[ \left( \frac{f_\alpha}{\sigma_{f_\alpha}} \right)^2 + \left( \frac{\tau}{\sigma_\tau} \right)^2 \right] \right] \right\} \\ &\triangleq H_{\sigma_{f_\alpha}}(f_\alpha) H_{\sigma_\tau}(\tau), \end{aligned} \quad (15)$$

where  $\sigma_{f_\alpha}$  and  $\sigma_\tau$  are the standard deviations along  $f_\alpha, \tau$  axes that parameterize the filter. Small  $\sigma_{f_\alpha}$  and  $\sigma_\tau$  values lead to higher localization resolution along the  $f_\alpha, \tau$  axes. The Gaussian filter can be replaced by other filters depending on the purpose of processing. As an example, the resulting wavelet coefficients before and after TF filtering are illustrated in Figs 6 and 7, respectively. The strong energy events (yellow-orange patches) represent the reflections generated by the target reflectors, whereas the weaker blue-green events with random shapes represent scattered noise. Removing the spurious events is desirable to enhance the overall S/N prior to correlation. Figure 7a demonstrates that  $H(f_\alpha, \tau)$  strengthens these deep targets by suppressing the background noise. The inverted time series before and after  $H(f_\alpha, \tau)$  filtering is shown in Figs 6b and 7b, respectively.

We extend 2D TF  $H(f_\alpha, \tau)$  into 3D TFK  $H(f_\alpha, \tau, k)$  filtering, which is expected to mitigate coherent noise in space/wavenumber dimension. Extending the 2D TF filtering given in Eq. (14), we propose a 3D filtering in the time-frequency-space (TFX) domain as

$$WV_o(f_\alpha, \tau, x) = H(f_\alpha, \tau, x) *_{f_\alpha, \tau} WV_i(f_\alpha, \tau, x) \quad (16)$$





**Figure 7.** As shown in Fig. 6, but after soft-thresholding filtering of the amplitude spectrum in a sliding window.

As a simple example, a 3D filter for  $H(f_\alpha, \tau, k)$  can be expressed as:

$$\begin{aligned} H(f_\alpha, \tau, k) &= \frac{\partial G_{f_\alpha}}{\partial \alpha} \frac{\partial G_\tau}{\partial \tau} H_{k\alpha}(f_\alpha, k) \\ &= H_{\sigma_{f_\alpha}}(f_\alpha) H_{\sigma_\tau}(\tau) H_{k f_\alpha}(f_\alpha, k), \end{aligned} \quad (17)$$

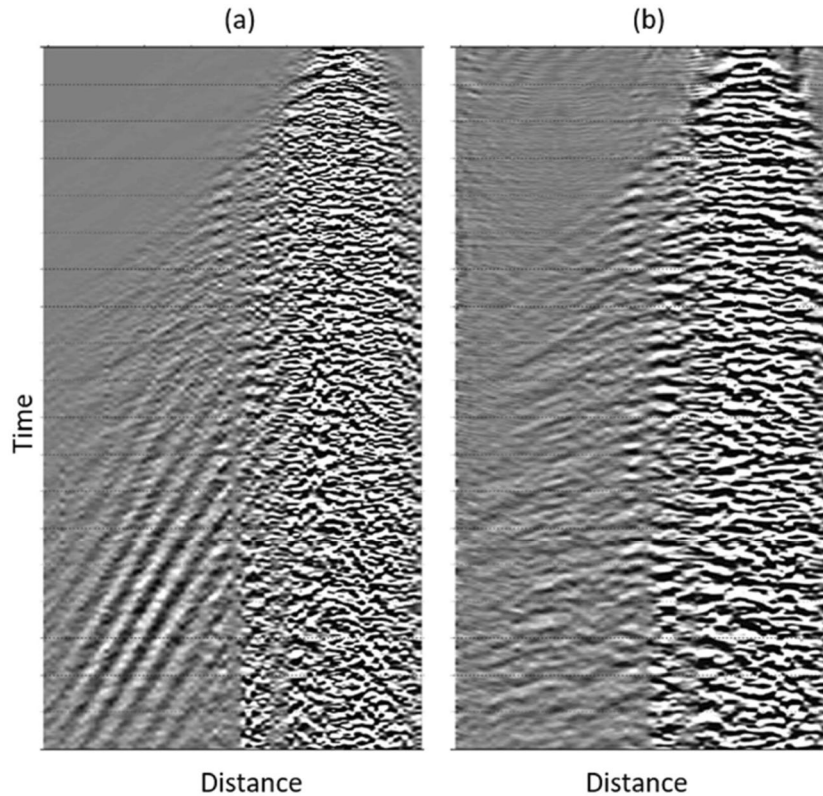
where  $H_{k\alpha}(f_\alpha, k)$  can be a simple  $f_\alpha, k$  bandpass filter allowing a particular propagation velocity to pass while filtering out other wavefield components. As another example, we apply the multi-channel filter  $H(f_\alpha, \tau, k)$  to process an entire VS shot gather contaminated by surface waves (Fig. 8a). A TFX cube is generated by gathering slices from each receiver in response to a common virtual shot to form the TFX cube. The data is then Fourier transformed along the receiver dimension to obtain the corresponding wavenumber spectra. Figure 8b shows the recovered signal after applying the inverse wavelet transform. Figure 8 demonstrates that the wavelet cross-correlation was able to eliminate residual surface waves with a simple TFX filter, where the traditional VS failed. We summarize this section of step 2 in a workflow, colored in blue, in Fig. 9.

**Step 3: Radiation-pattern correction for unbalanced illumination.** To address the unbalanced illumination issue, we used a method<sup>30</sup> which iteratively constructs and applies a matched filter for an ideal 3D amplitude spectrum for direct downgoing arrivals. The simplest approach to estimate an isotropic radiation pattern is to compute the Green's function of direct P-waves in a homogenous model<sup>31</sup>. This homogenous model is 2.5D and built independently for every VS, spanning from the buried receiver up to the surface with the radius consisting of the contributing sources. By targeting an optimal approximation to the direct P-wave arrivals before correlating with the reflections, the correct image associated with a cleaner direct P-wave can be enhanced, while near surface induced artifacts are suppressed. The following workflow describes radiation-pattern correction.

Estimate the initial homogeneous model. The length of the straight ray connecting source and receiver is divided by the automatically picked onset arrivals. The output is the velocity estimate for each source-receiver pair. The local initial model above each VS is then derived as the average of the calculated velocities. Using this model, the synthetic direct arrivals are generated using an analytical 3D acoustic Green's function<sup>31</sup>

$$D_p(r_A|r_S; t) = \frac{\varphi(t - r\chi)}{4\pi r} e^{-\frac{\omega t}{2Q}}, \quad (18)$$

where  $D_p(r_A|r_S; t)$  represents the output synthetic direct P-waves.  $D_p(r_A|r_S; t)$  is then converted into FK domain  $\tilde{D}_{p,r_A|r_S}(k_r, \omega)$ , where  $k_r$  and  $\omega$  are wavenumber and angular frequency, respectively. Tildes above the symbols indicate quantities in the FK domain. The FK amplitude spectrum of the field data  $\tilde{D}_{r_A|r_S}(k_r, \omega)$  can be computed similarly.  $\varphi(t - r\chi)$  still denotes the source Ricker wavelet (Eq. (3)),  $r$  is the distance between the surface source  $r_S$  to the buried VS  $r_A$  with the constant slowness  $\chi$  (the inverse of P-wave velocity).  $1/4\pi r$  is the geometrical spreading of a spherical wavefield in homogenous medium.  $e^{-\frac{\omega t}{2Q}}$  represents the attenuation term of seismic amplitude of this analytical 3D acoustic Green's function, where  $Q$  is a constant seismic quality factor derived from the data to match waveform amplitudes, and estimated from a nearby acoustic log. The dominant frequency is 35 Hz to match the wavelet pattern of the target reservoir reflection. We solve Eq. (18) using the initial homogeneous velocity, and illustrate the target synthetic responses in Fig. 10.



**Figure 8.** A VS shot gather obtained from (a) crosscorrelation and (b) wavelet crosscorrelation followed by TF and TFK filtering.

Estimate the matched filter so that the filtered recorded wavefields match with the modeled wavefields. A cost function in the FK domain is designed to shape the radiation pattern of the field data to match that of the modeled synthetic data as follows

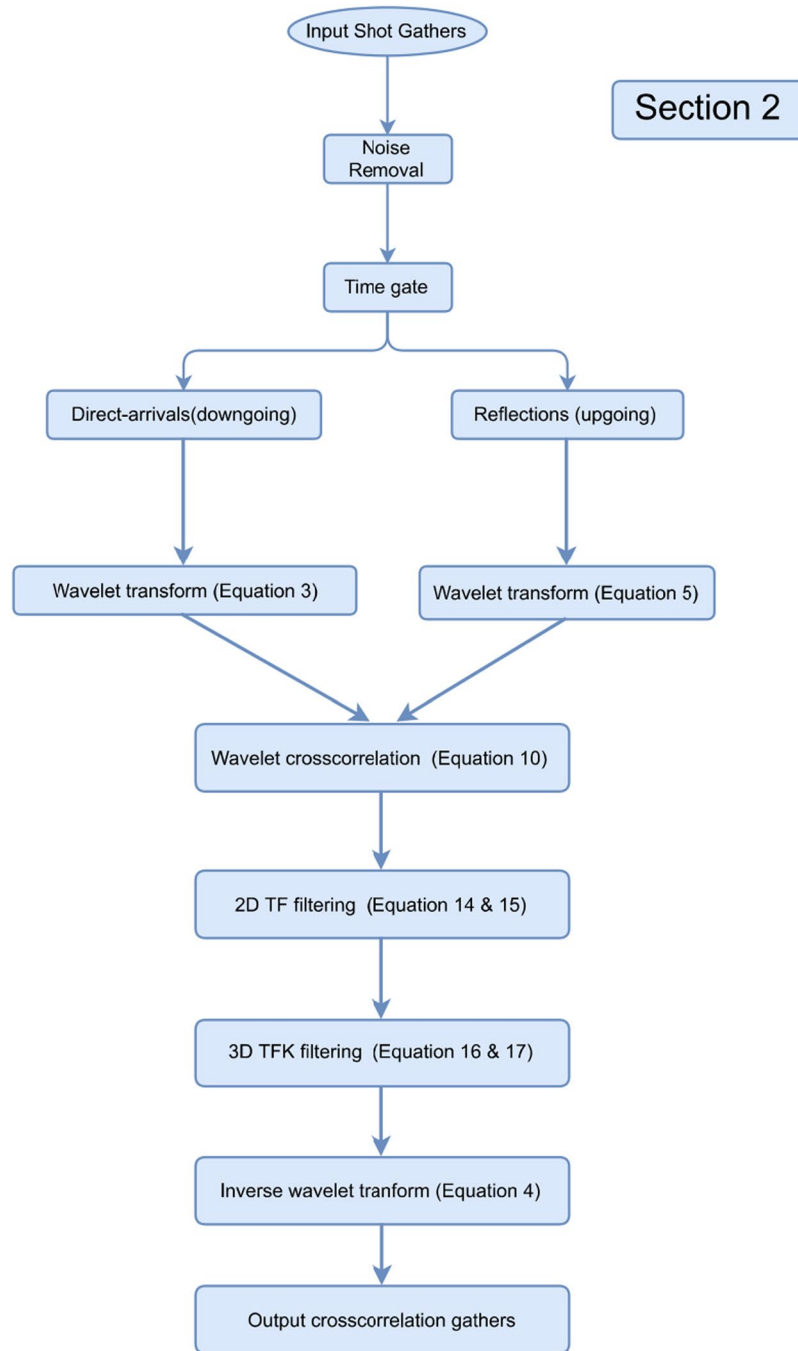
$$J(\tilde{\mathbf{H}}) = \min_{\tilde{\mathbf{H}}} \left\| \tilde{\mathbf{D}}_{r_A} \tilde{\mathbf{H}} - \tilde{\mathbf{D}}_{p,r_A} \right\|_2^2, \tag{19}$$

where  $\tilde{\mathbf{D}}_{r_A}$  and  $\tilde{\mathbf{D}}_{p,r_A}$  are the vector representations (including contributing surface sources  $r_s$ ) of the 3D amplitude spectrum of  $\tilde{D}_{r_A|r_s}(k_r, \omega)$  and  $\tilde{D}_{p,r_A|r_s}(k_r, \omega)$  associated with the field data  $D(r_A|r_s; t)$  and the synthetic  $D_p(r_A|r_s; t)$ , respectively. The corresponding matched filter  $\tilde{\mathbf{H}}$  is an L2 norm constrained least-square solution, and  $J(\tilde{\mathbf{H}})$  is the corresponding cost function. We define  $\tilde{\mathbf{H}}$  as vector multiplication in the FK domain, whereas the matched filter is traditionally defined as a convolution in the time-space (TX) domain. Since we operate on the amplitude spectrum and leave the phase intact,  $\tilde{\mathbf{D}}_{r_A}$ ,  $\tilde{\mathbf{D}}_{p,r_A}$ , and  $\tilde{\mathbf{H}}$  are real numbers. Noting that Eq. (19) is solved iteratively,  $\tilde{\mathbf{D}}_{p,r_A}$  is also updated with a new velocity model, and therefore is a function of the matched filter from the previous iteration.  $\tilde{\mathbf{H}}$  is updated iteratively using:

$$\tilde{\mathbf{H}}_{i+1} = \tilde{\mathbf{H}}_i + (\tilde{\mathbf{D}}_{r_A}^T \tilde{\mathbf{D}}_{r_A})^{-1} \tilde{\mathbf{D}}_{r_A}^T (\tilde{\mathbf{D}}_{p,r_A} - \tilde{\mathbf{D}}_{r_A} \tilde{\mathbf{H}}_i). \tag{20}$$

This is an iterative process since  $\tilde{\mathbf{D}}_{p,r_A}$  is also updated as a function of  $\tilde{\mathbf{H}}$ . The matched filter for the 3D amplitude spectrum  $\tilde{\mathbf{D}}_{r_A}$  and  $\tilde{\mathbf{D}}_{p,r_A}$  may be considered as a zero-phase FK filter, enforcing an isotropic radiation pattern associated with the desired direct P-wave, while leaving the phase intact. Since this field example has a good initial model estimation, only five iterations are needed for the inversion of the velocity and the matched filter to converge. The stopping criteria is the summation of the misfit less than  $1e^{-2}$ . Figure 11 illustrates the gated original direct arrivals of the 3D field common-receiver array (aperture 30 m at 7.5 m sampling interval).

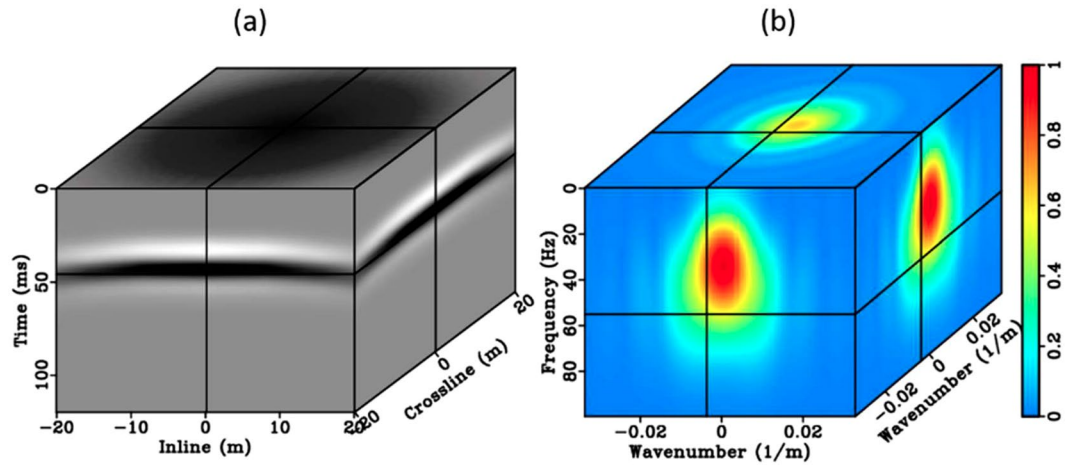
Update the homogeneous model velocity. The homogeneous overburden model is iteratively updated, and coupled with the matched filter update. The velocity model is updated by solving the same minimization cost function, Eq. (20), but with the matched filter set to the value obtained from the previous iteration and optimized over slowness  $\chi$ . After the initial model is built, the slowness can be updated iteratively



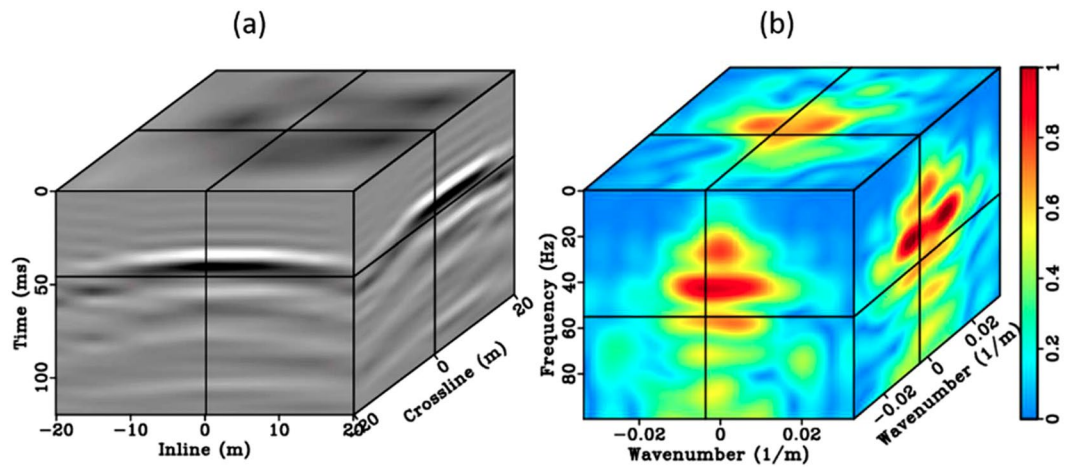
**Figure 9.** The summarized workflow is colored in blue for the section of step 2. The input data is from the buried geophones, and the output comprises crosscorrelation gather after wavelet filtering.

$$\begin{aligned}
 \chi_{i+1} &= \chi_i - \alpha \langle \nabla \mathbf{g}_{\chi_i} \rangle \\
 \nabla \mathbf{g}_{\chi_i} &= \left( \frac{\omega \mathbf{r}}{2Q} - \frac{\varphi'(\mathbf{t} - \mathbf{r}\chi_i)\mathbf{r}}{\varphi(\mathbf{t} - \mathbf{r}\chi_i)} \right) \cdot \mathbf{F}_{t,r \rightarrow \omega, k_r}^{-1} \{ \tilde{\mathbf{e}}^{j\varphi} [ \tilde{\mathbf{D}}_{P, r_A}^T (\tilde{\mathbf{D}}_{r_A} \tilde{\mathbf{H}} - \tilde{\mathbf{D}}_{P, r_A}) ] \},
 \end{aligned}
 \tag{21}$$

where  $\chi_i$  is the constant slowness model, and  $\nabla \mathbf{g}_{\chi_i}$  is the gradient for the  $i_{th}$  iteration. The FK phase term of the original direct wavefields  $\tilde{\mathbf{e}}^{j\varphi}$  is retained through the workflow, while the amplitude spectrum is iteratively updated with  $\tilde{\mathbf{H}}$ . An inverse multi-dimension Fourier transformation  $\mathbf{F}_{t,r \rightarrow \omega, k_r}^{-1}$  brings data from the FK to TX domain.  $\varphi(\mathbf{t} - \mathbf{r}\chi_i)$  still denotes the Ricker source wavelets (Eq. (18)) in a vector notation (including all contributing surface sources) with respect to the VS. Since  $\chi_i$  is a constant value, symbol  $\langle \rangle$  simply averages the gradient



**Figure 10.** The target wavefield generated from the homogenous model (a) and its 3D FK spectrum (b).



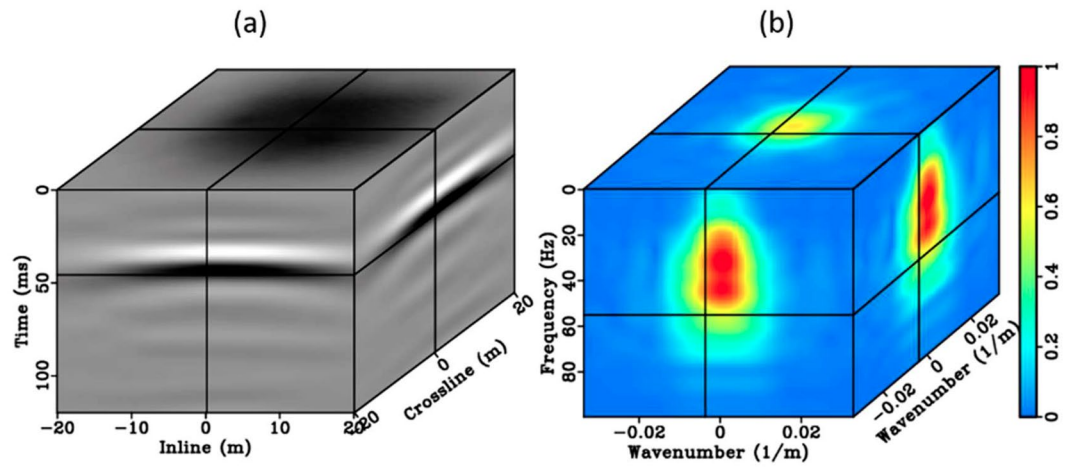
**Figure 11.** 3D representation of the original direct wavefield (a) and its 3D FK spectrum (b).

$\nabla \mathbf{g}_{\chi_i}$  for all contributing sources. We adapt a linear search to determine the step length  $\alpha$ , and  $\chi_i$  is then iteratively updated in TX domain. An updated  $\chi_i$  is plugged back to Eq. (18), and repeats until the iterations are finished.

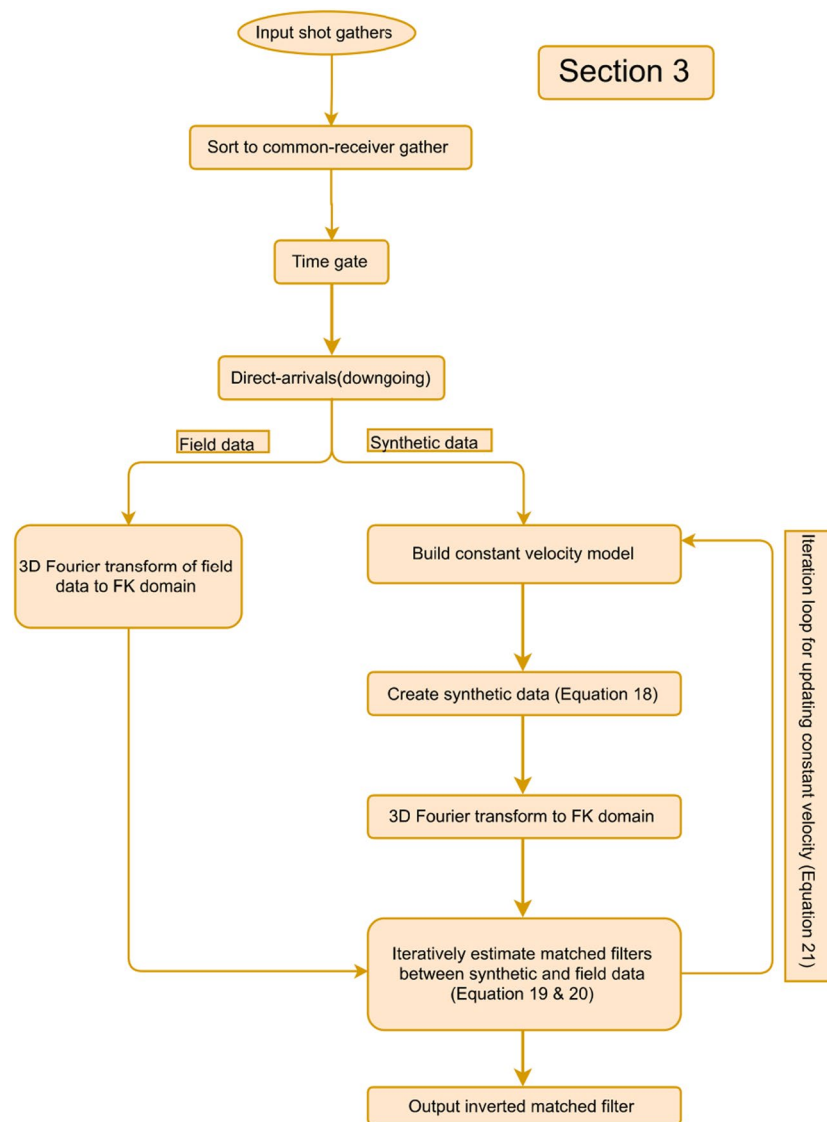
Apply the estimated matched filter to direct wavefields and perform VS. With a few iterations,  $\tilde{\mathbf{H}}$  is applied from Eq. (20)

$$\begin{aligned} V(r_B|r_A; \omega) &= \sum_{r_s} g(r) \{ \mathbf{D}^h(r_A|r_s; \omega) \mathbf{U}(r_B|r_s; \omega) \} \\ &= \sum_{r_s} g(r) \{ (\mathbf{F}_{r \rightarrow k}^{-1} \{ \tilde{\mathbf{e}}^{j\varphi} [ \tilde{\mathbf{D}}_{r_A|r_s} \tilde{\mathbf{H}} ] \})^h \mathbf{U}(r_B|r_s; \omega) \}. \end{aligned} \tag{22}$$

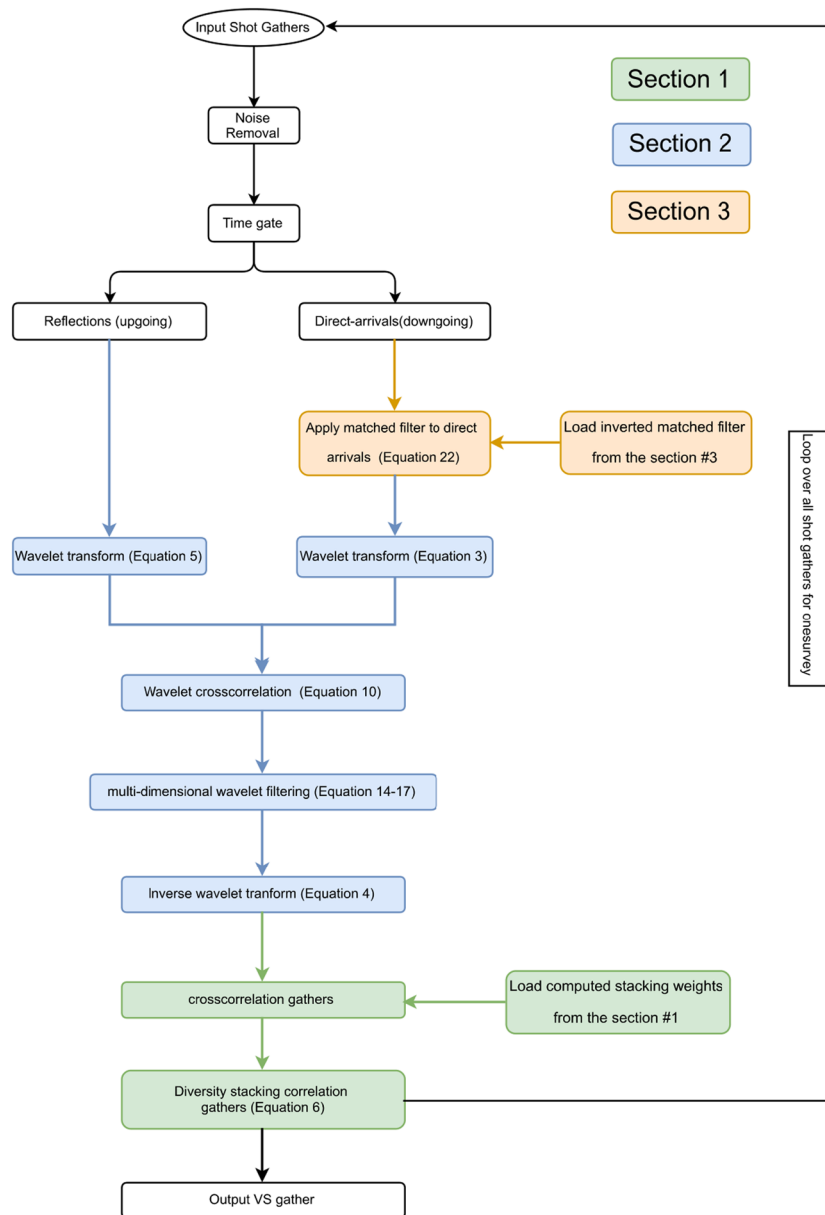
$V(r_B|r_A; \omega)$  is the final output with radiation pattern correction, and a final Fourier transformation back to  $V(r_B|r_A; t)$ . Again  $\tilde{\mathbf{e}}^{j\varphi}$  is isolated from the filter  $\tilde{\mathbf{H}}$  to maintain its originality. The adaptively updated matched filter is only applied to the direct-downgoing wavefield prior to crosscorrelation with the upgoing wavefield. Ideally, the matched filter should not affect any time-lapse signal carried by the reflected wavefield after crosscorrelation. Figure 12 is the matched filter  $\tilde{\mathbf{H}}$  output after the final iteration. In this example, first-arrival picks provide velocity updates with sufficient quality so that the output from the matched filter after five iterations already provides reasonable results. In the time domain, the iteratively computed matched filter effectively regularizes spectrum bandwidth across the gathers, suppresses the coda waves, and stabilizes the wavelet character (Fig. 12). The direct-downgoing wavefields (Fig. 11) gradually approach the ideal P-wave (Fig. 10) associated with the isotropic radiation pattern. The reason we are proposing an iterative inversion theme, is that there is always some error associated with auto-pickers. We summarize this section of step 3 as a workflow (orange) in Fig. 13.



**Figure 12.** The inverted direct wavefields generated with the matched filter applied (a) and its FK spectrum (b).



**Figure 13.** The summarized workflow is colored in orange for the section of step 3. The input data is from the buried geophones, and the output is comprised of an inverted matched filter after radiation-pattern correction.

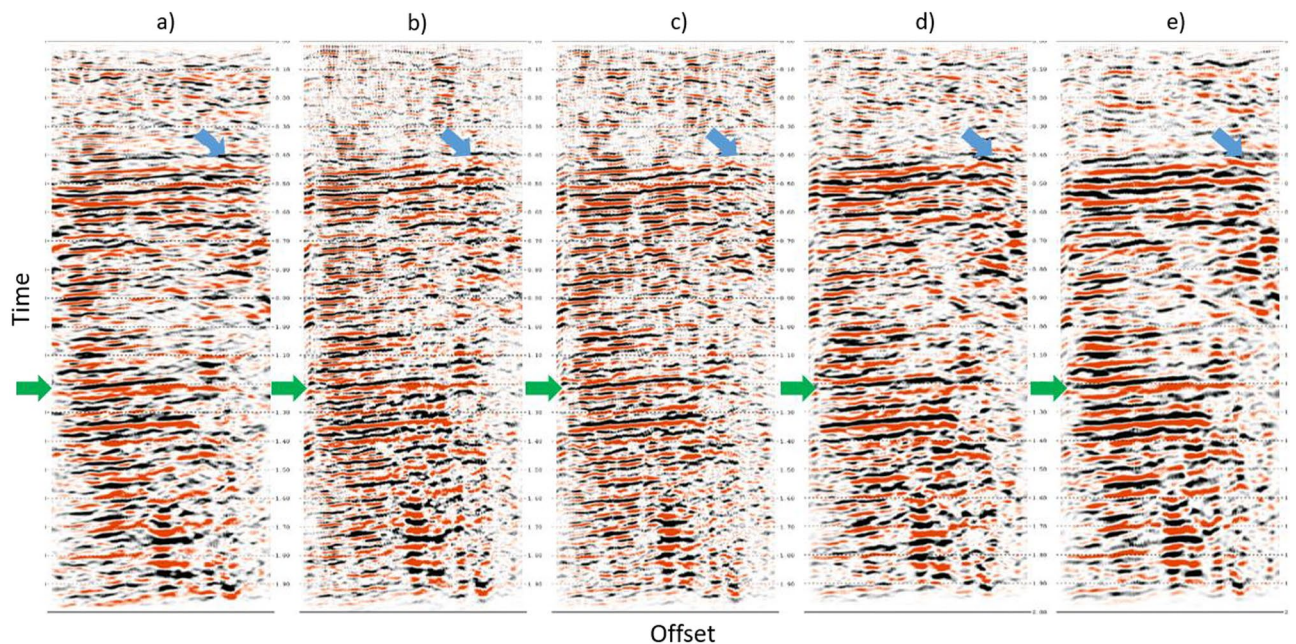


**Figure 14.** The overall workflow consist of step 1 (green), step 2 (blue), and step 3 (orange) for our comprehensive VS processing. The output VS gather combines all proposed three steps and to be used for subsequent imaging.

## Results

To assess the potential performance gain of this proposed workflow, the buried receiver data from the same series of field surveys was processed in three stages. These stages involved noise removal, VS redatuming, and CMP stacking. In this study we focus on the second stage, where four estimates of the proposed workflows were evaluated independently for each VS, including conventional procedures, using the offset stacking (the step 1), plus the wavelet-correlation filter (the steps 1 + 2), and the radiation-pattern correction (the steps 1 + 2 + 3). Figure 14 summarizes this proposed comprehensive workflow by organizing Fig. 5 (green), Fig. 9 (blue), and Fig. 13 (orange) in an orderly manner. During the VS processing, an auto-picking algorithm selected the first arrivals and windowed the direct-arrival energy in a 60 ms time window. All four outputs of each VS were obtained in parallel from post-correlation gathers stacked within common offsets. A non-VS CMP stack was created as well for comparison.

We plot the non-VS and proposed VS stacks of a selected survey in Fig. 15. The target horizon and selected areas are marked with green and blue arrows, respectively. The non-VS control section (Fig. 15a) shows good signal continuity on the marked reflectors, whereas the conventional VS stack (Fig. 15b) appears noisier with less reflection continuity in some zones. The VS stack (Fig. 15c) using the offset stack (step 1) shows better signal-to-noise ratio (SNR) and better continuity when compared with the conventional stack. Taken one step



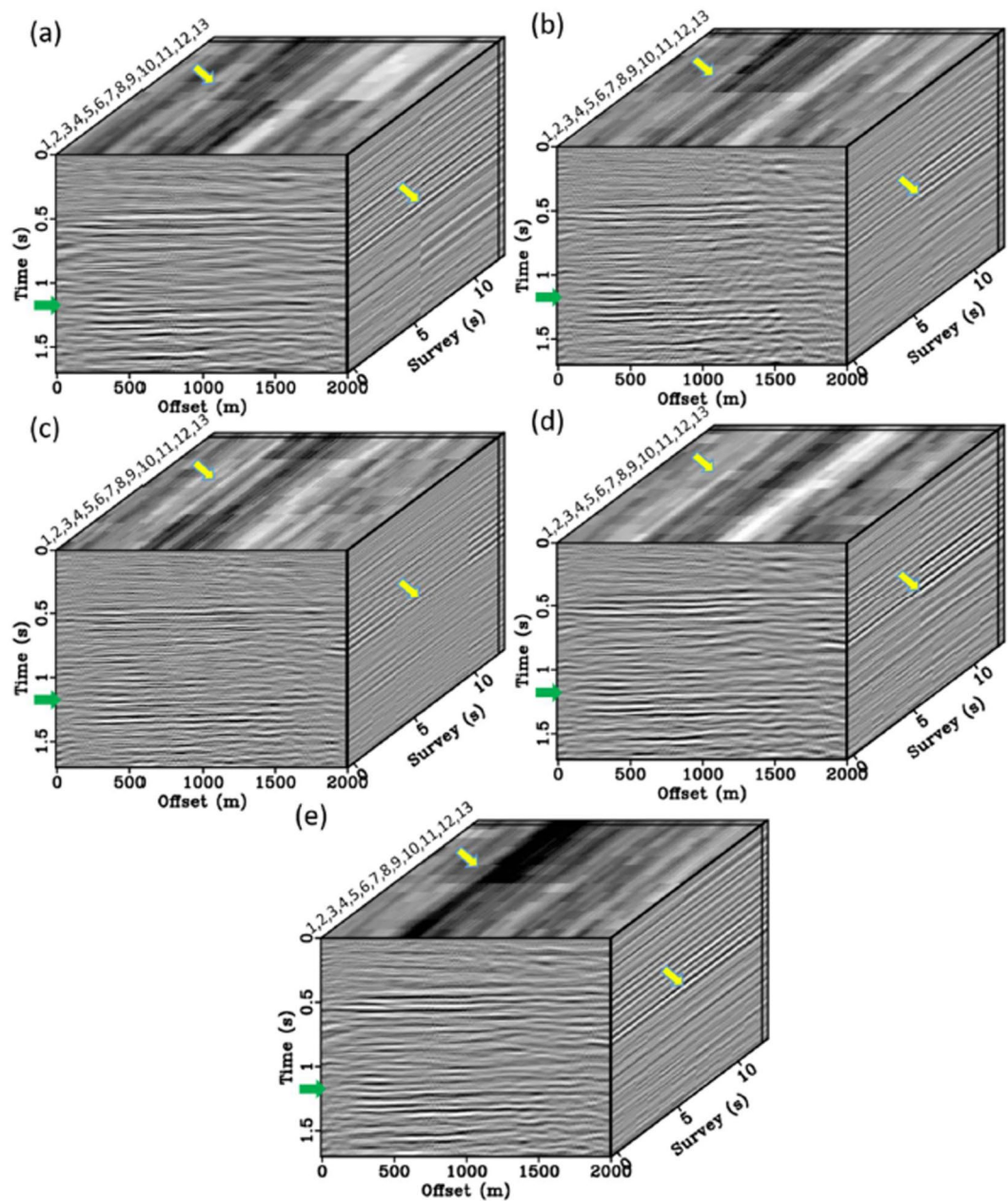
**Figure 15.** Survey 1 stacks obtained from (a) non-virtual source (non-VS), (b) conventional VS, (c) VS using the offset stack, (d) VS using the offset stack and the wavelet-crosscorrelation filtering, and (e) VS using the offset stack, the wavelet-crosscorrelation filtering and the radiation-pattern correction. The green and blue arrows indicate the SNR improvement on the target reflector and a selected shallow reflector, respectively. Panel (e) with the complete workflow has the best continuity of all five sections.

further, the VS stack (Fig. 15d), using the offset stack and the wavelet-correlation filter (the steps 1 + 2), demonstrates reduced contamination from the residual surface waves (steeply dipping, cross-cutting events) and scattering noise. Finally, the comprehensive VS stack (Fig. 15e), with the offset stack, the wavelet-correlation filter, and radiation-pattern correction (the steps 1 + 2 + 3), shows better SNR for the major reflectors.

Figure 16 shows a 3D volume comprised of all 13 surveys of 2D stacks in terms of a repeatability illustration. Recall that there is no time-lapse signal in this data, so we are focusing on repeatability only. Specifically, we can compare surveys across time at specific sections (the arrows in Fig. 16). Similar repeatability issues can be observed from survey S6 to survey S7 in the non-VS control section (Fig. 16a), particularly at the areas marked by the yellow arrows. This gap is caused by the 17-month acquisition break. The conventional VS (Fig. 16b) shows repeatability improvement but the gap still exists. In contrast, we observe that using our comprehensive workflow gradually mitigates these discontinuities (Fig. 16c,d). Finally, Fig. 16e eliminates this break using a series of our proposed steps (steps 1 + 2 + 3). It shows a marked improvement in repeatability at all reflectors, especially for the time slices at the target level.

As we discussed in the step 1, the offset-stacking theme is a fully data-driven and target-oriented for the reservoir target. As a result, we observe that Fig. 16c strengthens the time-lapse continuity of the reservoir target from survey S6 to survey S7. As a side effect, however, other depths were subjected to experience a loss of time-lapse repeatability, especially the discontinuities around surveys 5–6 and 11–12. There should be other data-driven reasons associated with seismic properties. Specifically, as discussed in step 1, the target response of the monitored reservoir is around 1.0 s–1.4 s with an effective bandwidth of about 20 Hz–60 Hz. The data-driven weight is therefore adjusted to capture its response pattern in the TFK domain (Eq. 9). As a result, this algorithm strengthens the continuity of those events associated with the target TFK properties, whereas other events (e.g. the high-frequency or shallow events) were subjected to repeatability loss. Currently, we are in a preliminary stage of implementation where the stacking weights are determined in a heuristic way instead of through systematic optimization. A more robust and self-adaptive weights is one of the on-going research directions.

We use a normalized root mean square (NRMS) computed over a small time window around the target zone to quantify the repeatability<sup>32</sup>. This is typically calculated between two seismic traces in a given window, divided by their average RMS, expressed as a percentage. NRMS is sensitive to the differences in seismic waveforms and extremely sensitive to the smallest of changes in the data. NRMS is computed between all surveys, which results in 78 NRMS combinations at each common-depth point (CDP). A histogram of NRMS values for each method is displayed in Fig. 17. The non-VS control section (blue line) shows a bi-modal NRMS distribution (with peaks around 20% and 65%). Not surprisingly, the larger NRMS values are for surveys split between the two survey groups S1–S6 and S7–S13. A similar distribution, though narrower and less pronounced than the non-VS, is also observed for the conventional VS (red) and the offset stack VS (green). In contrast, the wavelet VS (black), using steps 1 + 2, bridge the bi-modal gap and result in a single peak value 29%. In the end, the combined process (workflow steps 1 + 2 + 3) produces the best NRMS values centered at about 25% (light blue). These results



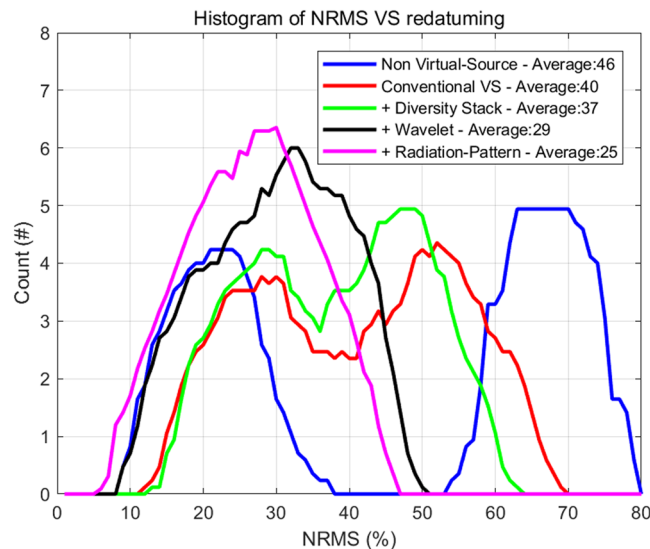
**Figure 16.** 3D volume plots of 2D CDP stacks of all 13 surveys including (a) non-virtual source (non-VS), (b) conventional VS, (c) VS using the offset stack, (d) VS using the offset stack and the wavelet-crosscorrelation filtering, and (e) VS using the offset stack, the wavelet-crosscorrelation filtering and the radiation-pattern correction. The green arrows indicate the SNR improvement on the target reservoir. Discontinuity issue between surveys S1-S6 and surveys S7-S13 marked by the yellow arrows. Panel (e) shows the best continuity along the time-lapse dimension.

support the observations made from Fig. 16 and indicate that the proposed strategy can effectively reduce overburden effects between surveys separated by a 17 month gap, and an improvement from the conventional VS.

## Discussion and Conclusion

We review three articles and summarize a comprehensive strategy to improve image quality and repeatability of virtual source redatuming in the presence of the complex near-surface. This VS method consists of a data-driven offset stack, wavelet-crosscorrelation, and radiation-pattern correction. In the first stage, the offset stack integrates wavelet transformation and cross-coherence analysis. The offset weight coefficients are directly computed from the data, based on the coherence similarity of the predicted upgoing wavefields and the original upgoing wavefields. The second step of the wavelet-crosscorrelation filtering integrates wavelet transformation, cross-correlation, non-stationary TF, and TFK filtering. This integration maintains signal coherence across frequency and wavenumber bands, and exploits scale dependency at each frequency and wavenumber to allow for better





**Figure 17.** Histogram showing NRMS computed using the four processing flows. Median values of each distribution is displayed in the legend.

noise filtering and signal separation. Effective noise suppression and high-resolution separation of nonstationary signals is achieved using TF, TFK, or TFX filtering of the wavelet-correlation coefficients. Lastly, prior to crosscorrelation, the direct arrivals are iteratively estimated using a zero-phase matched filter incorporating the 3D radiation pattern. This step modifies the radiation pattern of the virtual source, such that it closely resembles that of the ideal direct P-wave with isotropic radiation pattern. Iterative matched filtering effectively corrected the distorted radiation pattern of the direct-downgoing wavefields, and produced a significant improvement in illuminations.

Results from field data tests demonstrate that this comprehensive strategy can effectively attenuate virtual source artifacts, and produce distinct stack images without requiring a near-surface model. As time-lapse noise is mainly caused by near-surface variations over time, by reducing near-surface influences on reflection signals, the described methodology can improve time-lapse repeatability as well as imaging quality. In summary, offset stacking, crosscorrelation and denoising have been considered the three essential components of virtual source redatuming. Within this framework, we have demonstrated the value of our proposed workflows with a challenging onshore time-lapse application in a desert environment. Improvements were confirmed using the 13 time-lapse surveys that included a significant repeatability problem across a 17-month survey gap.

**Accession codes and data.** Data & Code is available at: <https://github.com/zhaoyangprof/waveletVS.git>.

Received: 22 January 2019; Accepted: 29 October 2019;

Published online: 13 November 2019

## References

- Bakulin, A., Smith, R., Jervis, M. & Burnstad, R. In *84th Annual International Meeting* 4843–4847 (SEG, Expanded Abstracts, 2014).
- Shapiro, N. M. & Campillo, M. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *31*, <https://doi.org/10.1029/2004gl019491> (2004).
- Lobkis, O. I. & Weaver, R. L. On the emergence of the Green's function in the correlations of a diffuse field. *110*, 3011–3017, <https://doi.org/10.1121/1.1417528> (2001).
- Grechka, V. & Zhao, Y. Microseismic interferometry. *The Leading Edge* **31**, 1478–1483, <https://doi.org/10.1190/tle31121478.1> (2012).
- Bakulin, A. & Calvert, R. The virtual source method: Theory and case study. *Geophysics* **71**, SI139–SI150, <https://doi.org/10.1190/1.2216190> (2006).
- Schuster, G. T. & Zhou, M. A theoretical overview of model-based and correlation-based redatuming methods. *Geophysics* **71**, SI103–SI110 (2006).
- Kelamis, P. G., Erickson, K. E., Verschuur, D. J. & Berkhout, A. Velocity-independent redatuming: A new approach to the near-surface problem in land seismic data processing. *The Leading Edge* **21**, 730–735 (2002).
- Wapenaar, K. Retrieving the Elastodynamic Green's Function of an Arbitrary Inhomogeneous Medium by Cross Correlation. *Physical Review Letters* **93**, 254301 (2004).
- van der Neut, J., Thorbecke, J., Mehta, K., Slob, E. & Wapenaar, K. Controlled-source interferometric redatuming by crosscorrelation and multidimensional deconvolution in elastic media. *Geophysics* **76**, SA63–SA76 (2011).
- Wapenaar, K., Fokkema, J. & Snieder, R. Retrieving the Green's function in an open system by cross correlation: A comparison of approaches (L). *The Journal of the Acoustical Society of America* **118**, 2783–2786, <https://doi.org/10.1121/1.2046847> (2005).
- Korneev, V. & Bakulin, A. On the fundamentals of the virtual source method. *Geophysics* **71**, A13–A17 (2006).
- Snieder, R. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Physical Review E* **69**, 046610, <https://doi.org/10.1103/PhysRevE.69.046610> (2004).
- Halliday, D. & Curtis, A. Seismic interferometry, surface waves and source distribution. *Geophysical Journal International* **175**, 1067–1087 (2008).
- Snieder, R., Wapenaar, K. & Larner, K. Spurious multiples in seismic interferometry of primaries. *Geophysics* **71**, SI111–SI124 (2006).
- Snieder, R. Extracting the Green's function of attenuating heterogeneous acoustic media from uncorrelated waves. *The Journal of the Acoustical Society of America* **121**, 2637–2643, <https://doi.org/10.1121/1.2713673> (2007).

16. Wapenaar, K. & Neut, J. V. D. A representation for Green's function retrieval by multidimensional deconvolution. *The Journal of the Acoustical Society of America* **128**, EL366–EL371, <https://doi.org/10.1121/1.3509797> (2010).
17. Mehta, K., Bakulin, A., Sheiman, J., Calvert, R. & Snieder, R. Improving the virtual source method by wavefield separation. *Geophysics* **72**, V79–V86, <https://doi.org/10.1190/1.2733020> (2007).
18. Mehta, K., Sheiman, J. L., Snieder, R. & Calvert, R. Strengthening the virtual-source method for time-lapse monitoring. *Geophysics* **73**, S73–S80, <https://doi.org/10.1190/1.2894468> (2008).
19. van der Neut, J., Alexandrov, D. & Bakulin, A. Shallow virtual source redatuming by multi-dimensional deconvolution. *Geophysical Prospecting* **64**, 4–18, <https://doi.org/10.1111/1365-2478.12258> (2016).
20. Alexandrov, D., Bakulin, A., Burnstad, R. & Kashtan, B. Improving imaging and repeatability on land using virtual source redatuming with shallow buried receivers. *Geophysics* **80**, Q15–Q26, <https://doi.org/10.1190/geo2014-0373.1> (2015).
21. Zhao, Y., Zhang, H., Liu, H., Zhang, D. & Sengupta, M. Target-oriented diversity stacking for virtual-source imaging and monitoring. *The Leading Edge* **37**, 306–314 (2018).
22. Kristeková, M., Kristek, J., Moczo, P. & Day, S. M. Misfit Criteria for Quantitative Comparison of Seismograms. *Bulletin of the Seismological Society of America* **96**, 1836–1850, <https://doi.org/10.1785/0120060012%J> Bulletin of the Seismological Society of America (2006).
23. Mallat, S. G. & Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing* **41**, 3397–3415 (1993).
24. Diallo, M. S., Kulesh, M., Holschneider, M., Scherbaum, F. & Adler, F. Characterization of polarization attributes of seismic waves using continuous wavelet transforms. *Geophysics* **71**, V67–V77, <https://doi.org/10.1190/1.2194511> (2006).
25. Labat, D. Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology* **314**, 275–288 (2005).
26. Li, H. & Nozaki, T. Application of Wavelet Cross-Correlation Analysis to a Plane Turbulent Jet. *JSME International Journal Series B* **40**, 58–66, <https://doi.org/10.1299/jsmeb.40.58> (1997).
27. Torrence, C. & Compo, G. P. A Practical Guide to Wavelet Analysis. **79**, 61–78, [https://doi.org/10.1175/1520-0477\(1998\)079<0061:Apgtwa>2.0.Co;2](https://doi.org/10.1175/1520-0477(1998)079<0061:Apgtwa>2.0.Co;2) (1998).
28. Zhang, H., Thurber, C. & Rowe, C. Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bulletin of the Seismological Society of America* **93**, 1904–1912 (2003).
29. Zhao, Y. & Li, W. Wavelet-crosscorrelation-based interferometric redatuming in 4D seismic. *Geophysics* **83**, Q37–Q47, <https://doi.org/10.1190/geo2017-0489.1> (2018).
30. Zhao, Y. & Li, W. Model-based radiation pattern correction for interferometric redatuming in 4D seismic. *Geophysics* **83**, Q25–Q35, <https://doi.org/10.1190/geo2017-0731.1> (2018).
31. Madariaga, R. In *Geophysics* 1129–1133 No title here. (Springer US, 1989).
32. Kragh, E. & Christie, P. Seismic repeatability, normalized rms, and predictability, **21**, 640–647, <https://doi.org/10.1190/1.1497316> (2002).

## Acknowledgements

We would like to thank Saudi Aramco for permission to publish this work. We specifically extend our appreciation to Cristina Young for her critical suggestion to improve the quality of this manuscript. This study is jointly supported by the National Basic Research Program of China (no. 2015CB250903), and the Science Foundation of China University of Petroleum, Beijing (no. 2462019YJRC007).

## Author contributions

Dr. Yang Zhao conceived of the presented idea, developed the theory, and performed the computations. He wrote the initial version of manuscript and verified the preliminary results. He changed employers during the time that manuscript revisions were being written. Dr. Tao Liu organized these four workflow diagrams, and delivered many insightful discussions with respect to the first-round reviewer comments. Prof. Genyang Tang provided the data platforms and computational resources for the corresponding author during his employment transition. This ensured that the revised manuscript could be resubmitted before the deadline. Dr. Houzhu Zhang assisted in preparing the updated data and codes, since the corresponding author was unable to access the resources which were associated with his previous employer. Dr. Madhumita Sengupta provided English language editing service and fixed grammar issues.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Y.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019