Characterization of drilling-related noise and curvelet-based evaluation of seismic-interferometric reflections for imaging of iron-bearing formations in Pilbara, Western Australia

Chamarczuk, M.; Malinowski, M.; Draganov, D.S.; Grant, A.; Asgharzadeh, M.; Urosevic, M.

DOI
10.1093/gji/ggab059

Publication date
2021

Document Version
Final published version

Published in
Geophysical Journal International

Geophysical Journal International

Geophys. J. Int. (2021) **226,** 377–404
Advance Access publication 2021 March 17
GJI Seismology

doi: 10.1093/gji/ggab059

Characterization of drilling-related noise and curvelet-based evaluation of seismic-interferometric reflections for imaging of iron-bearing formations in Pilbara, Western Australia

M. Chamarczuk[1], M. Malinowski[1], D. Draganov,[2] A. Grant,[3] M. Asgharzadeh[4] and M. Urosevic[4]

[1] *Institute of Geophysics PAS,* 01-452 *Warsaw, Poland. E-mail: mchamarczuk@igf.edu.pl*
[2] *Department of Geoscience and Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology,* 2600 *GA Delft, The Netherlands*
[3] *BHP Minerals, Perth, Western Australia, Australia*
[4] *Exploration Geophysics, Western Australian School of Mines, Curtin University, Perth, Australia*

Accepted 2021 February 9. Received 2021 February 3; in original form 2020 May 25

SUMMARY

Here we analyse ambient noise (AN) data generated during drilling of exploration boreholes and recorded using a dense array deployed over one of the numerous shallow iron-ore mineralization targets in the Pilbara region (Western Australia). Drilling and drilling-related operations were reoccurring in a sequence as described by the drillers' field notes, which created the rare opportunity to analyse AN data in time segments when only one type of technical process was predominantly active. Consequently, most of the recorded AN sources did not overlap in time and space. We extract the recordings in 15-min-long segments matching the time-span of single field-note entry and identify individually acting AN sources associated with specific field operations. The temporal variations of noise spectrograms and AN cross-correlations show dependency on the sequence of a few consecutive field operations and specific frequency–amplitude patterns associated with single field operations. These changes are directly reflected by the events visible in the retrieved virtual-source gathers (VSG), implying significant changes in noise temporal and spatial stationarity. Some VSGs represent the mixed contributions of surface and air waves. To remove the contributions of these arrivals to the reflection imaging, we visually inspect all data and select only field operations acting as stationary-phase sources specifically for the reflection retrieval. This was done for different receiver configurations inside PilbArray, and as a result, we obtain a collection of VSGs containing coherent body-wave reflections. Database of visually inspected VSGs is used to develop and benchmark a semi-automatic curvelet-based method for accurate parametrization of the reflection events retrieved from passive data and to compare the imaging quality of the different field operations. Common-midpoint stacks from manually and automatically selected VSGs show reflectivity consistent with the one obtained from the active-source data and related to the structure hosting shallow iron mineralization. Our results demonstrate the capacity of AN seismic interferometry to retrieve body-wave reflections and image shallow mineralization. They also provide an intermediate step toward automating the passive reflection imaging with similar data sets.

Key words: Australia; Body waves; Induced seismicity; Seismic interferometry; Seismic noise.

1 INTRODUCTION

Continuous recordings of seismic ambient noise (AN) allow to study natural and anthropogenic seismicity as well as to retrieve information about the subsurface using seismic interferometry (SI, e.g. Shapiro & Campillo 2004; Sens-Schönfelder & Wegler 2006; Nakata *et al.* 2015; Brenguier *et al.* 2020). The SI principle states that by cross-correlation (CC) of seismic recordings at two receivers, it is possible to retrieve an estimate of the impulse response between them as if one of the receivers were acting as a so-called virtual source (Schuster 2001; Wapenaar 2003; Wapenaar & Fokkema 2006). The resulting virtual-source response is influenced by the content of the AN wavefield and pre-processing applied before the CC (Curtis *et al.* 2006; Wapenaar *et al.* 2008).

Studies have shown that AN induced by local anthropogenic seismicity is suitable for subsurface imaging purposes. First successful applications utilized AN induced by trucks moving in the tunnels beneath a hill to retrieve its *P*-wave reflectivity image (Matsuoka *et al.* 2006). Further investigations of AN induced by road traffic were used to extract surface waves (Halliday *et al.* 2008; Nakata *et al.* 2011; Behm *et al.* 2014) and shear waves (Nakata *et al.* 2011). Quiros *et al.* (2016) demonstrated that AN induced by trains could be utilized to extract both body and surface waves and produce virtual shots similar to those from active seismic data. The application of railroad ANSI was further extended to crustal depths by Brenguier *et al.* (2019), who showed that passing trains generate direct *P* waves useful for monitoring of active faults. Another contribution to anthropogenic AN comes from industrial activities. Initially, Olivier *et al.* (2015) showed that underground recordings of AN induced by mine activity could be used to obtain the *S*-wave velocity structure of the mine. Further studies showed that mine activities could be utilized for imaging also with arrays at the surface (Cheraghi *et al.* 2015; Roots *et al.* 2017; Girard & Shragge 2019), primarily by retrieval of reflected *P* waves.

In contrast to the complicated AN wavefield generated by vehicle movement and mine activity, continuous recordings of borehole drilling are dominated mostly by the repetitive type of arrivals with approximately known source positions (Liu *et al.* 2016). This type of data is often referred to as active noise (e.g. Liu *et al.* 2016) and dedicated methods were developed for its processing (e.g. Seismic While Drilling; Rector & Marion 1991; Poletto & Miranda 2004). These methods usually require pilot signals to remove the drilling-induced signatures. However, even without a pilot trace, it is possible to retrieve the reflection response from drilling-induced AN using SI (Liu *et al.* 2016; Asgharzadeh *et al.* 2019).

Currently, drilling-related AN studies utilize only a small fraction of the recorded data, that is those directly associated with hammering action of a drill-bit (e.g. Miyazawa *et al.* 2008; Asgharzadeh *et al.* 2019). Incorporating contributions from AN sources acting between borehole locations could yield more isotropic illumination of the medium, extending the capacity of AN to image parts of the subsurface uncovered by drilling-induced noise. Furthermore, in conjunction with continuous recordings from an array of sensors, SI allows retrieving multiple virtual-source responses for a single receiver position. In this context, the additional advantage of utilizing AN sources not related to hammering action comes from the fact that such sequential virtual-source responses might be retrieved at receiver locations which do not fall in stationary regions (Snieder 2004) of seismicity induced by actual drilling.

The repeatability of the virtual-source response patterns would depend on the statistical repeatability of AN oscillations over specific time windows (Sanchez-Pastor *et al.* 2019). If the AN conditions vary between time windows, the resulting virtual sources are the function of both the underlying subsurface and the noise conditions. Provided the variations of AN are repeatable, it should allow to identify repeatable patterns of virtual sources and select a specific type of pattern directly related to subsurface properties. Such repetitive AN is known to be generated by technical processes related to the drilling of boreholes. The recent study performed at a Reykjanes geothermal field (Sànchez-Pastor *et al.* 2019), where multiple injection boreholes were drilled, indicated the possibility to identify short-term AN fluctuations associated with significant changes in the drilling process. However, because the multiple injection boreholes were drilled simultaneously in different locations of the study area, the associated AN variations could not be directly linked with a single process causing them. If a technical process involves periods when an only single type of field operation (e.g. drilling of an exploration borehole in a single location) predominantly contributes to seismic recordings, it would be possible to establish a causal link between specific variations of AN and individual processes.

In SI used for reflection retrieval, the virtual-source response, for which the trace from one receiver position (referred to as a master trace) is cross-correlated with the traces from all other receivers of the recording array, is referred to as virtual-shot gather (VSG, Draganov *et al.* 2009). In this configuration, SI processing aims to obtain a VSG resembling its counterpart from active-source exploration surveys, that is characterized by the similar slope of the first-arrivals and similar appearance of reflections (characterized by hyperbolic moveout in case of horizontal layers). Body-wave arrivals could be detected using well-established methods like beamforming (Johnson & Dudgeon 1993) or illumination diagnosis (Almagro-Vidal *et al.* 2014). Compared to body-wave illumination, appraisal of the reflectivity content is more challenging, because reflections might exhibit coherency only across few traces and might be characterized by a lower signal-to-noise ratio (SNR) in the VSGs as compared to the same events in the active-source data (e.g. Draganov *et al.* 2013). Determining thresholds of the first-arrivals slope, and SNR of reflection arrivals usually requires using either *a priori* knowledge about the medium (e.g. Olivier *et al.* 2015) and/or synthetic tests (Ruigrok *et al.* 2010; Minato *et al.* 2012). Having a manually selected set of VSGs, we would be able to perform reverse process and define the parameters characterizing the specific radiation pattern and presence of coherent events in the VSGs, which already satisfy the criteria for reflection imaging. Several methods exist that indirectly evaluate reflection content (e.g. see convergence measure proposed by Girard & Shragge 2019); however, to our knowledge, there are no SI array-processing methods that directly measure the kinematics and spatio-temporal location of reflection events retrieved in VSGs.

The aim of this study is twofold. First, we investigate how to effectively parametrize and characterize AN data to be able to distinguish between different anthropogenic processes. Then, we use this knowledge for generating VSGs containing coherent reflection events. Initially, we optimize the reflectivity content in the VSGs by manual selection. The visual verification allows to confirm objectively the presence of reflections in VSGs, which in turn means that the AN source, whose time segment was cross-correlated to obtain the given VSG, was located in the stationary-phase region for the retrieval of a reflection. Next, we introduce an automatic data-driven approach using the curvelet transform to maximize the reflectivity content in the VSGs. Our method selects the VSGs that have the highest probability of containing the correct reflections out of all analysed data (meaning that the targeted reflection event has the highest SNR out of all VSGs inspected for a given master-trace position).

We use visual inspection of every calculated VSG to assure that we are not mixing surface, direct and reflected waves: (i) We perform analysis of the recorded noise to see if some field operations contribute constructively to the VSGs (not specifically to surface or body waves, but coherent arrivals in general). (ii) During the visual inspection, we start to distinguish between arrivals useful for our reflection imaging study and those that are not useful. (iii) For the curvelet evaluation, we use as a benchmark data that is already visually verified as the ground truth.

Avoiding mixing of arrivals is ensured by the visual inspection (we check whether every single correlated panel contains specific arrivals). We prove this by showing multiple AN quality-control (QC) panels, that represent the majority of the recorded data. The processing done in this study is visualized for a single master-trace position, and the imaging results are the consequence of applying the same procedure to every other receiver position. In the visual inspection, we focus on finding only specific arrivals that are important for the reflection imaging in this study, that is arrivals with high apparent (body-wave) velocity and events with a hyperbolic moveout (later automatically scanned using illumination diagnosis and curvelet evaluation, respectively). Naturally, when we observed other features than those mentioned above, we classified such a panel accordingly and stored the information for future studies. However, for the final reflection imaging in this study, all panels without events specific for reflection retrieval are discarded from further analysis.

We use the PilbArray data set recorded for evaluating the possibility of imaging iron-bearing formations in the Pilbara region (Western Australia) using drill-bit-induced seismic noise (Asgharzadeh *et al.* 2019). This data set is complemented with field notes describing field operations in 15-min-long segments. This catalogue, in conjunction with continuous, dense measurements of the sequence of field operations performed in multiple locations inside the array, allows us to identify temporal changes in the AN wavefield associated with specific processes. Because the drillers field notes contain only time occurrence of the performed field operations, we cannot *a priori* select the source-receiver pairs for every performed field operation. Thus, for this study, the only reliable and trustful way to say that a given source is located in the stationary-phase region for reflection retrieval is the data-driven verification that the correlation gather contains the retrieved reflections.

The paper is structured as follows. First, we focus on temporal changes of the AN recorded by PilbArray. We analyse noise spectrograms, CC functions (CCF) and VSGs to understand the frequency-amplitude AN characteristics and changes in the dominant noise sources stationarity (both temporal and spatial) and illumination. In the next step, these characteristics are used to analyse the influence of specific field operations on AN. The second part is focused on processing and imaging using SI. We obtain VSGs using AN segments related to specific field operations and evaluate their illumination and reflection content. Subsequently, we design an automatic tool for estimating VSGs quality using the curvelet transform, which is used to assess the imaging quality of different field operations objectively. Finally, we show a comparison of imaging results using active-source data, manually selected VSGs and VSGs selected after automatic evaluation. The acronyms we use in the manuscript, other than acronyms of the drilling-related operations, are listed in Table 1.

Table 1. Definitions of acronyms used in the manuscript.

Acronym	Meaning
AN	Ambient noise
CC	Cross-correlation
CCF	Cross-correlation function
PSD	Power-spectral density
SNR	Signal-to-noise ratio
SI	Seismic interferometry
QC	Quality control
VSG	Virtual-source gather

2 DATA SET AND FIELD OPERATIONS

In June 2017, a dense (∼1000 receivers) array (PilbArray) was deployed by BHP Minerals Australia in Pilbara region of Western Australia. It recorded drilling of 19 shallow exploration boreholes and associated field operations (Asgharzadeh *et al.* 2019). The primary aim of this experiment was to establish a cost-effective methodology for imaging the iron mineralization in the Pilbara region (Fig. 1) using drill-bit-induced seismic noise.

PilbArray consisted of 945 single-component point receivers (10-Hz geophones), distributed in 14 east–west oriented receiver lines, with an alternating number of 90 and 45 receiver stations along receiver lines 1–7 and 8–14, respectively (see Fig. 1). Receiver spacing was 3 and 6 m over lines 1–7 and 8–14, respectively, with receiver line spacing of 12.5 m. The subset of receivers used for analysing of the AN characteristics and those used for the generation of VSGs is shown in Figs 2(a) and (b), respectively. In total, 61.5 hr of AN data, spread over seven days of operations, were recorded with a sampling rate of 2 ms. The drilling type was reverse circulation and the bit action was a combination of percussion and rotation. The average depth of boreholes was approximately 75 m.

Continuous measurements enabled recording AN coming from periods when the boreholes were drilled as well as all the events occurring in-between. The drillers' field notes account the activities every 15 min. The following categories were listed: Standby (ST), Pre-start (PS), Drilling (DR), Collaring (CO), Geophysical Survey (SR), Pull rods (PL), Move (MV), Safety (SF) and Maintenance (MT, see Fig. 3 for their statistics). Throughout the paper, we refer to the type of activity using its acronym. Additionally, for AN segments recorded in the period of a single field operation, we provide its ordinal number counted from the beginning of the survey. The dominant activity is drilling (51 per cent); however, another part of the data comprises recordings related to different operations (with SR, PL, CO and MV activities occurring more than ten times, see Fig. 3). The data were initially stored in 2-min-long panels which were subsequently concatenated into 15-min-long segments and time-aligned to match the duration of activities provided in the field notes.

3 IMPACT OF DRILLING-RELATED OPERATIONS ON AN CHARACTERISTICS AND VIRTUAL-SOURCE RESPONSES

In this section, we analyse the AN temporal behaviour and data characteristics, without interpreting the type of waves that are retrieved (e.g. surface or body waves). We investigate the possibility to observe the constant retrieval of coherent arrivals (whatever type they are). The essence of data comparison in this section is linking the observed changes in AN (Figs 4 and 5) to changes in field operations, and associated virtual-source gathers (Fig. 6): (i) We evaluate the data consistency and the general behaviour of CCF using the AN QC panel shown in Fig. 4. (ii) We prove the dependency of the noise spectrograms (Fig. 5) and VSGs (Fig. 6) on the changes in the field operations. (iii) We find time periods when the noise sources appear in such positions that stationary-phase requirements are satisfied (because we do not know *a priori* the location of the noise sources, and we do not track their location).

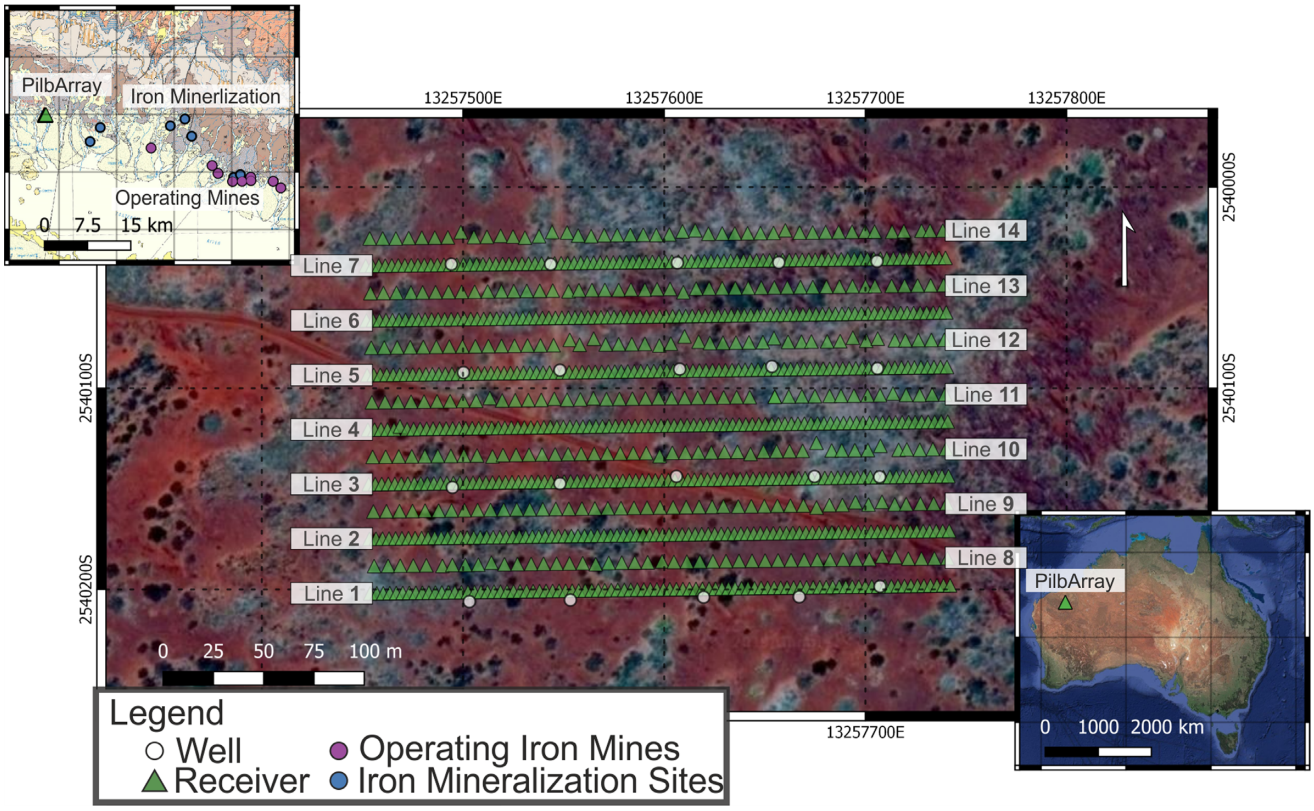

Figure 1. Location of the PilbArray. Green triangles denote 945 receiver stations used in this study.

Fig. 4 shows whether we can expect any dependency between changes in the field operations and recorded AN. The information in Fig. 4 allows observing temporal changes of (i) CCF and (ii) the frequency-amplitude character of the data (now only with the DR operation). To compute the CCFs in Fig. 4, we used receivers 3046 and 5046 (shown with blue triangles in Fig. 2a) oriented perpendicularly to the inline direction. With this spatially limited information, we can not judge whether the observed peaks are associated with surface or body waves. However, because of the presence of other peaks in CCFs, and varying PSD characteristics, the analysis in Fig. 4 highlights the potential to observe changes also associated with other field operations if: (i) more receivers would be used and (ii) preferably in the inline direction. With this in mind, Figs 5 and 6 are showing: the consistency of the data across the whole array (Fig. 5), the differences between different periods (Figs 5 and 6) and VSGs calculated between all receivers along receiver line 4 (Fig. 6). For these purposes, it is not needed to see the zoomed, detailed PSDs, but it is sufficient to comprehend the recorded data as a whole, such that the repeatability of similar patterns can be observed. Fig. 6 provides a direct depiction of the visual inspection process. We use the visual inspection shown in Fig. 6 as a tool for QC, because the visual scanning of VSGs allows (i) a quick glimpse of the whole range of retrievable events and (ii) choosing a suitable SI processing approach for the whole array.

3.1 CCF pre-processing

Before investigating the temporal evolution of CCFs and VSGs, we tested one-bit amplitude normalization and spectral whitening in conjunction with several frequency bandpass filters for the pre-processing of noise panels (see Fig. A1 for the summary of pre-processing routines comparison). To produce all CCFs and VSGs results shown in this study, we applied spectral whitening followed by a bandpass filter (10-20-90-110 Hz). This bandpass filter was adjusted to the frequency content of the PilbArray, with the lower limit of 10 Hz coinciding with the corner frequency of the geophones, while the higher limit of 110 Hz was dictated by the most energetic segments identified in the noise spectrograms shown in Figs 4 and 5b.

3.2 Sequence of field operations

Compared to the active-source studies, processing and visualization of passive data must include additional information—the temporal behaviour of data. In order to investigate whether activities listed in the field notes are identifiable in the AN data, we calculated power-spectral density (PSD, McNamara & Buland 2004) for the central crossline of PilbArray (see yellow receivers in Fig. 2a) for every 15-min-long AN segment (see bottom panels in Fig. 4). Generally, amplitude variations of the observed AN fall in the range between $\sim$10 and $\sim$110 Hz. The most significant contribution comes from drilling, which can be observed as an increase in PSD amplitudes between 15 and 40 Hz and is denoted with the violet dashed box in Fig. 4(a). In correspondence with the noise spectrograms, drilling has a clear impact also on the CCF (indicated with the black dashed box in Fig. 4a), manifesting itself as a strong peak between 0 and $\pm$0.1 s lag time throughout most of the recording time. Besides, we provide in Supporting Information Fig. S1 a plot of the RMS amplitude values. It shows data consistency over the whole recording time, as well as the temporarily increased amplitude levels consistent with the drilling itself.

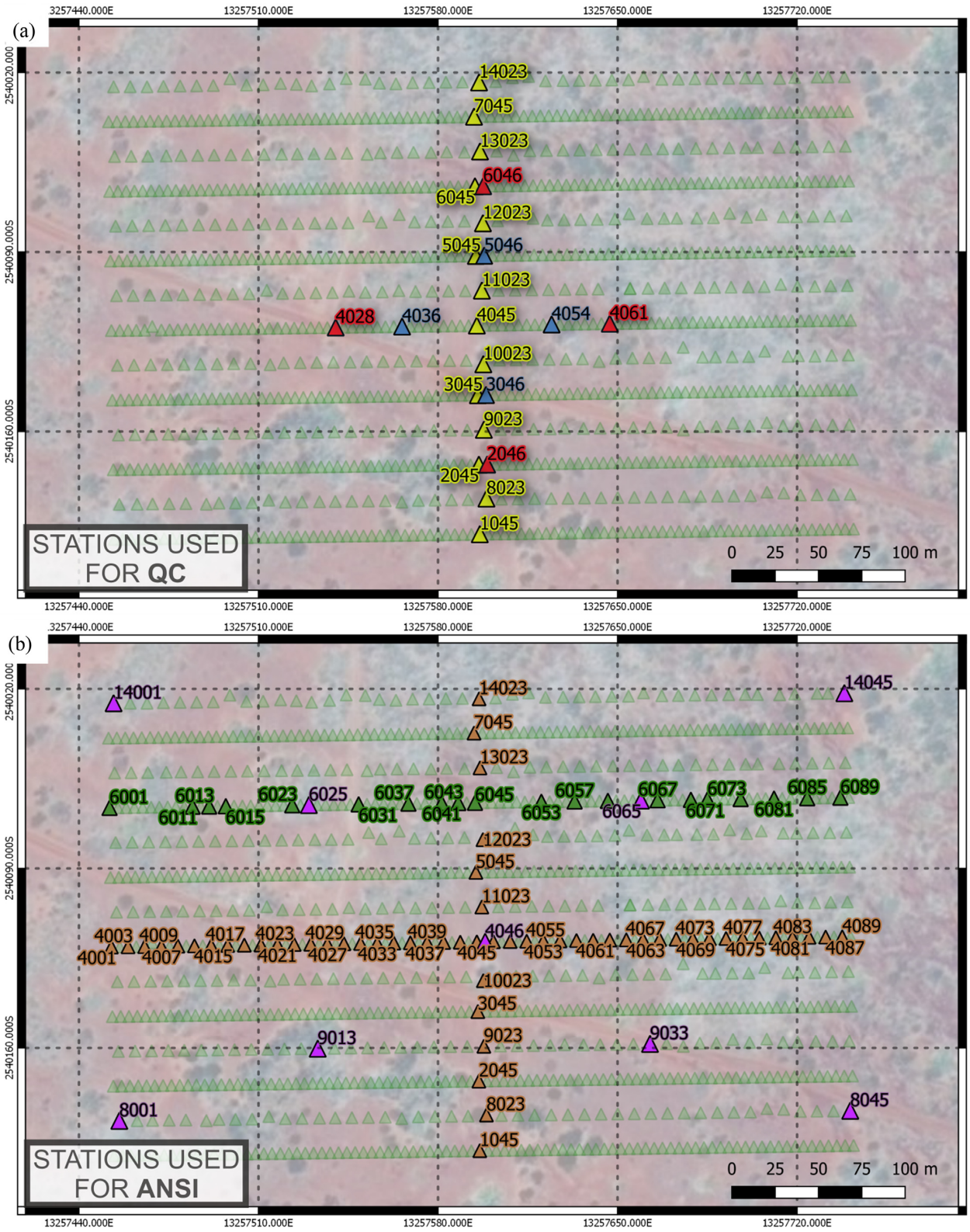

Figure 2. (a) The subset of receivers selected for analyses of ambient noise (AN) temporal variation: cross-correlation function (CCF) and power spectral density (PSD) analysis in Fig. 4 (blue and yellow triangles, respectively) and CCF pre-processing techniques comparison in Fig. A1 (red triangles). (b) The subset of receivers acting as master traces for the generation of virtual-source gathers (VSG). Receiver line 4, and receiver 4045 for which VSG temporal evolution (Fig. 6), as well as VSGs shown in Figs 7 and 8 are obtained, are denoted with horizontally aligned brown triangles. Green triangles denote the flagship receiver line 6 shown in Fig. 10 and used later on for imaging. Violet triangles denote master-trace positions used to indicate consistency between inline- and crossline-receiver direction VSGs in Fig. S5. VSG used for causal- and operation-wise stacking in Fig. S9 are denoted with brown triangles (both perpendicular lines).

(a)

FIELD OPERATION	ACRONYM	COUNT
STANDBY	ST	7
PRE-START	PS	6
DRILLING	DR	128
COLLARING	CO	19
GEOPH. SUR.	SR	32
PULL RODS	PL	25
MOVE	MV	19
SAFETY	SF	2
MAINTENANCE	MT	8
TOTAL		246

(b)

STANDBY (ST)
Client Standby for safety discussions/observations, driller standby due to operational issues before the maintenance crew arrives on-site.

PRE-START (PS)
Checking of rig before drilling starts (i.e. instrument /equipment checks, safety checks, etc.).

DRILLING (DR)
Drilling - actual drilling - i.e. rods are turning, hammer is firing etc.

COLLARING (CO)
Actual drilling the start of the hole (usually the first 3-6m) often with a larger hammer bit or tri-cone bit if the near-surface is soft.

GEOPHYSICAL SURVEY (SR)
In-rod gamma survey

PULL RODS (PL)
Removing the rods from the hole

MOVE (MV)
Rig moving from drill pad to drill pad

SAFETY (SF)
Toolbox meetings, safety observations, hazard assessment, Job safety and environmental analysis

MAINTENANCE (MT)
Maintenance of drill rig and equipment

(c)

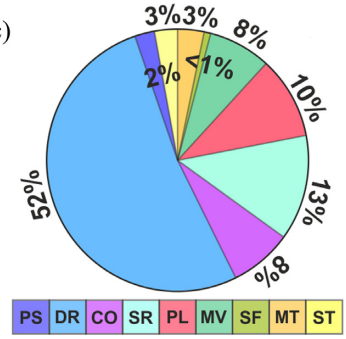

Figure 3. (a) Acronyms and the number of occurrences of each field operation recorded by the PilbArray. (b) Contribution of each type of field operation to the overall number of all recorded activities. The same colour scale is used in Fig. 9 showing field operations used to retrieve VSGs for imaging. (c) Description of drilling-related operations based on daily operation reports. Note that due to the accuracy of the daily reports provided by drillers, during 15-min-long segment, more than one activity could have occurred.

The purpose of Fig. 4 is the initial recognition of the temporal AN evolution (whether and how frequently it changes) and its eventual sensitivity to changes in the field operation as provided by field notes. Thus, at this point, we are interested in answering the question whether the field operations induce AN sources strong enough to have a visible result in a CCF, rather than whether the retrieved events are predominantly body or surface waves.

AN for days 1–3 (Fig. 4a) is dominated mostly by continuous drilling, represented by a trend of a single dominant peak on the CCF at positive time lags (note the consistent temporal trend for the 2nd day of recordings) and a horizontal band of high amplitudes on the PSD plots limited to a narrow frequency range. At the beginning of the 3rd day, the sequence of PS, PL and MV activities indicates the restart of the drilling process and relocation of the drill-rig to a new location. This sequence temporarily breaks the CCF continuity. From the beginning of the 3rd recording day, the typical sequence of DR, SR, MV, DR, CO operations indicated by a triangular shape of the activity curve (plotted as the magenta line in Fig. 4) is established. With a few exceptions (e.g. events SR101, MT104 and PS105), this triangular pattern of activities (equivalent to 75 min of AN recordings) is persistent until the end of the recordings. It is visible in the noise spectrograms for AN recorded during days 4–7 (Fig. 4b, bottom panel), where the frequency range of increased amplitude variations follows the oscillatory trend matching the repeatable changes in the drilling-related operations. Those changes are less evident in the CCFs (Fig. 4b, top panel). However, as compared to days 1–2, more events at time lags larger than those characteristic for DR activities are observed (note the peaks at ∼–0.2s lag time and the sequence of lower-amplitude peaks between 0.2 and 0.4 s lag time marked with black arrows in Fig. 4(b), usually coinciding with a sequence of MV-DR-CO and DR-CO-DR). Another contribution observed in the CCFs for the second acquisition period are events retrieved at negative time lags (see hours 7–10:30 am of the fifth and sixth day) suggesting significant spatial shifts of the dominant noise sources. Such temporal changes of CCFs indicate that selective stacking of individual correlation gathers rather than a summation of all recorded data is required in order to better retrieve reflections from our data.

Because of the limited spatial information in Fig. 4 (only single receiver pair in the crossline direction), we could not observe changes associated with field operations other than DR. For the same reason, the arrivals in Fig. 4 that are repeatedly retrieved between the two receivers can be any kind of wave, and it is not possible to interpret them at this point. Thus, to provide a more detailed analysis, in the next subsection we investigate the AN characteristic using the whole PilbArray.

3.3 Temporal evolution of full-array noise spectrograms

In order to identify the specific patterns of AN data associated with certain activities and their consistency on the scale of the whole array, we compute the PSD at every receiver position for each recorded AN segment. In Fig. 5, we show the full array PSD plots obtained for the first half of the data (the remaining part is provided in Fig. S2). High amplitudes between ∼15 and 40 Hz associated with the DR activity are evident along the complete array, and represent the most frequently occurring pattern (see the top panel in Fig. 5c). Field operations other than drilling are represented by more complicated PSD patterns (i.e. when the 15–40 Hz peak is not dominant, other patterns become more visible); the most characteristic ones are indicated in Fig. 5(d). Some of these panels exhibit patterns consistent across the whole array (for example note the horizontally and vertically distributed PSDs for panels CO18 and MV110, respectively). More spatially focused PSD patterns can be identified in panels SR101 and PS105. Exemplary noise spectrograms with

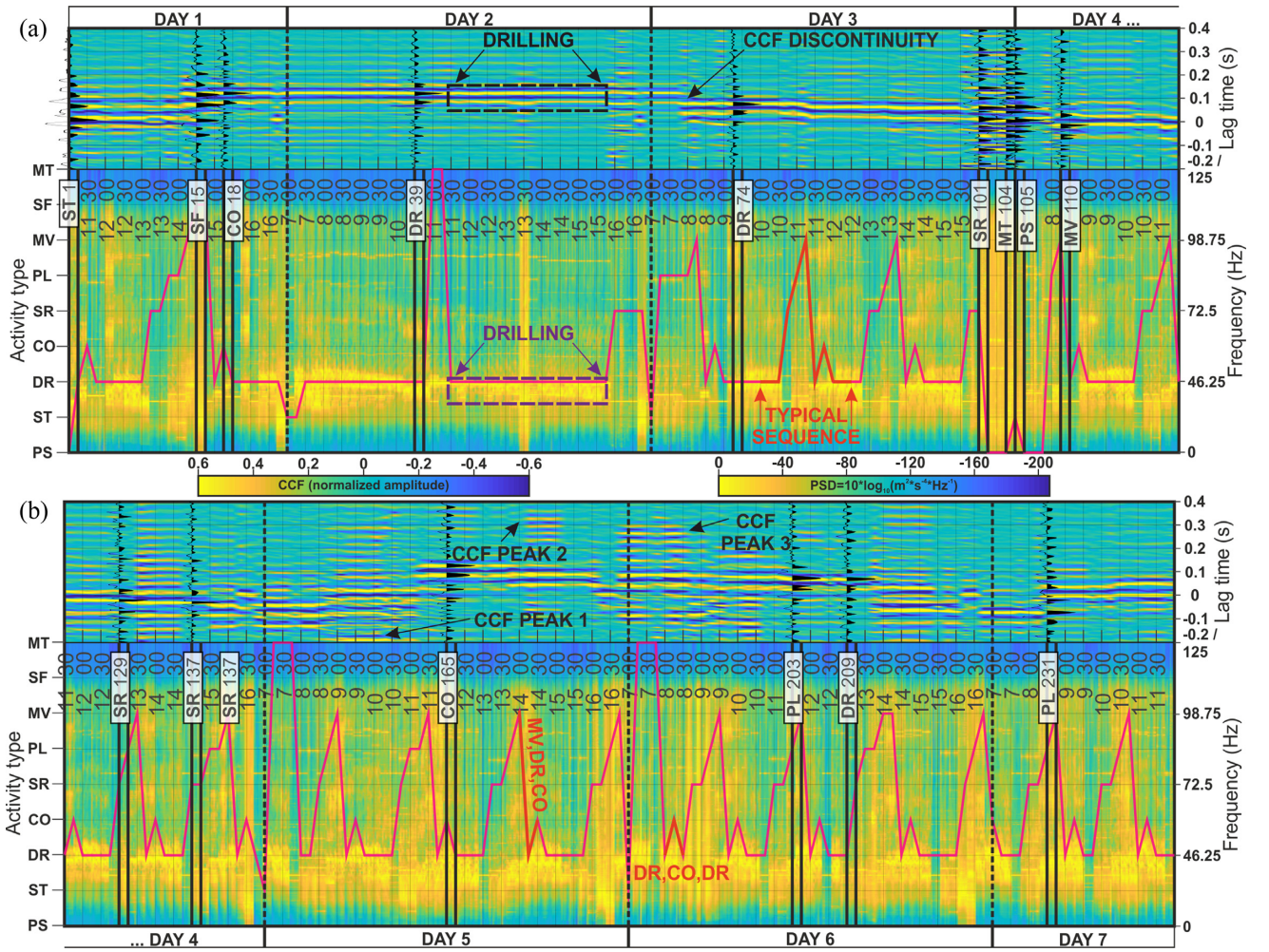

Figure 4. PSD and CCF temporal evolution computed in 15-min-long intervals and time-aligned with ongoing field operations recorded by the PilbArray. First (a) and second (b) half of the recording period. Colour plots show the comparison of the temporal changes of CCFs between receiver pair 3046 and 5046 (top panels) and noise spectrograms (bottom panels) computed for the central crossline of the PilbArray (yellow triangles in Fig. 2a). The magenta line is the activity sequence plot, indicating the type of field operation performed in each 15 min time window. It is aligned such that each column of the CCF and PSD plots can be directly associated with the ongoing operation. Representative AN segments selected for further analysis are highlighted with white labels and wiggle traces.

uniform distribution of PSD values can be observed for SF15 and MT104 segments.

The above analysis reveals a possibility to identify repeatable types of noise spectrograms, with few of them associated with a specific type of field operation. However, most of the AN segments are represented by PSD panels being a conglomerate of those repeatable patterns (note multiple AN contributions in panel DR74 or CO165 highlighted in Fig. 5d). It means that AN segments recorded during a single type of field operation (e.g. MT102, MT103 and MT104) are expected to have similar PSD patterns with likely contamination of patterns identified for other types of activities. We link these inconsistencies with the imprint of longer AN temporal trends associated with gradual changes in operation type identified in previous segments and affecting a sequence of AN segments.

Inspecting VSGs for periods of AN associated with the repeatable patterns identified in Fig. 5 may allow retrieving specific contributions typical for different drilling-related operations. Thus, before applying SI to the whole array, we focus on the inspection of VSGs obtained for representative AN segments shown in Fig. 5(d). Consequently, we use them to design a suitable processing and stacking approach. Note that we use representative AN segments for evaluating the results of SI by comparing different stacks of data. This was done using a data-driven approach, and thus at the moment, we do not interpret the events shown in Figs 4–6.

3.4 Temporal changes of virtual-source gathers

Specific sources contributing to AN wavefield can be considered as non-zero contributions to correlation gathers (e.g. Mehta *et al.* 2008). Thus, the correlation gather pattern depends also on the reference receiver. In VSGs, these contributions are represented by coherent events with different moveouts and amplitudes. Those events which pass through $t = 0$ s at the position of a master trace carry information about the illumination characteristics of sources contributing to the retrieval of the virtual-source response (Almagro-Vidal *et al.* 2014).

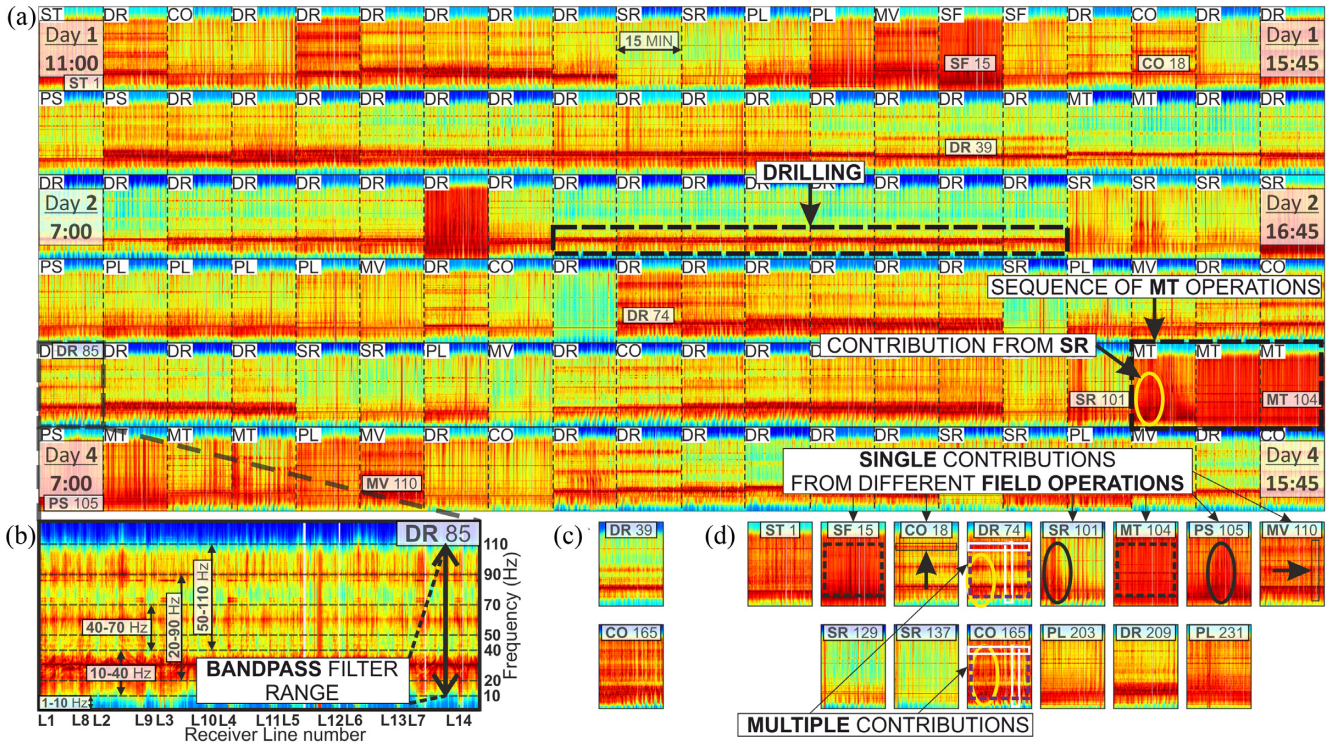

Figure 5. (a) Temporal variation of noise spectrograms during the first half of the recording time. Each panel represents PSD computed for a 15-min-long AN segment recorded by the whole PilbArray. Acronyms in white rectangles denote the type of field operation performed at the time of recording and its ordinal number counting from the start of the recording. (b) A zoomed version of a PSD plot for DR85. The dashed lines indicate the range of frequencies corresponding to CCF pre-processing results shown in Fig. A1. Panels (c) and (d) represent segments used for CCF pre-processing tests and further analysis, respectively. VSGs for segments in (d) are shown in Figs 6(d) and 7(a).

In order to associate virtual-source responses with the variations in the AN wavefield, we analyse the temporal changes of VSGs obtained using the same location of a master trace (receiver 4045) and all remaining receivers from line 4 used as input for CC (see Fig. 6 for VSGs representing the first half of the data and Fig. S3 for the remaining part). The VSGs are aligned in the same chronological order as the noise spectrograms for the full array. Cross-correlating AN data segmented in the time windows aligned with the period of the field operations reveals several repeatable patterns of VSGs. As for the PSD plots, we can highlight a few most frequently occurring patterns (see Fig. 6d with VSGs for characteristic AN segments) representing specific events retrieved in the virtual-source responses. The most notable contributions visible in the VSGs are high-amplitude air wave with apparent velocity around 340 m s^{-1} (see panels SF15 and SR129), first arrivals with apparent velocity of $\sim$790 m s^{-1} (see left-hand part of panel DR74), higher-velocity symmetric first arrivals with an apparent velocity around 1500 m s^{-1} (see arrivals highlighted with yellow lines in panel CO165), weak coherent events below the line of first arrivals (CO165), inclined linear events similar to air waves (MT104) and panels without clearly interpretable coherent events (SR137). Note, that for the moment we do not try to interpret what type of wave a retrieved event represents; however, due to the apparent velocities of the retrieved events equal to 340, $\sim$790 and 1500 m s^{-1} we tentatively interpret them as air waves, surface waves and body waves, respectively.

In Figs 4–6 we can observe many parts of data dominated by drilling activity; however, to adequately represent the difference between drilling and other field operations we highlighted only one example that is the most extended, undisturbed sequence of only this single type of operation (see the label 'Drilling' in Figs 4–6). This period represents the pattern characteristic for DR, recorded by the whole array, and also persistently observed in several successive time windows. It is an example of the spatial and temporal consistency of the recorded AN data. It allows understanding all other types of AN variations by explaining their deviations from these uniform recordings.

4 AN SEISMIC INTERFEROMETRY PROCESSING

Sources contributing to the extremum of a correlated-signal curve in a correlation gather are referred to as stationary-phase contributors (i.e. stationary sources; Snieder 2004; Snieder *et al.* 2006). SI theory states that choosing the virtual-source responses from individual source can act as a filter to determine the slowness of the virtual source (Mehta *et al.* 2008). Thus, the changes in source locations will be evidenced by changes of the arrivals retrieved in the VSGs that pass the master trace at $t = 0$ s. That would be indicative of changes in the spatial distribution of the noise sources. With a sufficiently broad spatial distribution of field operations around the PilbArray, it should be possible to capture the stationary-phase source zones of different reference receiver stations. For the reflection retrieval investigated in this study, finding the source in the stationary-phase zone and VSGs with reflection content is the same task.

4.1 AN segments from single field operations

The temporal evolution of a given VSG is inherited from the illumination characteristics of the AN sources prevailing in different

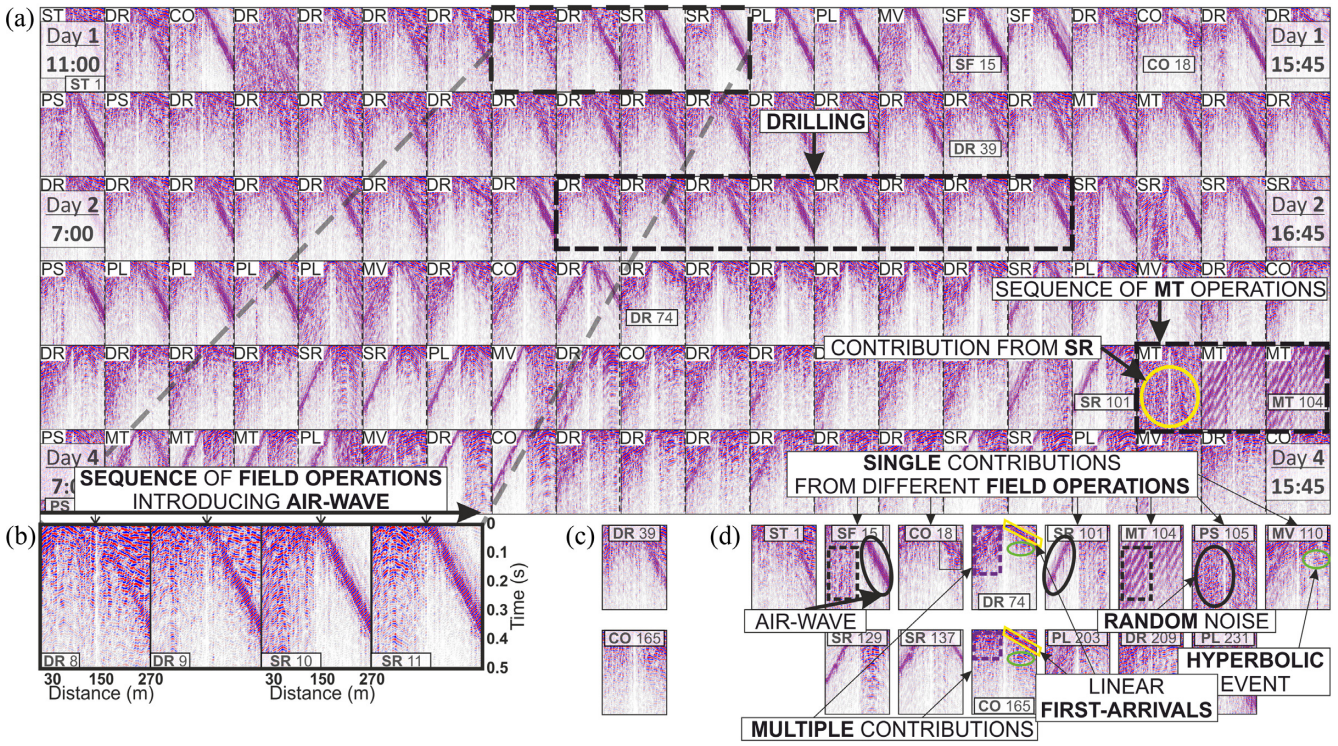

Figure 6. (a) Temporal variation of virtual-source responses during the first half of the recording aligned in the same way as the PSD plots in Fig. 5. Each panel represents a VSG obtained from a 15-min-long AN segment by cross-correlating the trace recorded at receiver 4045 with the traces from all receivers from line 4 (see brown triangles in Fig. 2b). CC of each segment was preceded by spectral whitening and bandpass filtering (10–20–90–110 Hz). Vertical and horizontal axes represent time samples and distance, respectively. (b) A zoomed version of four VSGs exhibiting gradual introduction of an air-wave contribution (see Fig. 13a). Panels (c) and (d) represent the VSGs obtained from the AN segments shown in Figs 5(c) and (d), respectively.

time segments and changing with the progress of the drilling-related operations.

Consequently, VSGs obtained from CC of separate AN segments would exhibit different retrieved contributions (note, e.g. VSGs for SF15 and SR129 in Fig. 7a). In Fig. 7(a) we show only acausal parts, because in this case, they illustrate more clearly differences between VSGs from specific field operations, while still yielding the kinematically correct arrivals. In Fig. 7(a), the most frequently retrieved contributions are surface-wave arrivals with velocities between 450 and $\sim$1038 m s^{-1} in CO18, DR74, PS105, possibly MV110 (partly hindered by artefacts), PL203 and DR209; and the air wave in ST1, SF15, SR101, MT104 (with many artefacts and multiples), MV110, SR129 and PL203. White arrows in Fig. 7(a) denote all types of non-stationary arrivals, which cannot be interpreted as neither air, surface, nor body waves. The non-stationary arrivals are evidenced by events in VGSs that have their apex outside the master trace location (indicated with a grey star in Fig. 7).

In Fig. 7(b), we show results representing the most fundamental stacking ideas: (i) using all panels, (ii) using several selected panels, (iii) using a single panel and (iv) using a combination of causal/acausal parts. To illustrate the selective stacking more clearly, we use the same VSGs as shown in Figs 6(d) and 7(a). To be more specific, in Fig. 7(b) we show from left- to right-hand panels: (i) result of stacking the same panels which are shown in Fig. 7(a) (acausal parts of VSG from 14 field operations); (ii) causal part of VSGs from all recorded panels; (iii) acausal part from all recorded panels; (iv) a combination of spatially limited acausal segments (with respect to the master trace) of VSGs from 2 field operations and (v) similar approach but extended to take into account the causal part. The panel in Fig. 7(b) with red and violet lines (stacked acausal part using VSGs from all recorded panels) denotes the mixed contributions resulting from stacking of all data. The events marked by the red lines in Fig. 7(b) have apparent velocities between $\sim$590 and 1038 m s^{-1} and thus appear to belong to higher-frequency dispersive surface waves. In the same panel, we see the lower-frequency surface wave (with velocities between 450 and 1038 m s^{-1}). The event denoted with the violet line in the same panel, below the red lines, is the air wave, which is clearly distinguished from the surface waves due to its lower apparent velocity. Representative panels shown in Fig. 7(a) highlight that the most frequently retrieved events in the Roy Hill area are the air waves (see all events denoted with violet lines) and dispersive surface waves (see all events marked with red lines). Consequently, when a VSG is obtained by stacking all data, these two arrivals are the dominant contributions (see red and violet line in Fig. 7b). The broad range of velocities of the dispersive surface waves (>440 and <1040 m s^{-1}) in the stacked VSGs (Fig. 7b) is explained by the fact that stacking of individual VSGs, which contain surface waves with different apparent velocities (see velocity labels in Fig. 7a), yields a conglomerate of all arrivals retrieved in VSGs. To avoid mixing of the different seismic phases in the VSGs, as well as to retrieve reflection events, we need to perform the selective stacking, which allows selecting only the parts of recorded data when neither surface waves nor air waves were dominant.

The waveforms associated with AN segments used to obtain the VSGs in Fig. 7(a), and exhibiting varying moveouts and hyperbola apexes, are shown in Fig. S4. At this point, it is becoming

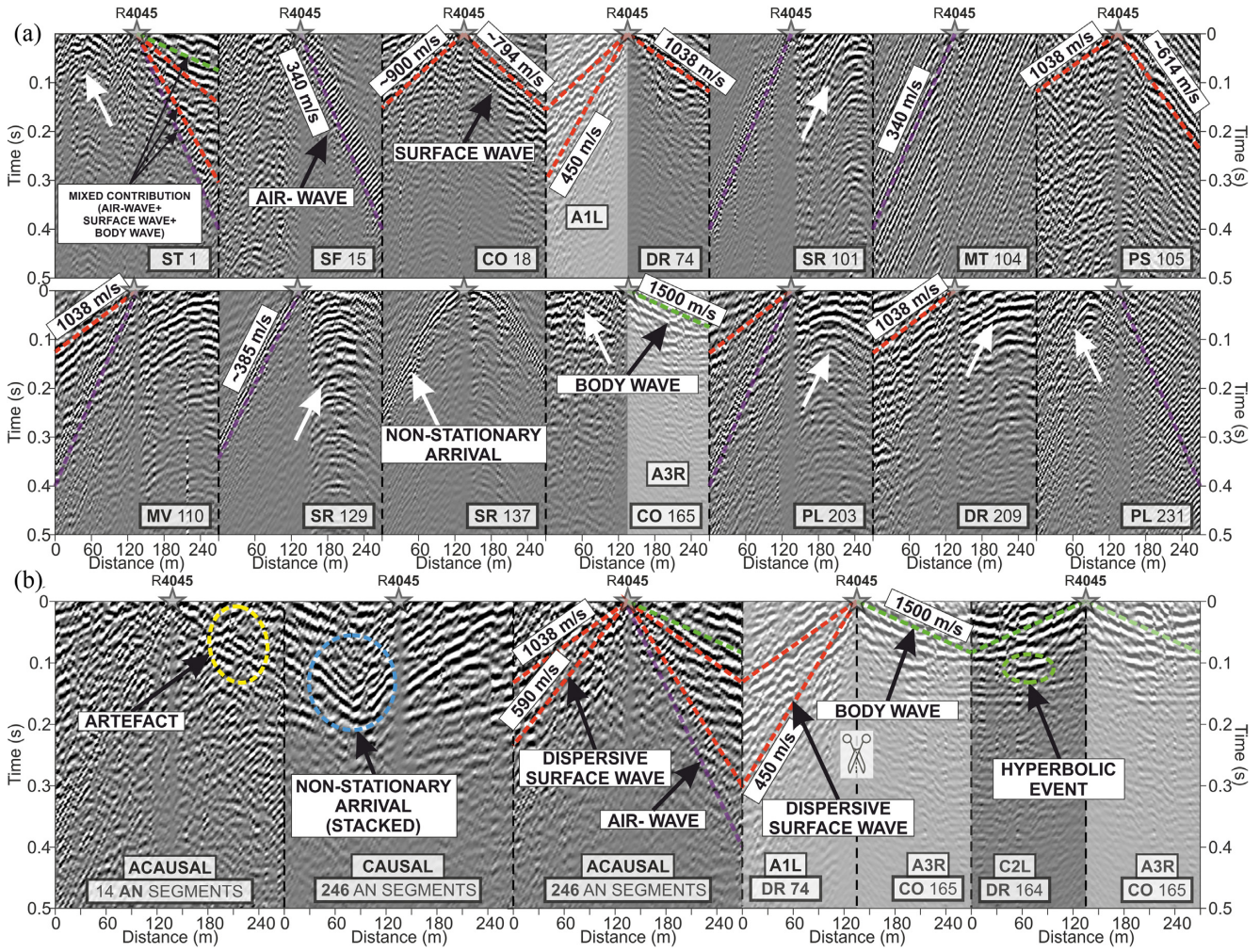

Figure 7. VSGs obtained by (a) cross-correlating AN segments recorded during the time span of a single field operation and (b) various stacking approaches. (b) From left- to right-hand panels: VSG representing stack from all the 14 panels shown in (a), using only the acausal parts; as in the left-most panel, but using only the causal part; as in the left-most panel but using all 246 segments; and two VSGs obtained by combining records from two different field operations. Grey shaded areas indicate parts of VSGs used for selective stacking, the scissors symbol denotes separation of VSG into two parts. Labels A1L and A3R denote time-reversed acausal traces to the left and to the right of the master trace, respectively. Numbers 1–3 denote DR74, DR 164 and CO165 panels, respectively, from which the parts of the VSGs in the shaded areas were extracted.

clear that some of the activities could act as stationary-phase sources for an appropriately chosen set of receivers (Mehta *et al.* 2008).

Restricting our analysis to single receiver lines requires either selecting a single field operation that produces AN sources located in the assumed stationary-phase area (Roux & Kuperman 2004), or simultaneous analysis of the positive and negative (causal and acausal) parts of the VSGs retrieved from different field operations (e.g. Draganov *et al.* 2013). We refer to these two approaches as operation-wise and causal-wise stacking, respectively. A virtual-source response with a complete reflection event in both the casual and acausal part of the CCF can be obtained when the CC results are summed over uniformly distributed sources along a surrounding surface (Wapenaar & Fokkema 2006). In reality, summing over a limited source aperture could be sufficient to retrieve a specific event but would also result in retrieval of artefacts (see the yellow ellipse in the VSGs obtained from stacking 14 activities in Fig. 7b). Summation over all available AN segments results in suppression of artefacts, but surface wave contributions dominate the retrieved VSGs (see the red ellipse in Fig. 7b), hindering the identification of reflection arrivals. The observation that the VSGs obtained from summing all available data (Fig. 7b) do not exhibit identifiable coherent reflection arrivals is another justification for applying the selective stacking in this study (besides the temporal changes of the CCFs discussed in Section 3.2).

The selective stacking, based on the selection of individual AN sources related to assumed stationary-phase regions, allows retrieving events with kinematics characteristic of reflected body waves but with distorted amplitudes (see the green ellipse in Fig. 7b). In such context, cross-correlating AN segment from a single field operation has a two-fold implication on the retrieval of VSGs: (i) illumination due to the activity performed in a specific area is one-sided (Wapenaar 2006) and provides an estimate of time-asymmetric virtual-source response (either only causal or acausal part) and (ii) a single field operation could supply stationary-phase sources for a limited number of receiver-pair configurations.

Because of the above reasons, we temporarily neglect the sparsity of the spatial sampling in the crossline direction (only 14 receivers with 12.5 m interval are available) and analyse AN segments of two field operations recorded by one receiver line in the inline and crossline directions passing through receiver 4045. This approach can provide additional information about the 3-D nature of the recorded AN wavefield (Almagro-Vidal *et al.* 2014; Panea *et al.* 2014). For instance, the events retrieved from the AN segment CO165 (see the shaded area on the VSG from CO165 in Fig. 7a) and its preceding segment DR164 are related in the noise panels to unaliased noise events in the limited range of wavenumbers in the frequency–wavenumber (f–k) domain (Figs 8a and b). Retrieving the VSGs for receiver position 4045, located at the intersection of these perpendicular lines (brown receivers in Fig. 2b), by using these unaliased noise events as input to CC results in similar moveouts and the presence of coherent events below the line of first arrivals (see the black arrows in Fig. 8c).

The purpose of Fig. 8 is to show that the waveforms associated with field operations, identified thanks to the field notes, are spatially well-sampled in the crossline and inline directions, and exhibit high-frequency content indicating their potential for the successful retrieval of reflections in the Roy hill area.

VSGs further sustain the consistency along the inline and crossline directions for master traces located at the intersection of perpendicular receiver lines (violet receivers in Fig. 2b) covering most of the PilbArray. These results are shown in Fig. S5.

4.2 Visual inspection of the virtual-source gathers

In the previous section, we demonstrated that changes in drilling-related operations affect both the noise spectrograms and the retrieved events in the VSGs. The associated temporal changes of the retrieved virtual-source responses require careful selection of AN segments concerning the selected virtual-source/receiver configuration. According to the theory, AN sources at the surface will mainly contribute to retrieval of surface waves, while AN sources relatively deeper in the subsurface will mainly contribute to retrieval of body waves (Wapenaar & Fokkema 2006). Drilling operations, as opposed to conventional industrial AN sources, are mostly occurring in the subsurface and thus can be used advantageously for retrieval of body waves with SI, as recordings from the drilling operations will not be dominated by surface waves (e.g. Liu *et al.* 2016). For a fixed AN source position in the subsurface, the slope of a retrieved reflection event depends on the propagation velocity, depth of the receivers, depth of the reflector and the angles of incidence and reflection (Snieder 2004). Therefore, calculation of the region of the stationary-phase sources for retrieval of a specific reflected wave between a virtual source and a receiver would require sufficiently accurate knowledge of the subsurface velocity model (e.g. Minato *et al.* 2012). On the other hand, assuming a more or less layered subsurface, one would expect the projection at the surface of the subsurface region of the stationary-phase sources to be along the surface line (in a wavelength sense) connecting a specific pair of a virtual source and a receiver. We use precisely this approach in our study and look for AN sources, whose projections at the surface would lie along such lines. We refer to such AN sources as falling in the stationary-phase region. Consequently, for the reflection retrieval in this study, the most objective way to say that a given source is located in the stationary-phase region for reflection retrieval is the visual verification because it allows confirming the presence of reflections in VSGs objectively.

After the analysis shown in Fig. 6 for a single reference position, it becomes clear that performing similar analysis for any other receiver position should bring at least one time period during which a stationary-phase source was present. Furthermore, AN source from a drilling operation that is deemed stationary will contribute to retrieval of reflection events not only for one pair of virtual-source/receiver but for multiple receivers for the same virtual source due to multiple reflections in the subsurface (Mehta *et al.* 2008), that is it will contribute to retrieval of reflection events from one virtual source at multiple offsets. In this subsection, we aim to select the period of a field operation acting as a stationary-phase source for a specific receiver position of the PilbArray.

The drillers field notes do not contain the information on coordinates of the performed field operations. For this reason, we decided that the most reliable approach would be to visually determine the noise source, which is the closest approximation of the stationary-phase source, for each reference-receiver station separately, and specifically for the reflection retrieval. In order to investigate the potential of the recorded drilling operations for retrieval of reflections, we first apply SI processing to every recorded activity at each receiver belonging to the same receiver line. For a specific receiver position acting as a master trace, we compute a collection of VSGs using all available field operations (i.e., 246 15-min-long AN segments). This procedure is repeated for every receiver position along the line. As a result, we obtain 246 VSGs for each of the 630 analysed receiver positions (see the map of field operations acting as stationary-phase sources in Fig. 9a). From the PSD analysis and VSG temporal changes, we expect to find at least one field operation acting as a stationary-phase source for retrieval of reflections for each master-trace location. The limited total number of noise segments allow us to analyse the virtual-source response for each consecutive noise panel. Consequently, we perform a visual inspection of the VSGs and for every analysed receiver position we choose a single VSG satisfying the following criteria: (i) radiation pattern typical for body waves and emerging from the master-trace location and (ii) identifiable coherent events with approximately hyperbolic moveouts.

4.3 Field operations acting as stationary-phase sources

As explained above, before seeing the data, the stationarity can be only assumed, and the verification comes with the visual inspection. In this study, the visual inspection relied on the inspection of the correlation gathers, from each time segment, and for every reference receiver separately (after inspecting all of the recorded segments, we move the master trace used for CC along the line one step at a time). The map in Fig. 9(a) shows which field operation was selected during the visual inspection as the operation providing the stationary-phase source for reflection retrieval at a given master-trace position. From the reason explained in Section 4.2, the stationary-phase requirement for reflection retrieval at a given reference receiver might be satisfied by sources at several different locations. Hence, while we conclude about the stationarity of a given field operation, its real location is outside the scope of the analysis in Fig. 9.

AN segments used to obtain the manually selected VSGs were recorded mostly during the drilling itself (see the blue rectangles in Fig. 9a, and representing 453 receiver positions at which DR was used), which comprise 72 per cent of all processed master-trace positions. Other field operations providing a significant number of stationary-phase sources are CO and MT with 93 and 38 covered

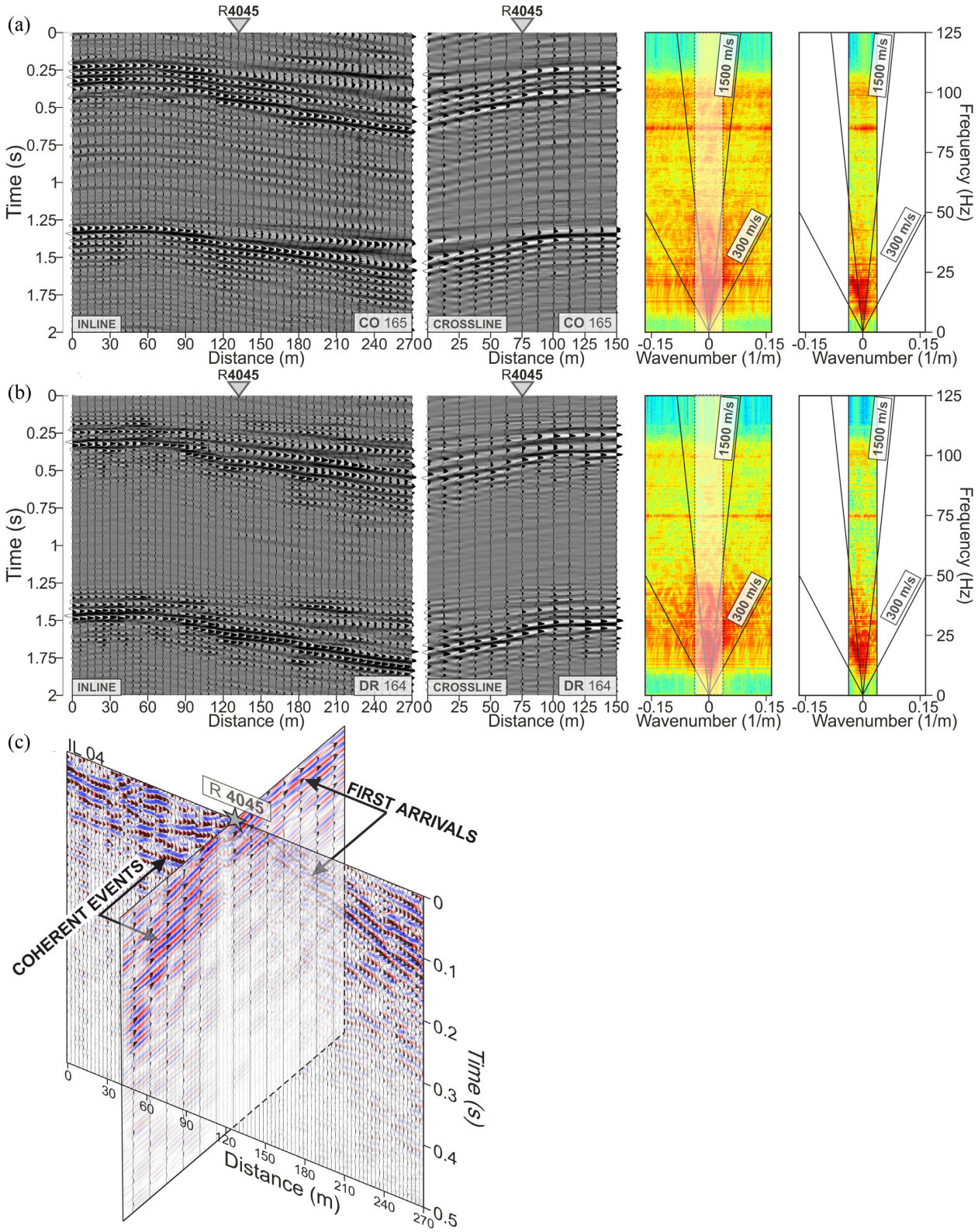

Figure 8. Comparison of (a, b) noise panels and (c) VSGs extracted along receiver line 4 and in the crossline direction using records of two consecutive field operations: (a) CO165, (b) DR164. Waveforms and their *f–k* representations are shown. (c) VSGs obtained by combining the two events. Note consistent moveouts of the first arrivals and the presence of coherent events in both inline and crossline directions.

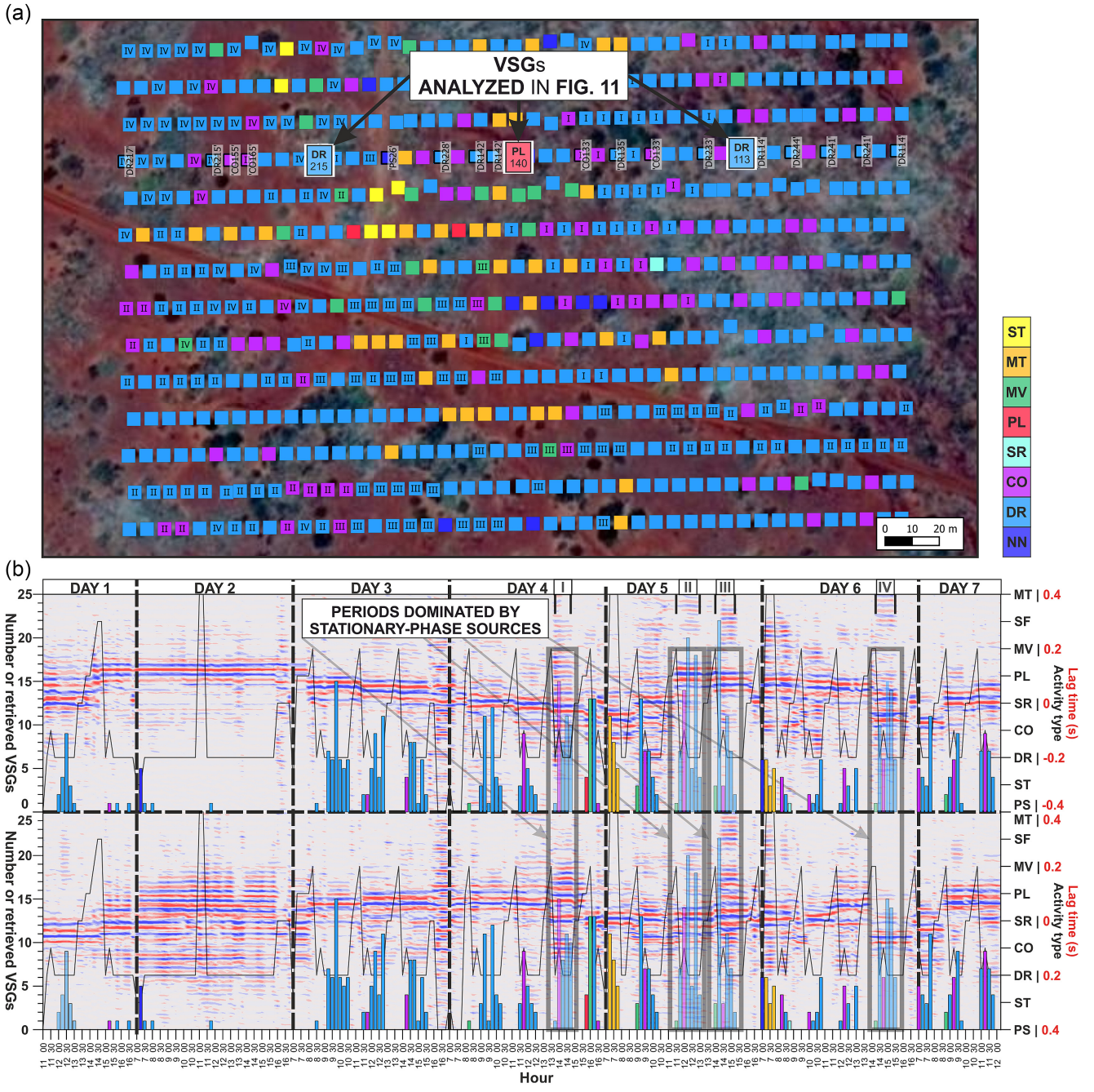

Figure 9. Analysis of field operations identified during visual inspection as the contributing stationary-phase sources with respect to the (a) locations of master-trace receivers, and (b) progress of drilling survey. To ensure consistency between the number of VSGs between alternating receiver lines for each dense receiver line (lines 1–7) every second receiver was used as a master trace. (a) Resulting 630 receiver positions highlighted in colour, where the block colour corresponds to the type of field operation used to obtain VSGs at that position. (b) The number of VSGs retrieved for each consecutive time window, where the bar colour corresponds to the type of field operation. The background colour plot shows CCF along inline (stations 4036 and 4045, top panel) and crossline (stations 3046 and 5046, bottom panel) receiver pairs. Note the sequence of activities denoted with Roman numbers, showing four sequences of field operations, contributing to ∼34 per cent of all VSG positions and coinciding with a consistent pattern of arrivals in the crossline and inline direction.

receiver positions, respectively. In general, field operations acting as stationary-phase sources usually appear in a repeatable sequence of several consecutive AN segments (note the transparent bars in Fig. 9b). It suggests, similarly to the noise spectrograms, that the stationarity of VSGs is affected by short- and long-term changes in the field-operations sequence.

We use VSGs from receiver line 6 (receivers marked in green in Fig. 2b) to showcase the performance of SI applied to AN segments from separate field operations. The retrieved VSGs are characterized by symmetrical radiation pattern and coherent events (see the first arrival emerging from the master-trace locations and parts of the VSGs falling inside the windows plotted in Fig. 10). In the following

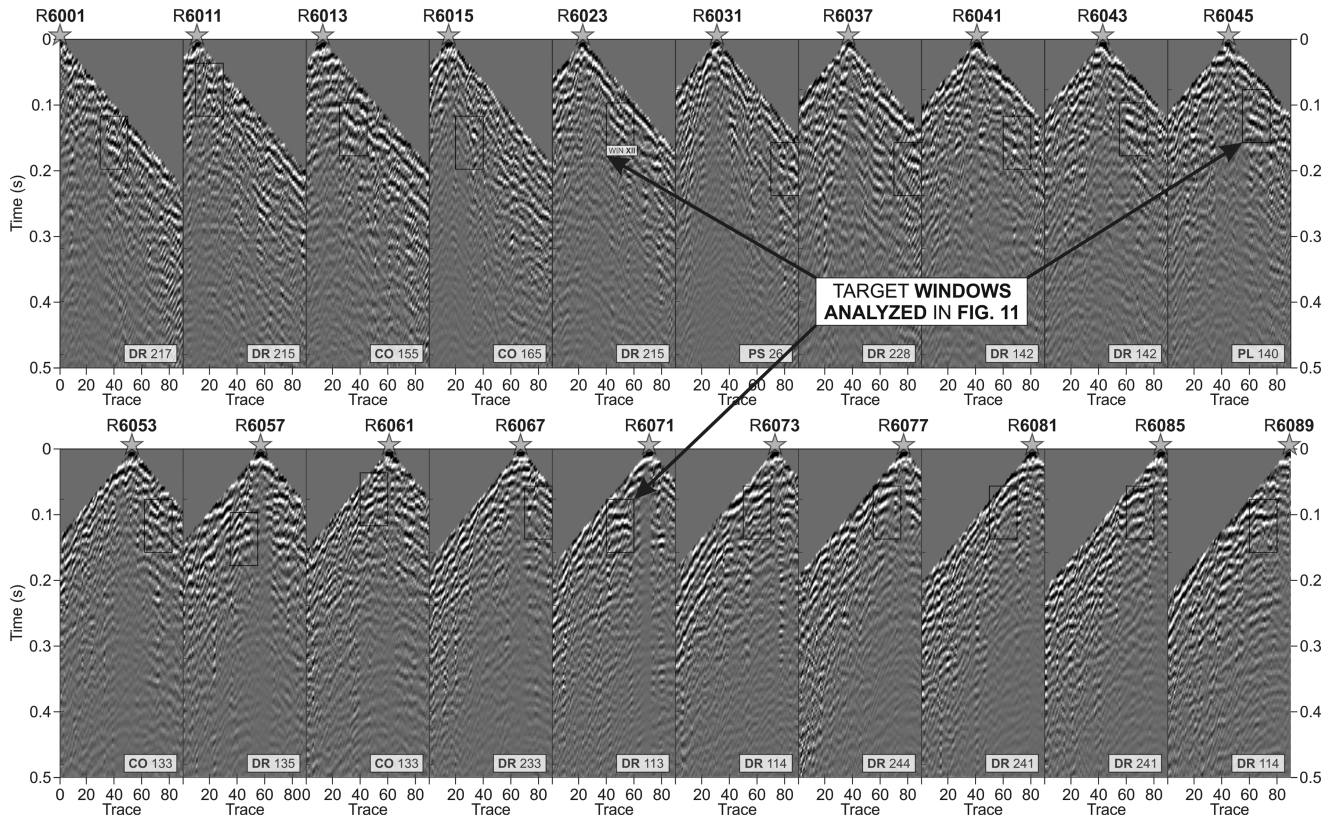

Figure 10. VSGs for 20 master-trace positions along receiver line 6. For each VSG, we provide field operation used for the CC. Black rectangles indicate windows used in the curvelet analysis. For the subset of the VSGs shown in this figure, the following field operations contributed as stationary-phase sources: 14 DR, 4 CO, 1 PL and 1 PS.

section, we parametrize such features to assess the possibility of quantifying the VSGs reflection content automatically.

5 AUTOMATIC EVALUATION OF BODY-WAVE REFLECTION CONTENT IN VIRTUAL-SOURCE GATHERS

We define VSGs useful for reflection imaging as follows: the slope of the retrieved apparent direct waves falls in the range of *P*-wave velocities expected for the study area (Almagro-Vidal *et al.* 2014; Panea *et al.* 2014) and arrivals with hyperbolic moveouts and amplitudes slightly different than those in active seismic data (see, e.g. Draganov *et al.* 2013 for comparison between active- and virtual-source reflectivity). Thus, assessing the reflection content requires a method accounting for the events' location, amplitude and angle.

5.1 Illumination-diagnosis and curvelet transform

The radiation pattern of a retrieved virtual source is determined by the slope of the dominant events emerging from the virtual-source position (Van der Neut 2013). Following Almagro-Vidal *et al.* (2014) and Chamarczuk *et al.* (2019), we use a linear slant-stack transform (Chapman 1981) to characterize the illumination characteristic of VSGs. Contrary to energetic first arrivals recorded on all traces, the reflection events are visible only on a few adjacent traces. One of the methods commonly used in seismic signal denoising, allowing to distinguish coherent events from incoherent noise, is the curvelet transform (Candés & Donoho 1999; Gòrszczyk *et al.* 2014). Using curvelet basis functions, dipping events in the time-offset domain map to the coefficients of specific angular wedges whereas noise maps over all angles (compare the top and bottom row in Fig. 11, and see the example of the discrete curvelet partitioning of the 2-D Fourier plane into scale wedges denoted with grey mesh in Fig. 12c). The choice of using the curvelet transform was motivated by the limited capacity of the *f–k* transform to distinguish between field operations inducing waveforms with similar moveouts (see Fig. 8). Here, we use the 2-D discrete curvelet transform [see Candés & Donoho (1999), for details about theory and implementation], which enables to accurately describe localized curved and complicated events (Kumar *et al.* 2011). Curvelet-based denoising can also enhance the prominence of reflections in passive data (compare the top and middle rows in Fig. 11); however, in this case, we are only interested in using the curvelet domain to characterize the reflectivity present in the VSGs.

Narrowing the evaluation of VSGs to parts of the data falling only inside rectangular windows and muting everything outside these windows allows associating a given reflection event with a single angular wedge (Fig. 11, bottom row). For the VSGs obtained at every receiver position along receiver line 6, we place a rectangular window at the most visually prominent coherent events (see black rectangles in Fig. 10 and window distribution in Fig. 12a) and evaluate the curvelet parameters. The curvelet analysis reveals that all coherent events visible in those VSGs (Fig. 12b) can be associated with *f–k* maxima within a narrow range of angular wedges (Fig. 12c).

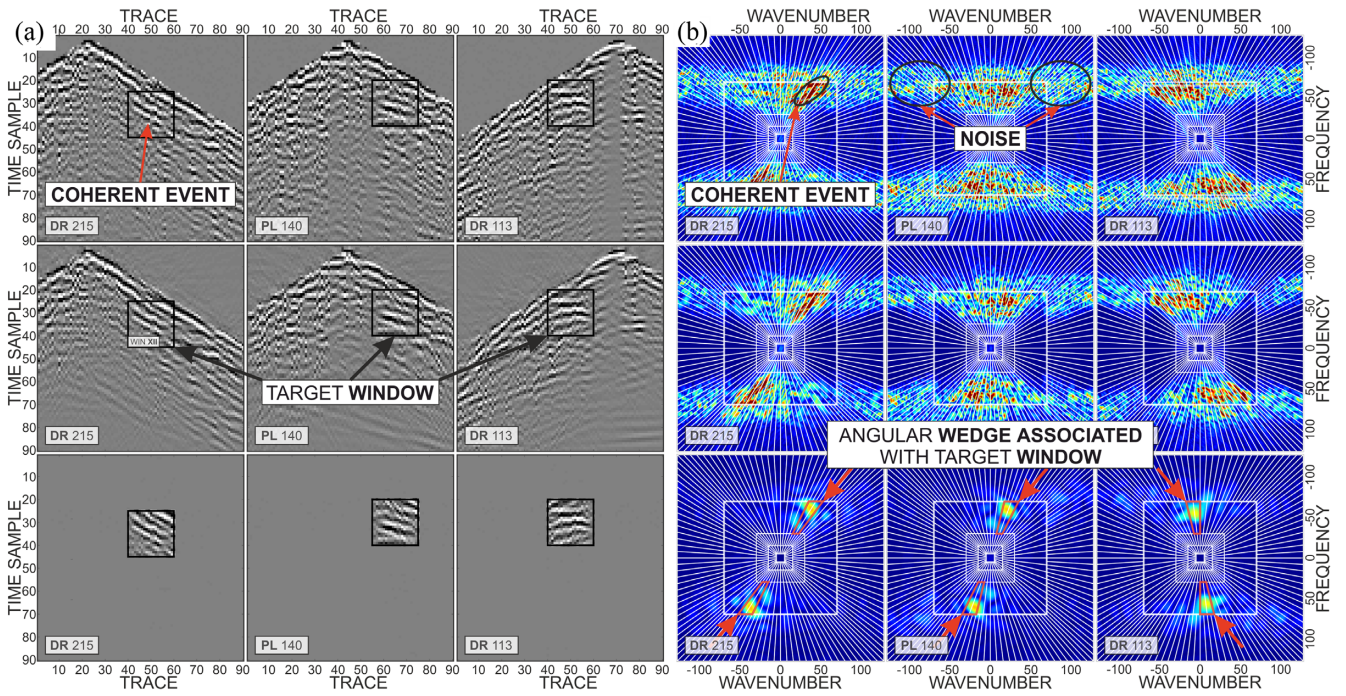

Figure 11. (a) VSGs and (b) their *f–k* representation. (a) Trace and time sampling of VSGs is 3 m and 2 ms, respectively. (b) White lines indicate the discrete curvelet partitioning of the 2-D Fourier plane into second dyadic coronae: division into scales (denoted with white boxes) and subpartitioning of the coronae into angular wedges (white radiating lines). We use 5 scales with 32 angles in the 2nd scale and 128 in the 5th scale. Curvelet analysis for the raw VSGs (top row), VSGs after filtering using global thresholding (all scales) by rejecting 1 per cent of wedges with the smallest amplitudes (middle row), and a part of VSGs extracted from the windows denoted with black rectangles (bottom row). Note that most of a VSG's energy in the *f–k* domain is related to the 4th scale (indicated with bold white line). Rejection of wedges improves the coherence of the dipping linear events, as well as those with the hyperbolic moveout. Narrowing the analysis to a part of VSGs falling in the selected windows allows to indicate the narrow range of angular wedges: see the red polygons in the bottom panel in (b), associated with the maximum values in the *f–k* domain.

5.2 Transition virtual-source gathers

The recorded data covered sequences of different field operations, inducing the change of AN stationary-phase source positions. For instance, a sequence of field operations involving equipment transportation or vehicle movements (MV) may cause spatial shifts in the illumination direction of the dominant noise sources and the introduction of new noise sources. CC of AN segments recorded during such consecutive activities results in VSGs in which the retrieved contributions and their SNR are informative about the changes in the field operations. So far, we demonstrated that the coherent events visible in VSGs could be described by several angular wedges in the curvelet domain. This observation allows quantifying the reflection content of a single VSG with otherwise unclear events. In this subsection, we test whether the curvelet parametrization enables detecting the specific contributions in VSGs and distinguishing between VSGs containing coherent events with similar kinematics but of variable SNR.

To address the detection of specific contributions, we use examples of VSGs retrieved for a single master-trace position (receiver 6023), where gradual retrieval of air waves and coherent events is observed within the period of four consecutive 15-min-long panels (we refer to these as transition panels). The retrieval of the air wave with gradually increasing SNR contaminates the VSGs and eventually masks all other arrivals (see Fig. 13a). A similar process, but for coherent events, is shown in Fig. 13(b), where transition panels starting from a VSG dominated by random noise (leftmost panel in Fig. 13b), exhibit coherent events with progressively increasing SNR. Note that the rightmost panel in Fig. 13(b) represents the VSG manually selected for receiver position 6023. To distinguish between VSGs visually similar to the one manually selected, we use the four panels shown in Fig. 13(c).

We use the difference between parametrization of wedges associated with a target window containing coherent events and wedges from remaining windows to design wedge parameters for automatic evaluation of VSGs reflectivity. For each window, we assess the reflectivity with two wedge parameters derived from the maximum amplitude of the inverse curvelet transform and the maximum amplitude of the forward curvelet transform (see Appendix B for the details of window selection, and definition of the wedge parameters W_{P1} and W_{P2} used in the following). The curvelet parametrization enables to clearly distinguish three groups of VSGs: (i) Those dominated by an air wave (DR8, DR9, SR10 and SR11), (ii) Dominated by events with higher velocities but low reflectivity (MV213, MV214, DR5 and DR218), (iii) Containing identifiable reflection content (DR215 and CO216). This observation is valid for all VSGs highlighted for line 6. A concise graphical display of these results is provided in Fig. S6.

5.3 Automatic evaluation of imaging potential of different field operations

As shown in the previous sections, the majority of the manually selected VSGs were retrieved during DR operations. However, the

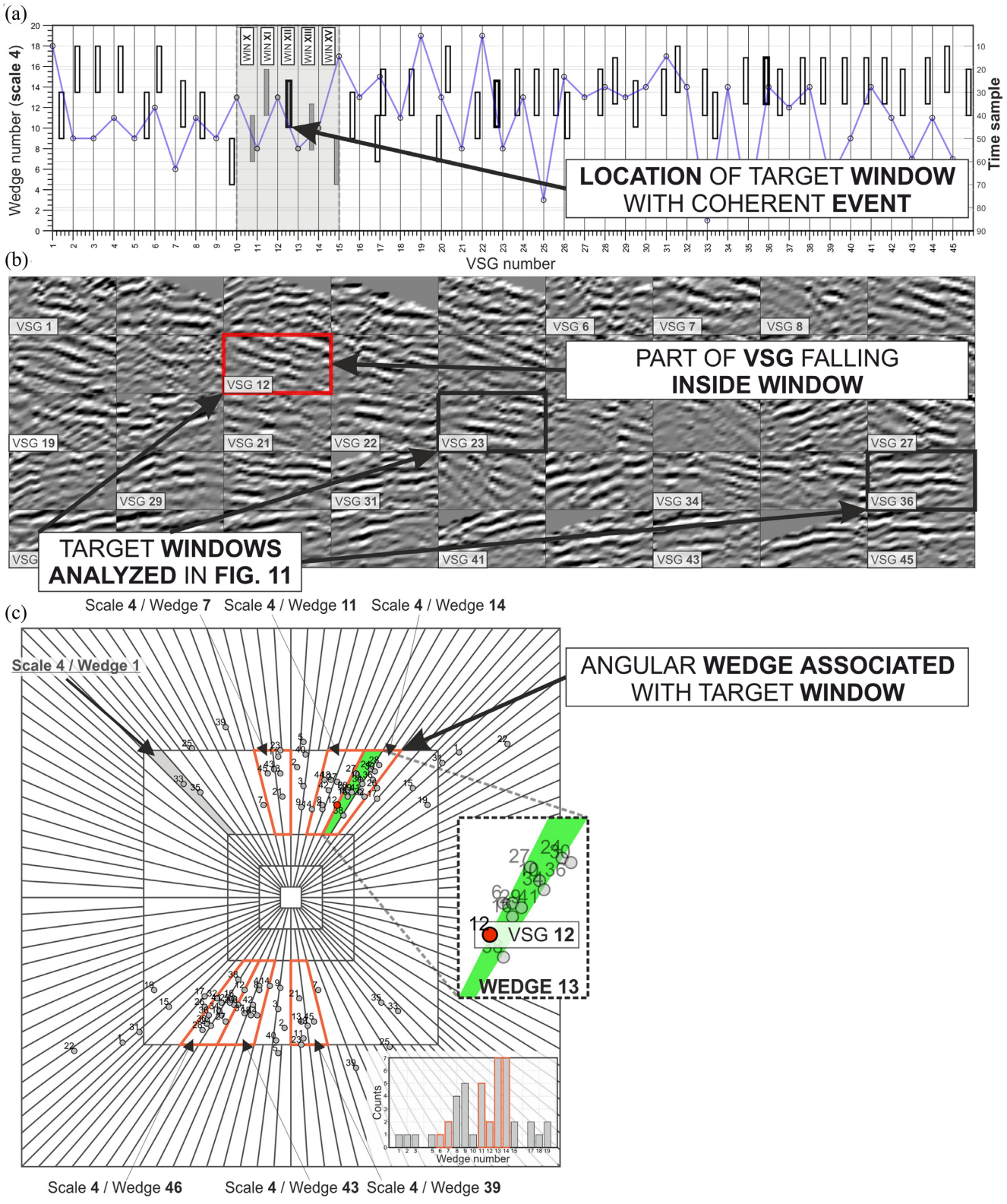

Figure 12. Automatic selection of angular wedges associated with maximum values in the *f*–*k* domain. (a) Location of windows selected for each VSG (denoted with grey boxes) and the number of angular wedges associated with a given window. The numbering of wedges refers to their position in the 4th scale and starts at 1 for the top left edge indicated with a grey-shaded polygon in (c). (b) Pre-selected windows containing coherent events extracted from every VSG along receiver line 6. The red- and black-outlined boxes correspond to the VSGs shown in Fig. 10, while boxes with only white labels are referring to VSGs shown in Fig. 11. The windows are characterized by a constant size of 20×20 samples and are individually designed for each VSG. The numbering of VSGs refers to the position along line 6. (c) Automatically picked maximum values of the panels shown in (b), projected on the mesh representing the 2-D Fourier plane partitioned by a discrete curvelet transform. Angular wedges outlined in orange are associated with the highest number of automatic picks. The numbering of data points refers to the position of VSGs along line 6. The green polygon indicates the angular wedge associated with the window selected for VSG 12.

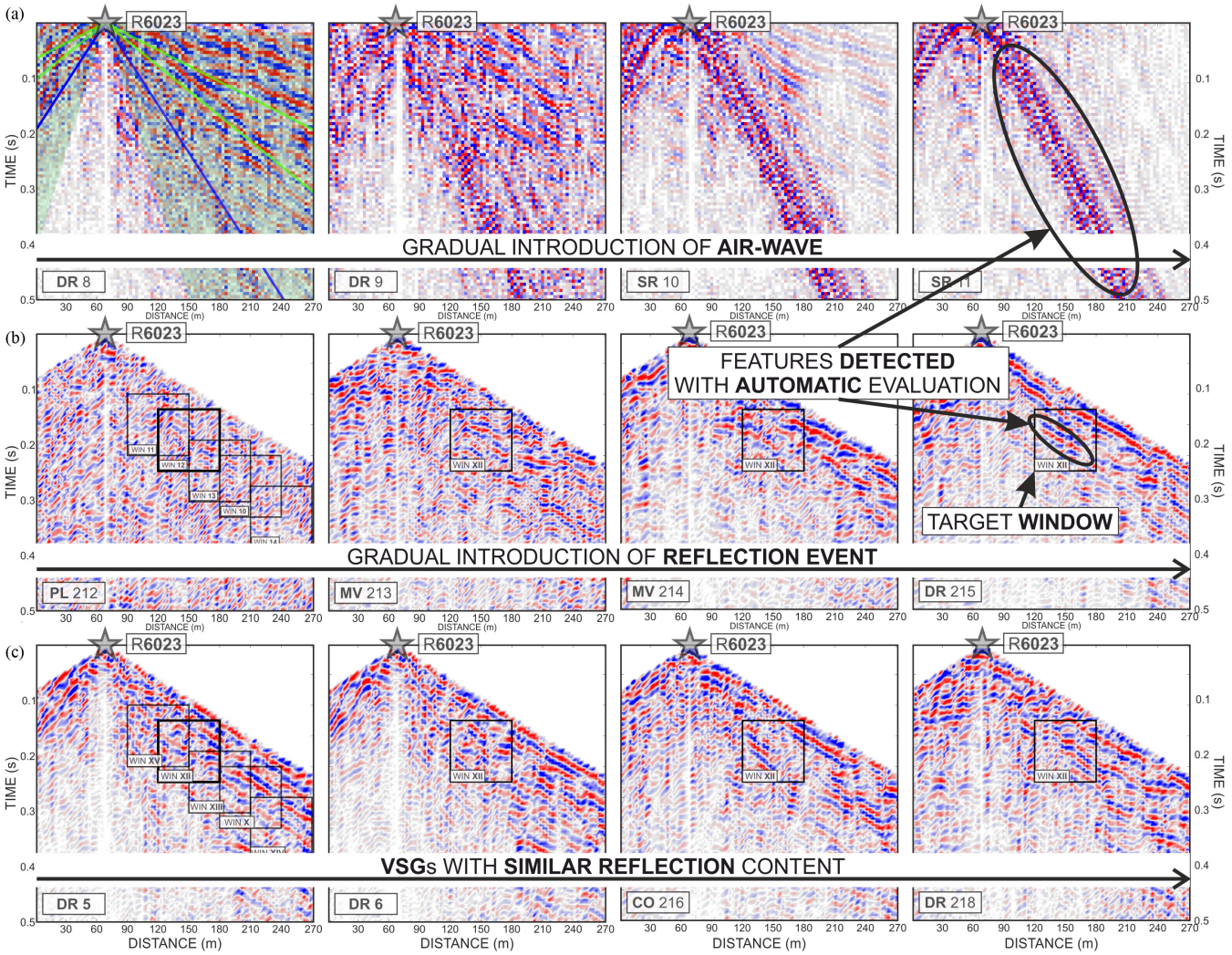

Figure 13. VSGs representing the gradual introduction of a single field-operation contribution (transition gathers) (a), (b) and VSGs representing similarity in radiation pattern and coherent events in the target zone (c). VSGs in (a) and (b) are used to assess the capacity of the illumination diagnosis and curvelet parametrization to detect specific contributions and determine threshold values representative for the study area. VSGs in (c) are used to address the ambiguity in VSG selection when a subset of VSGs with similar patterns is analysed. Note the progressive retrieval of the air-wave contribution in (a) related to the sequence of four consecutive field operations. Green lines indicate the range of scanned slownesses in the illumination diagnosis, while the blue line denotes the slope of the air-wave contribution. (b) VSGs illustrating the retrieval of a coherent event in the windows marked by a black rectangle. (c) VSGs obtained at the beginning of the survey (DR5 and DR6) and close to the end of the recording (CO216 and DR218). These panels have a similar radiation pattern and reflection content, thus their automatic evaluation requires both illumination diagnosis and curvelet parametrization. VSGs in (c), apart from the coherent events, contain a significant amount of artefacts, possibly contributing destructively during imaging.

presence of virtual sources characterized by similar reflectivity content for different VSGs at a single receiver position (Fig. 13c) and those with the high score in wedge parametrization (see Fig. S6) suggest the possibility to use other field operations for the retrieval of reflections. In this section, we aim to automatically quantify the potential of every drilling-related operation to generate VSGs resembling those in Fig. 10. In order to evaluate the reflectivity, we use two wedge parameters. For estimating kinematics of first arrivals we use illumination diagnosis [see Fig. S7 for the results of combined wedge/illumination-diagnosis characterization of VSGs representing gradual changes in the field operations shown in Fig. 13: PL212, MV213, MV214 and DR215 (Fig. 13b), and those with similar reflection content: DR5, DR6, CO216 and DR218 (Fig. 13c)].

In this way, it is possible to obtain the representation of VSGs obtained for every master trace along receiver lines 1–7 for every field operation in the 3-D space of illumination/wedge-parameters domain. VSGs obtained from the manual selection exhibit 2-D distribution in the 3-D space of illumination/wedge-parameters domain. These results are shown in Fig. S8. Generally, if a 3-D illumination/wedge-parameters characterization of a particular VSG falls inside the parameter subspace defined by the range of values typical for the manually selected VSGs then this particular VSG is expected to exhibit reflection content and first arrivals similar to the manually selected VSGs.

To highlight the influence of specific parts of the drilling survey in the top panel of Fig. 14(a), we show the activity sequence plot, where colour changes represent the W_{P1} values averaged over each

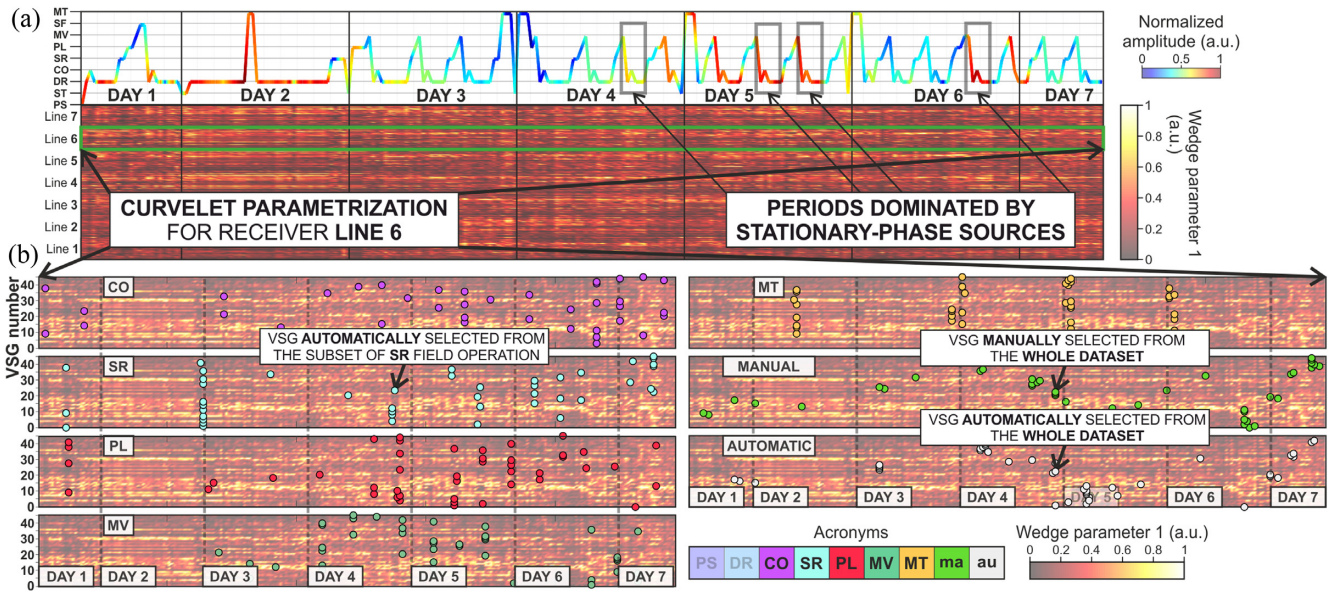

Figure 14. Automatic evaluation of the VSG reflection content and imaging potential of different field operations. The horizontal axes show the hours of the day, and the colour scales correspond to the normalized value of W_{P1}. (a) Top panel shows activity sequence plot. Line colour corresponds to the value of W_{P1} averaged over every master-trace position of the receiver lines 1–7, for a given time instance. Note the increase in W_{P1} coinciding with the right flank of the triangular sequence of activities during day 5. Bottom panels show curvelet parametrization for VSGs obtained at each of the lines 1–7. Receiver line 6 is denoted with the green rectangle. The maxima denoted with brighter colours represent the increase in W_{P1} in spatio-temporal windows of expected reflection events. (b) Comparison of VSGs curvelet parametrization obtained for receiver line 6. Colour dots denote time occurrence of VSGs selected with the manual and automatic approach used for the production of common midpoint (CMP) stacks in Fig. 16. White-dot VSGs are picked by selecting the highest values of W_{P1} using all available AN segments for retrieval. The rest of the automatically picked VSGs are picked by selecting the highest values from the subset of data related to a specific field operation. Note the similarity of picks distribution between the automatic and manual approaches.

AN segment for every analysed master-trace position (i.e. every column in the bottom panel of Fig. 14a). In Fig. 14(b) we show the W_{P1} distribution for receiver line 6, and dots denote the automatically picked VSGs with highest scores (for a single master-trace position we pick one VSG). The bottom panel in Fig. 14(a) shows the values of W_{P1} computed for master trace positions for every recorded AN segment. Thus, it represents the temporal variation of the reflectivity parametrization during the progress of the drilling survey. The maxima denoted with brighter colours represent the increase in wedge amplitudes in spatio-temporal windows of expected reflection events.

In the next section we assess the validity of the automatic selection of VSGs (Fig. 15) and 2-D stacks (Fig. 16). Stacks in Fig. 16 are obtained along receiver line 6 using the following sets of VSGs: (i) manually picked VSGs, (ii) VSGs with maximum values of the wedge parametrization automatically selected from all available AN segments and (iii) VSGs obtained from picking points with the highest curvelet parametrization from a subset of AN segments related to a single type of field operation. The automatic picking is done by analysing VSGs from all AN segments for every master-trace position along receiver line 6. For the 'Automatic approach', we pick the highest value of wedge parametrization out of all 246 available AN segments. In order to pick the VSGs for the specific field-operation (denoted with field-operation acronyms in Fig. 14b), we choose VSGs with the highest value of wedge parametrization only from a subset of AN segments related to the given field operation. Please note that in Fig. 14 for the display purposes we show only the values of W_{P1}, while for the actual selection we used both parameters: W_{P1} and W_{P2}.

5.4 Reflection imaging

The purpose of Fig. 11 was to highlight the capacity of the curvelet transform for the localized analysis. As shown in Fig. 11, the curvelet transform allows to focus on the specific curved events and assess their dip, amplitude and position. Considering a single VSG, the processing done in this subsection is as follows: we first analyse the predefined window; when the event in the pre-defined window meets the criteria, for further processing we choose the complete VSG. That is, the windowed part is only analysed during the curvelet scanning. The further processing sequence (strictly speaking, the conventional seismic imaging approach) is meant to be identical to the active-source data, so as input we use VSGs in the form which most resembles the active data.

The availability of coincident active-source data allows us to compare the active-source gathers with VSGs obtained using the automatic selection. Fig. 15 shows a comparison of the common-source gathers obtained along receiver line 6. The orange arrows on both active- and virtual-source gathers indicate a reflection event associated with the base of the iron-ore formation. Despite the retrieval of the main reflection event identified in the active-source data, the VSGs exhibit non-stationary arrivals and a moderate amount of noise. The differences in (i) the amplitudes in the passive and active shot gathers (yet with correct kinematics), (ii) the appearance of the reflector and (iii) variations of amplitudes amongst traces in the passive data, all arise mainly due to expected discrepancies from the requirements of the SI theory (and typical for field data). Primarily, to get the correct relative amplitude, the sources should surround the receivers and be in the far-field from them (Wapenaar & Fokkema 2006; see also Section 4.1 for a more detailed explanation). In the

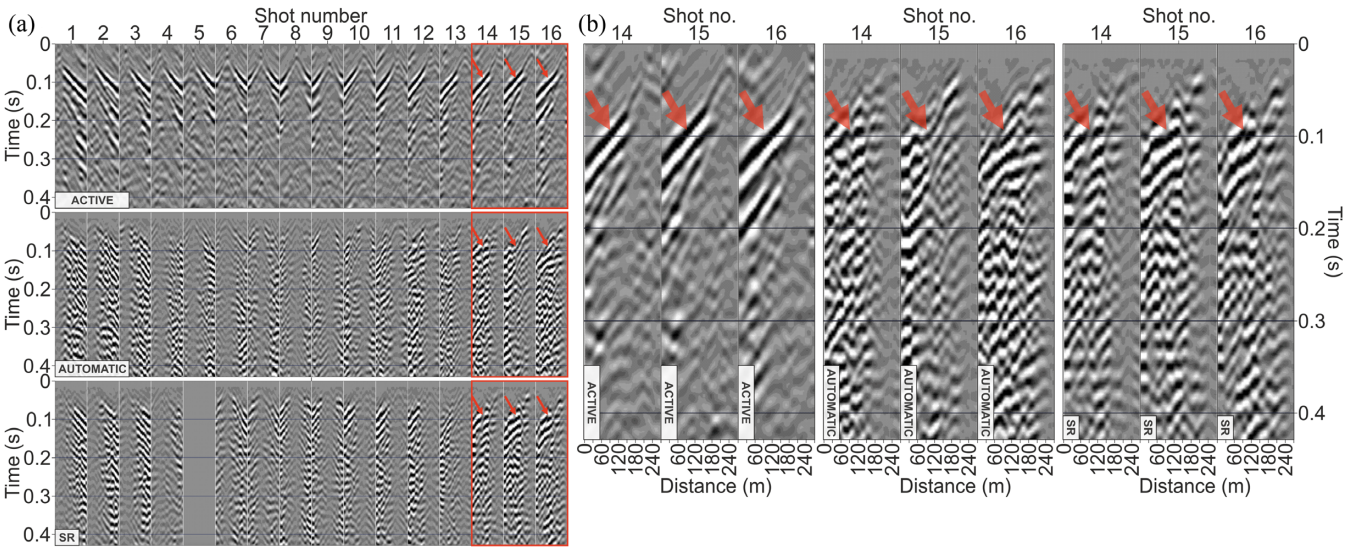

Figure 15. Comparison of collocated common-source gathers along receiver line 6 obtained using active and passive data. (a) Juxtaposition of 16 common-source gathers obtained using active-source data (top row), automatically selected VSGs (middle row), and automatically selected VSGs using only noise segments recorded during SR field operations (bottom row). For the 5th panel showing the VSG obtained using SR (bottom panel), retrieval of reflections was not possible, because all master-trace recordings for this VSG were dominated by random noise. (b) Common-source gathers from the orange boxes in (a). The orange arrows highlight reflection arrivals contributing to the main reflector observed in CMP stacks in Fig. 16.

second place, the small discrepancy between the same reflectors observed in passive and active data is related to the inherent difference in frequency between passive and active sources, as well as possible attenuation effects. We discuss this in more detail in Section 6.3.

When a relatively high SNR characterizes the main reflection event, the quality of the retrieved results should suffice for imaging purposes. The presence of both the clear reflection event and the artefacts in the VSGs can be attributed to the content of the 15-min-noise segments used for CC. As expected from the temporal changes of AN identified in Section 3, the dominant contribution comes from a particular single field operation that provides the stationary-phase sources. The secondary contributions, treated here as artefacts, result from the CC of the recorded fragments of other field operations performed outside the assumed stationary-phase area, which adjoin the main category listed in the field notes (see Fig. 3 for the summary of field notes).

We compare the VSGs stacks with the stack along line 6 from the 3-D active-source survey (Fig. 16). Note that for the latter, we used only 18 shot points (inline shots only), whereas for the passive data stacks we took VSGs at every 2nd receiver station (*ca*. 2 times higher stacking fold). The processing sequence, identical for the passive and active data, includes muting of first-arrivals, trace-by-trace amplitude balancing, mild *f*–*k* filtering to remove dipping events, bandpass filtering (10–20–50–70 Hz), constant velocity ($V = 2200$ m s^{-1}) normal-moveout correction and common-midpoint stacking. The active and passive stacks show reflectivity down to 0.3 s two-way traveltime with the strongest reflectivity present in the top 0.15 s. The most prominent event in the active data is the horizontal reflector at approximately 0.1 s, stretching along the whole section.

This reflector is associated with the base of the detrital contact forming a shallow unconformity; beneath this contact is the iron formation. Arrows 1, 2 and 3 show the base of the detrital contact. On the stacks from VSGs, the same event is best retrieved using manually selected and automatically picked VSGs. The operation-wise stacks exhibit sparser reflectivity content as compared to the manual and automatic approaches; however, the trend of horizontal reflectivity at around 0.1 s is consistently retrieved for every drilling-related operation. Apart from the strong reflector at 0.1 s, the shallow reflectivity above this reflector is related to changing lithology of the detrital and erosional surfaces, such as channels. The reflectivity within the iron formation is more variable and is likely to be related to the changing lithology, in particular the increasing shale content, and the increased density associated with high iron content. Discontinuities in the reflectivity in the iron formation can be attributed to faults and joints. Several features have been interpreted in the active seismic image that was also seen in the AN images. Arrows 4 and 5 may indicate a fault within the iron formation that does not extend up through the unconformity into the detrital, while arrow 6 indicates the base of the iron formation.

In Fig. 14(a), we showed the summary of the curvelet parametrization performed for each dense receiver line. Now, because the curvelet parametrization targets directly the reflection hyperbolae, the increase of values shown in Fig. 14 might be interpreted as an increase of the 'imaging potential' of the given noise segment. Here, we show stacks obtained only for receiver line 6 (denoted with the green box in Fig. 14a), but we emphasize the similarity of the curvelet parametrization for the remaining receiver lines. It means that VSGs obtained for those lines exhibit similar reflectivity, and consequently would result in similar stacks.

6 DISCUSSION

6.1 Temporal patterns in the recorded AN

The analysis of the temporal changes in the noise spectrograms, single receiver-pair CCF and VSGs revealed a strong dependency on the progress and changes in the processes associated with the drilling operations. The amount of data for Figs 4–6 was chosen based on the trial and error approach, such that it highlights the variability of recorded passive data, but at the same time, the possibility to observe the regularity of the impact of drilling-related

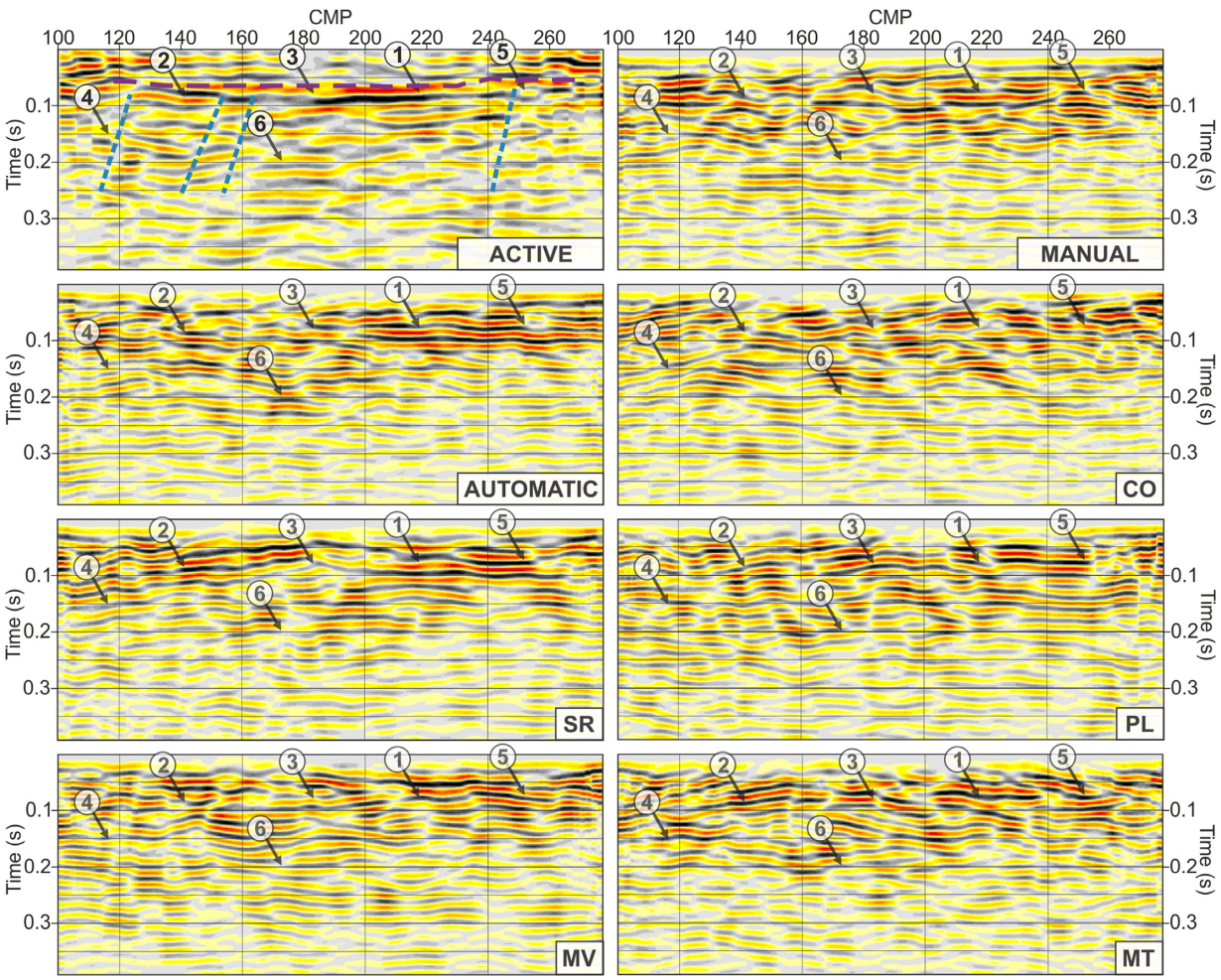

Figure 16. Comparison of CMP stacks of active-source gathers and VSGs selected in Fig. 14(b). Arrows 1, 2 and 3 show the base of the detrital contact, arrow 4 and 5 may indicate a fault within the iron formation that does not extend up through the unconformity into the detrital, and arrow 6 indicates the base of the iron formation.

operations on the recorded AN. The shape of the events retrieved in VSGs is inherited from the waveforms present in the noise panels used for CC. The field operations recorded by PilbArray contribute to retrieval of VSGs dominated by the air wave (mostly SF and SR activity), hyperbolic events (DR and CO), and random noise (MT and PS). Please note here, that the above interpretation is limited by the accuracy of the field notes, and the assumption that during a 15-min-long segment no more than one activity could have occurred. Nevertheless, the potential explanations for the operation-related changes observed in CCFs, and PSDs, may be derived from Fig. 3(c), where the detailed descriptions of each field operation are provided. Only a few of the activities provided VSGs with illumination characteristics and reflection content resembling the active-source data. The field operations considered here covered most of the spatial extent of the PilbArray. It gave the rare opportunity to pinpoint stationary-phase sources at almost all the receiver positions (note that the receiver stations collocated with boreholes were excluded from the analysis).

The inspection of noise spectrograms revealed an increase of amplitude in specific frequency ranges depending on the current operation. It proved the capacity of using the AN from PilbArray to track certain field operations relevant for our study. Restricting the analysis to the inspection of AN segments recorded during the time-span of a single operation suggested that the predominant contribution comes from spatially localized noise sources generating waveforms with a very similar moveout (waveforms shown in Fig. S4 represent predominant contributions to associated 15-min-long AN segments). The above observations explain the convergence and repeatability in the virtual-source patterns. Repeatable patterns observed in the noise spectrograms allow to distinguish between field operations and to relate a specific type of patterns to virtual-source responses. Analysis of Figs 6 and 7(a) reveals that surface and air waves were the most persistent arrivals during the whole recording time, thus, as evidenced by the VSG obtained from stacking over all data (see red ellipses and violet line in Fig. 7b), they preclude the retrieval of reflections when selective stacking is not used. Consequently, the visual inspection of virtual-source responses obtained from noise panels dominated by a single type of AN source allows to selectively stack only those panels which are dominated by body waves and assure that no parts of AN data dominated by surface waves are used for stacking.

6.2 AN processing

The length of the noise segments in this study (15 min) was chosen to match the field-operation description in the observer notes and

sufficient convergence of CCFs. Selection of shorter segments (see Almagro Vidal *et al.* (2014) for criteria of choosing the length of correlation panels for imaging) would result in much more virtual-source responses available for every receiver position (e.g. for 10-s-long segments the number of VSGs retrieved for a single receiver position would increase from the current 246 to 22 140). Since the virtual-source responses are directly connected to the repeatable patterns in the recorded AN, it would be possible to process AN in shorter segments by automatic detection of patterns recognized primarily in 15-min-long segments.

A further processing improvement might come from accounting for both the stationary-phase region and positive and negative times during the visual inspection of VSGs. The purpose of Fig. 7 is to explain: (i) the difference between selective stacking and using the all-noise approach and (ii) to show different approaches of stacking the causal and acausal parts. In Fig. 7(b), we showed how different processing approaches could yield a specific contribution. By showing different retrieved events, with some of them resembling the conventional seismic events, it is easy to comprehend the idea of the impact of drilling-related noise on the reflection imaging using VSGs. While for the PilbArray it was sufficient to utilize a single field operation per receiver station serving as a master trace, for other similar data sets the VSGs might need to be obtained using either AN segments from two different field operations and/or combination of the causal and acausal part of the correlated panels. Such a joint analysis of the causal and acausal part allows selecting only those parts of the correlated panels with properly shaped events (i.e. hyperbolic reflection moveouts). For instance, as proposed by Ruigrok *et al.* (2010), if the dip of the retrieved events increases with offset, the VSG may be obtained from summing only specific offsets from the positive and negative times of the correlated panel. Exemplary results for this approach are shown in Fig. S9, where VSGs for a master trace position 4045 were obtained using combinations of 16 different segments of correlated panels.

Continuous retrieval of virtual-source responses for a fixed reference receiver allows retrieving VSGs characterized by a similar slope of the first arrivals and presence of coherent events (see Fig. 13c). These VSGs illustrate the ambiguity in the selection of stationary-phase sources: for each master traces position we can find few visually similar VSGs with first arrivals of a similar slope, and containing reflection events. As a remedy, we proposed an objective parametrization of the retrieved reflectivity using curvelet-transform-derived measures.

The semi-automatic curvelet tool is used with the aim to minimize the risk of selecting VSGs containing spurious arrivals in the analysed time window. In our automatic approach, we choose the result (VSG), which is the most trusted, and for this study, the most similar to the result selected during the visual inspection. We tested the robustness of the method by investigating the following factors affecting the automatic evaluation: (i) VSGs with different retrieved contributions (including spurious arrivals), (ii) different levels of noise present in the window and (iii) slightly varying curvatures and amplitudes of the reflection events. For these tests, we used transition VSGs shown in Fig. 13 and representing the factors mentioned above. The results of these tests are shown in Figs S6 and S7. The takeaway message from these figures is, that curvelet parametrization selects the same VSGs as previously recognized during the visual inspection (Fig. S6a) and that it allows to effectively distinguish between even very similar VSGs (Figs S6b and S7). Furthermore, the curvelet parametrization yields a whole range of values (represented by a cloud of grey points in Fig. S6; however, most VSGs cluster around similar scores and the targeted—visually recognized VSGs—are characterized by the relatively highest values, shown with green points in Fig. S6). Therefore, our method selects the VSGs having the highest probability of containing the correct reflections out of all analysed data.

The curvelet transform allowed us to omit the limitation of the global nature of the *f*–*k* transform by accurate partitioning of the 2-D Fourier representation of the seismic data into localized anisotropic wedges, related to the specific target zones in the VSGs. The efficiency of this approach is confirmed by the ability to detect spatiotemporally focused coherent events in VSGs using two wedge parameters. The same parametrization applied to the transition VSGs exhibiting the same retrieved events with gradually increasing SNR (Figs 13a and b) might be used to determine the threshold values specific for a given study and useful for retrieval of VSGs with the desired type of arrivals.

6.3 Field operations with lower imaging potential

As already mentioned, SI processing applied to different field operations resulted in the retrieval of different arrivals (Fig. 7a). The parts of the VSGs excluded from reflection imaging might be used for other monitoring or imaging purposes. In particular, panels related to MT102-104 seem feasible for surface wave extraction. The curvelet parametrization indicates the potential of using field operations other than drilling for retrieval of reflections (e.g. CO). The primary motivation behind utilizing AN segments not related to DR activity is to obtain virtual-source response at receiver locations which do not fall in stationary regions of DR-induced seismicity. Comparison of 2-D stacks obtained from operation-wise stacking indicates that some operations (e.g. SR and CO) have the potential to be used as standalone input for imaging. Some VSGs exhibit reflection arrivals earlier than in the active data, and consequently, some passive reflectors in Fig. 16 are also located earlier than the same reflectors in the active stacks. One of the main reasons for the above discrepancy is that the ‘original’ frequencies in VSGs and active-source gathers are different. Please note that this difference in frequencies appears despite using the same bandpass filter for the active and passive data. We have to accept that additional processing effort will not fully overcome the natural frequencies of recorded noise sources. Nevertheless, the nature of the retrieved passive reflections, from this point of view, is physically correct; but because of the difference in source types used for the active and passive data, the same targets are illuminated in slightly different ways. Another reason, which theoretically leads to differences between the same reflection observed in passive and active data, is the attenuation, possibly producing phases at different times. The last factor is related to the differences between the wavelets of passive and active data. The wavelet in the active data is not characterized by a zero-phase wavelet, while CC yields a zero-phase wavelet. Field operations with lower imaging potential (see CMP stacks in Fig. 16) exhibit sparse reflectivity collocated with events observed in the active-source data. Therefore, they might still be used to complement the stationary-source regions previously uncovered by illumination from DR activity. The preponderance of the CO (collaring) field operations for imaging might be related to the fact that similarly to DR operations, they represent the actual drilling activity (see Fig. 3c). Based on this observation, we propose a workflow of semi-automatic curvelet-based reflectivity evaluation (see flowchart in Fig. C1).

6.4 Practical implications

Here, we would like to emphasize four temporarily limited recording periods dominated by stationary-phase sources representing 6–7 AN segments (i.e. equivalent to 90–105 min of recording). These four periods (indicated in Fig. 9b) include recordings of 24 field operations (∼9 per cent of all recorded field operations) and contribute to the production of 214 VSGs (∼34 per cent of the analysed receiver positions). Aligning these segments with an activity sequence plot and CCF (Fig. 9b) reveals their coincidence with the same acquisition process occurring between the triangular patterns described in Section 3.1. Field operations contributing to these sequences are DR, CO and MT. We argue that introducing small adjustments to the drilling operations by performing these field operations additionally on the edges of the recording array could broaden the coverage of stationary-phase sources, such that every receiver position could be turned into a virtual source. The identified segments providing many stationary-phase sources for retrieval of reflections, cover a time window of 1.5 hr overall. Incorporating this observation when designing similar passive experiments could help reduce the necessary time of recording to only certain parts of the drilling campaign.

6.5 Possible explanations for the operation-related changes in CCF and PSD and the essence of their comparison

The temporal AN variations recorded by PilbArray were investigated in Figs 4–6. The explanations provided in Sections 3.2–3.4give our interpretation about linking AN changes with a single type of field operation (Section 3.2), with a sequence of field operations (Section 3.3) and how each of the two contributes to the virtual-source data (Section 3.4). Moreover, we explained that some field operations contribute to the retrieval of specific arrivals (these are highlighted with labels in Figs 5 and 6). To further complete these interpretations, in Section 6.3, we explain nuances about the MT and CO field operation, which contribute to the retrieval of surface and body waves, respectively. In the same subsection, we also discuss how other field operation can be useful for AN studies. To highlight the dependency of the AN behaviour on the changes of the drilling survey, we provide the exact exemplary panels (see the labels in Figs 5 and 6). These explanations are supported by the description of the field operations, detailing what actual activities contributed to a given operation (Fig. 3). Nevertheless, within segments of 15 min, more than one operation could have happened. AN characteristics for a single 15-min-long noise panel can be determined by: (i) single type of field operation, (ii) few overlapping types or (iii) joint single type and the longer, temporal trend associated with the repetitive sequence identified in Fig. 4. In Section 5.4, these phenomena are used to explain specific arrivals that are observed in our VSGs. By this, we manage to show that by focusing on specific features of AN, and their anthropogenic causes, it is possible to improve the performance of the passive imaging and aid the interpretation of otherwise unknown contributions in the retrieved virtual-source data. Finally, as summarized by Fig. 9 and Section 6.4, we once again focus on the specific sequences of field operations, but this time only on those which contribute to a high number of stationary-phase sources. We provide the exact type of field operations and highlight the AN patterns associated with them. This is done toward directly explaining which parts of the drilling survey should be carefully recorded, or eventually how to optimize the acquisition, by performing some field operations (assuming they do not significantly increase the cost of the survey) specifically for the purpose of obtaining a more isotropic illumination during an AN acquisition.

6.6 Further methodological developments

The semi-automatic curvelet-based reflectivity evaluation that is developed as a supporting tool for this study could be further automatized. For example, by robust picking of polygons in the *f*–*k* domain and associating them with the high-amplitude regions could allow to automatically cover targeted coherent events with the moving windows and remove the necessity for their initial manual selection. Further improvements could involve incorporating illumination diagnosis as an initial step for rejection of VSGs dominated by high-slowness arrivals (e.g. see Fig. 13a) and decrease the number of VSGs for curvelet scanning to only those, which fall inside the range of velocities typical for body waves. The summary of the illumination diagnosis for the PilbArray and the relevant velocity limits are provided in Fig. S10.

The segmentation of the continuous data chosen in this study enables visual inspection of all virtual-source responses. However, SI with AN could be utilized for more detailed analysis, using much shorter time windows (e.g. Cheraghi *et al.* 2017), which could potentially bring higher SNR in stacking virtual-source responses (see Draganov *et al.* (2013) for comparison of different stacking approaches). Applying curvelet wedge parametrization to VSGs obtained for such shorter segments could allow obtaining an extensive database of reflection-related parameters describing the AN wavefield. Even though the values would be specific for a given study, this database could be further used to add new insight into the processing of the data sets similar to the PilbArray. It should allow narrowing the analysis to clusters of virtual-source responses characterized by similar, highly scored values in the wedge parametrization. Therefore, it has the potential to be implemented in transfer learning. However, we argue that the primary focus of using machine learning in the proposed workflow is in turning the semi-automatic evaluation to automatic (see flowchart in Fig. C1, Chamarczuk *et al.* 2020).

7 CONCLUSIONS

The temporal variations of AN can be used to identify contributions of specific drilling and non-drilling related field operations accompanying the routine drilling campaigns. Using passive seismic interferometry, we showed that virtual-source responses similar to active-source data could be obtained by selection of noise segments associated with field operations related to specific AN characteristics. Next, we used this knowledge to obtain stationary VSGs for most positions inside the PilbArray. The visually inspected VSGs were subsequently used to derive data-driven parameters accounting for their reflection content. Finally, we designed a semi-automatic method for detection of VSGs with reflection content and used this approach to assess the imaging potential of different field operations. We note that the general conclusions and recommendations for future studies should be separated from the specific findings for the case study with the PilbArray data set. Consequently, considering the geological conditions of the Pilbara region, and specific type of passive seismic survey acquired during a drilling campaign, this study proved the ability to image shallow iron ore mineralization hosting strata. To provide the general conclusions for future

studies, we developed a generic workflow that allows assessing the reflection imaging potential of specific segments of AN data. The final, objective takeaway message is utilizing the whole recorded AN, rather than focusing only on drilling periods, to optimize subsurface imaging.

ACKNOWLEDGEMENTS

DCT algorithm was taken from the CURVELAB project (www.curvelab.org). We thank BHP field crew and drillers for providing this high-quality passive dataset and field notes allowing to reveal more insights about AN content. M. Malinowski and M. Chamarczuk were supported by the National Science Centre (Poland) grant no UMO-2018/30/Q/ST10/00680 (FULLIMAGE project). Comments by N. Nakata, M. Chmiel and two anonymous reviewers are much appreciated.

DATA AVAILABILITY

Based on the research collaboration agreement with our sponsor and considering the nature of the dataset used in this publication we are not able to make the dataset accessible to the public at the date of publication.

REFERENCES

Almagro-Vidal, C., Draganov, D., van der Neut, J., Drijkoningen, G. & Wapenaar, K., 2014. Retrieval of reflections from ambient noise using illumination diagnosis, *Geophys. J. Int.,* **198,** 1572–1584.

Asgharzadeh, M., Grant, A., Bòna, A. & Urosevic, M., 2019. Drill bit noise imaging without pilot trace, a near surface interferometry example, *Solid Earth,* **10,** 1015–1023.

Behm, M., Leahy, G.M. & Snieder, R., 2014. Retrieval of local surface wave velocities from traffic noise–an example from the La Barge basin (Wyoming), *Geophys. Prospect.,* **62,** 223–243.

Brenguier, F. *et al.*, 2019. Train traffic as a powerful noise source for monitoring active faults with seismic interferometry, *Geophys. Res. Lett.,* **46,** 9529–9536.

Brenguier, F. *et al.*, 2020. Noise-based ballistic wave passive seismic monitoring. Part 1: body waves, *Geophys. J. Int.,* **221,** 683–691.

Candès, E.J. & Donoho, D.L., 1999. Curvelets—a surprisingly effective non adaptive representation for objects with edges, in *Curves and Surface Fitting: Saint-Malo 1999,* eds Cohen, A., Rabut, C. & Schumaker, L., pp. 105–120, Vanderbilt Univ. Press.

Chapman, H., 1981. Generalized Radon transforms and slant stack, *Geophys. J. Int.,* **66,** 445–453.

Curtis, A., Gerstoft, P., Sato, H., Snieder, R. & Wapenaar, K., 2006. Seismic interferometry–turning noise into signal, *Leading Edge,* **25,** 1082–1092.

Chamarczuk, M. *et al.*, 2019. Automatic 3D illumination-diagnosis method for large-N arrays: robust data scanner and machine-learning feature provider, *Geophysics,* **84,** 1572–1584.

Chamarczuk, M., Nishitsuji, Y., Malinowski, M. & Draganov, D., 2020. Unsupervised learning used in automatic detection and classification of ambient–noise recordings from a Large-N array, *Seismol. Res. Lett.,* **91,** 370–389.

Cheraghi, S., Craven, J.A. & Bellefleur, G., 2015. Feasibility of virtual source reflection seismology using interferometry for mineral exploration: a test study in the Lalor Lake volcanogenic massive sulphide mining area, Manitoba, Canada, *Geophys. Prospect.,* **63,** 833–848.

Cheraghi, S., White, D.J., Draganov, D., Bellefleur, G., Craven, J.A. & Roberts, B., 2017. Passive seismic reflection interferometry: a case study from the Aquistore CO_2 storage site, Saskatchewan, Canada, *Geophysics,* **82,** B79–B93.

Draganov, D., Campman, X., Thorbecke, J., Verdel, A. & Wapenaar, K., 2009. Reflection images from ambient seismic noise, *Geophysics,* **74,** A63–A67.

Draganov, D., Campman, X., Thorbecke, J., Verdel, A. & Wapenaar, K., 2013. Seismic exploration-scale velocities and structure from ambient seismic noise (>1 Hz), *J. geophys. Res.,* **118,** 4345–4360.

Girard, A. & Shragge, J., 2019. Automated processing strategies for ambient seismic data, *Geophys. Prospect.,* **68,** 293–312.

Gòrszczyk, A., Adamczyk, A. & Malinowski, M., 2014. Application of curvelet denoising to 2D and 3D seismic data – practical considerations, *J. appl. Geophys.,* **105,** 78–94.

Halliday, D., Curtis, A. & Kragh, E., 2008. Seismic surface waves in a suburban environment: active and passive interferometric methods, *Leading Edge,* **27,** 210–218.

Johnson, D.H. & Dudgeon, D.E., 1993. *Array Signal Processing: Concepts and Techniques,* Pearson.

Kumar, V., Oueity, J., Clowes, R.M. & Herrmann, F.J., 2011. Enhancing crustal reflection data through curvelet denoising, *Tectonophysics,* **508,** 106–116.

Liu, Y., Draganov, D., Wapenaar, K. & Arntsen, B., 2016. Retrieving virtual reflection responses at drill-bit positions using seismic interferometry with drill-bit noise, *Geophys. Prospect.,* **64,** 348–360.

Matsuoka, T., Shiraishi, K., Onishi, K. & Aizawa, T., 2006. Application of seismic interferometry to subsurface imaging, in *Proceedings of 10th International Symposium on RAEG,* Kyoto University Geophysical Society, pp. 35–38.

McNamara, D.E. & Buland, R.P., 2004. Ambient noise levels in the continental United States, *Bull. seism. Soc. Am.,* **94,** 1517–1527.

Mehta, K., Snieder, R., Calvert, R. & Sheiman, J., 2008. Acquisition geometry requirements for generating virtual-source data, *Leading Edge,* **27,** 620–629.

Minato, S., Tsuji, T., Matsuoka, T. & Obana, K., 2012. Crosscorrelation of earthquake data using stationary phase evaluation: insight into reflection structures of oceanic crust surface in the Nankai trough, *Int. J. Geophys.,* **2012,** doi:10.1155/2012/101545.

Miyazawa, M., Snieder, R. & Venkataraman, A., 2008. Application of seismic interferometry to extract P- and S-wave propagation and observation of shear-wave splitting from noise data at Cold Lake, Alberta, Canada, *Geophysics,* **73,** D35–D40.

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

Nakata, N., Chang, J.P., Lawrence, J.F. & Boue, P., 2015. Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *J. geophys. Res.,* **120,** 1159–1173.

Olivier, G., Brenguier, F., Campillo, M., Lynch, R. & Roux, P., 2015. Body-wave reconstruction from ambient seismic noise correlations in an underground mine, *Geophysics,* **80,** KS11–KS25.

Panea, I., Draganov, D., Almagro-Vidal, C. & Mocanu, V., 2014. Seismic interferometry of railroad induced ground motions: body and surface wave imaging, *Geophysics,* **79,** Q31–Q42.

Poletto, F.B. & Miranda, F., 2004. *Seismic While Drilling: Fundamentals of Drill-Bit Seismic Exploration,* Vol. **35,** Elsevier.

Quiros, D.A., Brown, L.D. & Kim, D., 2016. Seismic interferometry of railroad induced ground motions: body and surface wave imaging, *Geophys. J. Int.,* **205,** 301–313.

Rector, J.W., III & Marion, B.P., 1991. The use of drill-bit energy as a downhole seismic source, *Geophysics,* **56,** 628–634.

Roots, E., Calvert, A. & Craven, J., 2017. Interferometric seismic imaging around the active Lalor mine in the Flin Flon greenstone belt, Canada, *Tectonophysics,* **718,** 92–104.

Roux, P. & Kuperman, W.A., 2004. Extracting coherent wavefronts from acoustic ambient noise in the ocean, *J. acoust Soc. Am.,* **116,** 1995–2003.

Ruigrok, E., Campman, X., Draganov, D. & Wapenaar, K., 2010. High-resolution lithospheric imaging with seismic interferometry, *Geophys. J. Int.,* **183,** 339–357.

Sànchez-Pastor, P., Obermann, A., Schimmel, M., Weemstra, C., Verdel, A. & Jousset, P., 2019. Short- and long-term variations in the Reykjanes geothermal reservoir from seismic noise interferometry, *Geophys. Res. Lett.,* **46,** 5788–5798.

Schuster, G., 2001. Theory of daylight/interferometric imaging: tutorial, in *Proceedings of the 63rd Conference & Technical Exhibition,* European Association of Geoscientists and Engineers, Extended Abstracts, Session A32.

Sens-Schönfelder, C. & Wegler, U., 2006. Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano, Indonesia, *Geophys. Res. Lett.,* **33,** L21302.

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

Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E.,* **69,** 46610.

Snieder, R., Wapenaar, K. & Larner, K., 2006. Spurious multiples in seismic interferometry of primaries, *Geophysics,* **71,** SI111–SI124.

van der Neut, J., 2013. Downhole interferometric illumination diagnosis and balancing, *Geophys. Prospect.,* **61,** 352–367.

Wapenaar, K., 2003. Synthesis of an inhomogeneous medium from its acoustic transmission response, *Geophysics,* **68,** 1756–1759.

Wapenaar, K., 2006. Green's function retrieval by cross-correlation in case of one-sided illumination, *Geophys. Res. Lett.,* **33,** L19304.

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

Wapenaar, K., van der Neut, J. & Ruigrok, E., 2008. Passive seismic interferometry by multidimensional deconvolution, *Geophysics,* **73,** A51–A56.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Data consistency of 61.5 hr of recording the ongoing drilling measured by RMS amplitude variations. Colour background shows amplitude variations per single receiver station, white lines are values averaged over receivers belonging to the same recording line. The bottom plot shows amplitude values averaged over the whole recording—mean value for single receiver station (blue), average over receiver lines (red). The right-hand panel shows the sequence of field operations provided in Fig. 3. Each field operation is considered to last 15 min and aligned such that every row of the colour plot represents RMS amplitude value of AN recorded during a given field operation. Grey shaded bars indicate AN segments for which noise spectrograms and VSGs are analysed in more detail in this study.

Figure S2. Temporal variation of noise spectrograms during the second half of the recording time. Each panel represents PSD computed for 15-min-long AN segments recorded by the whole PilbArray. The acronyms in white rectangles denote the type of field operation performed at the time of recording and its ordinal number counting from the start of recording time.

Figure S3. Temporal variation of virtual source responses during the second half of the recording time aligned in the same way as PSD plots in Fig. S2. Each panel represents VSG from 15-min-long AN segment, obtained by cross-correlating trace recorded at receiver 4045 with every other receiver from receiver line 4 (see brown triangles in Fig. 2b). The acronyms in white rectangles denote the type of field operation performed at the time of recording and its ordinal number counting from the start of recording time.

Figure S4. Waveforms recorded by the whole PilbArray during the first half of recording time, and extracted from AN segments shown in Fig. 5(d). These events are representative of the labelled noise segments. The dashed rectangle and grey triangle indicate the location of receiver line 4 and a master trace (receiver 4045), respectively (see horizontally aligned brown triangles in Fig. 2b), used to obtain virtual source responses (see Fig. 7a). The same part of the data was used to obtain *f–k* plots shown in the right panels.

Figure S5. VSGs after top-muting above the first arrivals, and computed for perpendicular lines crossing each other at the master-trace positions shown in Fig. 2(b). The locations of the master traces were chosen to include the presence of multiple stationary-phase areas provided by drilling-related field operations. For each master-trace position, we used the same AN segment in the crossline and inline direction. Note the similarity in the radiation pattern of all retrieved VSGs.

Figure S6. Curvelet parametrization of VSGs. Green dots denote manually selected VSGs, grey dots denote VSGs for all remaining AN segments for a master-trace position indicated in the top right part of plots, and blue dots denote VSGs for all AN segments, for every master-trace position along receiver lines 1–7. (a) Parametrization for VSGs obtained at the 23rd receiver in receiver lines 1-7 shown in Fig. 1. (b) Parametrization of 20 VSGs along line 6 (see Fig. 10 for relevant VSGs). Note the similar distribution of points for every master-trace position. The red outlining denotes master-trace position analysed in Figs 13 and B2. (c) The leftmost panel shows curvelet parametrization of transition VSGs representing gradual changes in the field operations shown in Fig. 13: PL212, MV213, MV214 and DR215 representing gradual changes in field operations (see Figs 13b and c), and those with similar reflection content: DR5, DR6, CO216 DR218 (Fig. 13c). This parametrization is further juxtaposed with all 246 VSGs obtained for the same receiver position (middle panel), and finally against parametrization of all receiver positions and AN segments from the PilbArray (right panel). Note the high values (above 0.1) in the curvelet parametrization of manually selected VSGs (denoted with green dots), as well as the relatively lower values of manually selected VSGs for other receiver lines.

Figure S7. Curvelet and illumination-diagnosis analysis of VSGs obtained at receiver position 6023. (a) Comparison of different 2-D juxtapositions of two wedge parameters (W_{P1}, W_{P2}; see Appendix B). Note that parameters derived from *f–k* wedge representation and wedge amplitudes in the time domain provide the most evident clustering of the two VGSs with identified reflection content. Dashed ellipse indicates the two VSGs which are expected to obtain the highest score. (b) Illumination-diagnosis values plotted against W_{P2}. Clearly, transition VSGs dominated by the air-wave contribution are separated from those with faster arrivals. (c) Combination of illumination diagnosis with wedge parameters: the illumination diagnosis clearly allows to separate air-wave-dominated panels from those dominated by higher-velocity arrivals. (d) The juxtaposition of two wedge parameters (W_{P1}, W_{P2}) selected in this study to compare the imaging potential of different field operations.

Figure S8. Combined illumination-diagnosis/curvelet-reflectivity evaluation of different field operations and their potential for reflection retrieval. We analyse VSGs obtained for every receiver position using 246 noise segments resulting in 77 400 data points distributed into groups of 9 field operations recorded by PilbArray (panels a–i). Each denoted data point (grey circle) represents a VSG obtained at a single receiver position using a single 15-minng AN segment.

VSGs are plotted in the 3D s-pace defined by parameters derived from illumination diagnosis (slowness) and curvelet parametrization (wedges). (j) The same parametrization of manually selected VSGs. The range of values associated with points in (j) is plotted as a dashed polygon in (a)–(i).

Figure S9. Causal-wise versus operation-wise stacking analysis using VSGs from AN segments recorded during two consecutive field operations and the combination of their causal and acausal parts. All VSGs are obtained for the same master-trace position 4045, which is cross-correlated with all the receivers from line 4 (see horizontally aligned brown receivers in Fig. 2b). (a) Full virtual-source response (causal and acausal part) for DR164 (left-hand panel). The right panel shows the combination of time-asymmetric parts providing proper kinematics of the retrieved events. (b) The same analysis for CO165. (c) VSGs obtained using various combinations of correlated gathers shown in (a) and (b).

Figure S10. Illumination diagnosis, consisting of the results for the VSGs at all receiver positions analysed in this study. Colour lines indicate the range of velocities typical for body waves in the study area (green line) and air wave (blue line) indicated in Fig. 13(a). Green dots denote data points representing manually selected VSGs. Note the consistent illumination diagnosis results between receiver lines.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

APPENDIX A: CCF PRE-PROCESSING

We compare the CCF pre-processing techniques applied to the two 15-min-long AN noise segments representing noise from a single transient source (extracted from DR39 and CO165 segments, respectively). We run one-bit normalization and spectral whitening and test filtering for several frequency ranges based on most energetic segments in the PSD plots and associated with contributions from different field operations (see Fig. 5b for relevant frequency ranges). Spectral whitening allows for sufficient convergence of CCF even for the short segment of data (compare results in Figs A1e and f). We use this approach in conjunction with bandpass filter preserving most frequency bands highlighted in Fig. 5(b) to investigate the temporal changes of VSGs for every 15-min period at the single location of the array (Fig. 6).

APPENDIX B: CURVELET WEDGE PARAMETERS

To obtain curvelet-domain wedge parameters for a specific virtual source gather (VSG) position v, we first compute the average amplitude of the inverse curvelet transform C_i^H from data represented by a single angular wedge:

$$\tilde{W}_i = \frac{\int_{x1}^{x2} \arg\max C_i^H(x)\mathrm{d}x}{x_2 - x_1}, \tag{B1}$$

where $\tilde{W}_i$ denotes the average amplitude for the ith analysed window, x_1 and x_2 are the first and last traces of a single angular wedge in the time–space domain, respectively. Then we define the wedge ratio parameter as

$$W_R = \frac{\tilde{W}_T}{\frac{1}{n}\sum_{i\in S}^{n} \tilde{W}_i}, \tag{B2}$$

where $\tilde{W}_T$ is the average amplitude for the window containing the target coherent event and S denotes the collection of indices related to n adjacent VSGs, from which the windows are used for weighting the amplitudes of $\tilde{W}_i$. Wedge parameter 1 (W_{P1}) is then obtained from the inverse curvelet transform:

$$W_{P1} = W_R \arg\max(C_T^H), \tag{B3}$$

where C_T^H is the inverse curvelet transform of the data from the target window (Fig. B1e). Wedge parameter 2 (W_{P2}) is defined as:

$$W_{P2} = W_R \arg\max(C_T) p_{sep}, \tag{B4}$$

where C_T is the forward curvelet transform of the data from the target window represented by a single angular wedge and p_{sep} is the distance between peaks associated with the maximum value of C_T (Fig. B1d).

In Fig. B2(a) we show the VSG manually selected for receiver position 6023, and its curvelet parametrization using the inverse (Fig. B2b) and forward (Fig. B2c) curvelet transform in five spatio-temporal windows. The same analysis performed for the transition VSGs from Figs 13(b) and (c) is shown in Fig. B2(d) (see also Fig. S7 for comparison of multiple parameters derived from curvelet analysis). We use two wedge parameters to assess the reflectivity of VSGs in the following configuration: (i) obtained from all 246 AN segments for receiver position 23 on every dense recording line (Fig. S6, (ii) obtained from all 246 AN segments for master-trace positions located at receiver line 6 shown in Fig. 10 (Fig. S6b) and (iii) transition VSGs (denoted with pink dots in Fig. S6c). In Fig. S6, green dots denote manually selected VSGs, grey dots denote VSGs for all remaining AN segments for a master-trace position indicated in the top right part of the plots and blue dots denote VSGs for all AN segments for every master-trace position along receiver lines 1–7. VSGs manually selected for receiver line 6 in Section 4, exhibit distinctly higher values than for VSGs obtained using the remaining 246 activities, which only partly is confirmed by analysing receivers from the remaining receiver lines (see points with high scores of wedge parametrization for receiver lines 1–5, and 7 in Fig. S6a).

APPENDIX C: SEMI-AUTOMATIC WORKFLOW TO EVALUATE REFLECTIVITY CONTENT

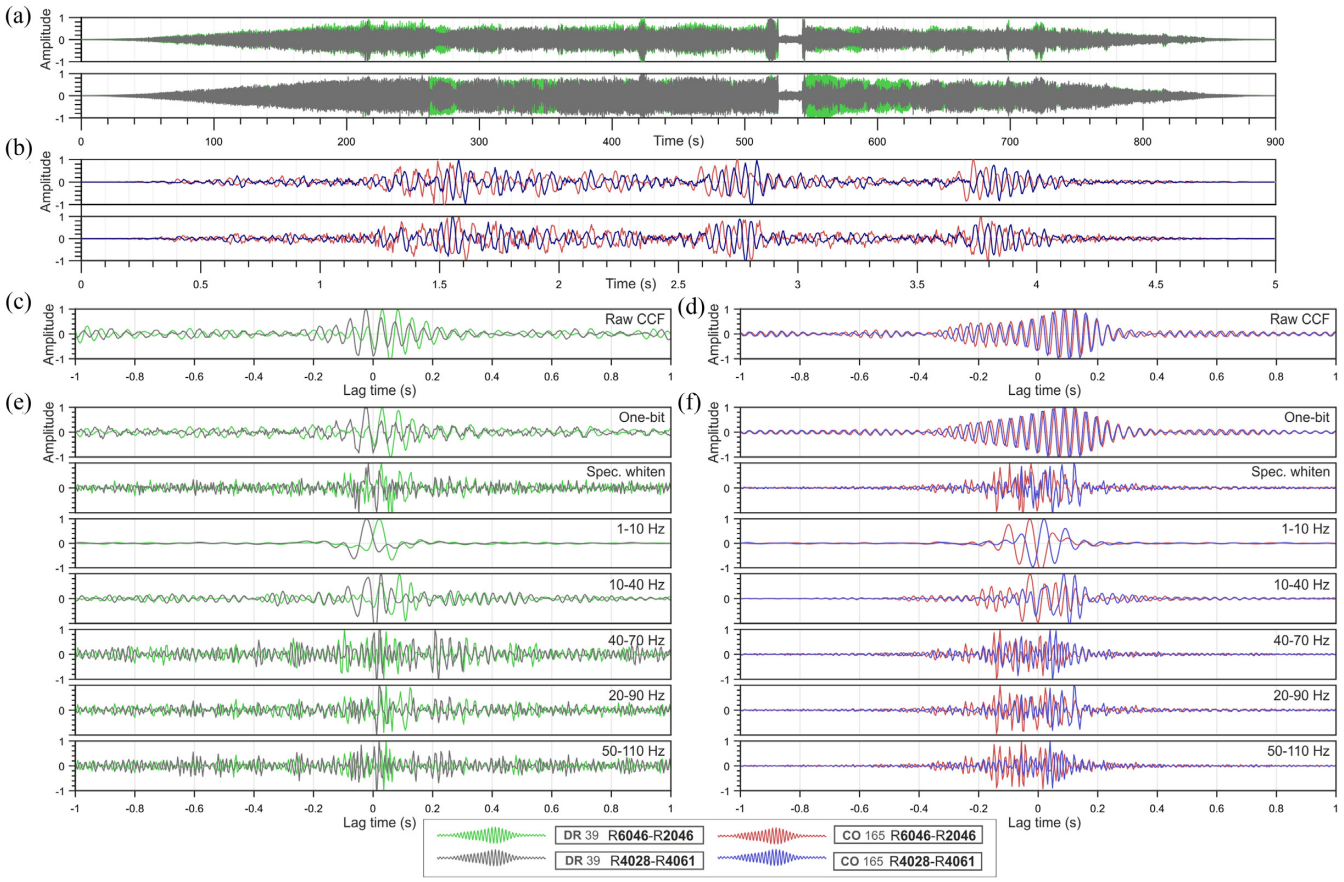

Figure A1. (a) Comparison of the cross-correlation functions (CCFs) obtained for two pairs of receivers in the crossline (denoted with green and red line) and inline (denoted with grey and blues lines) directions of the PilbArray using 10-s-long traces from CO165 activity [blue and red lines in panel (b)] and the full 15-min-long segment of DR39 activity [green and grey lines in panel (a)]. Panels (c) and (d) show CCF results without any pre-processing. Panels (e) and (f) show results of computing the CCF in using one-bit normalization (e), and spectral whitening (f), followed by several frequency bandpass filters.

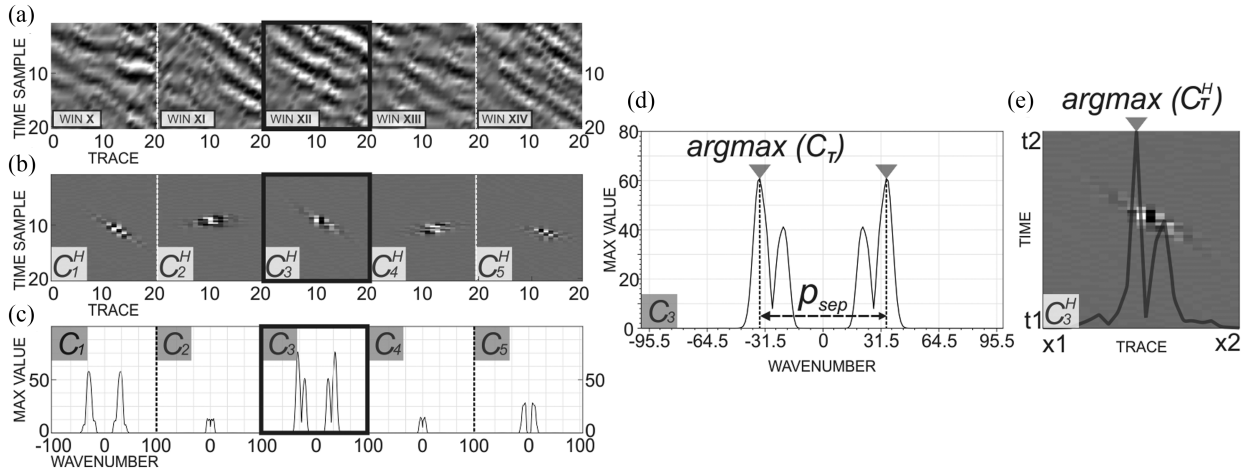

Figure B1. Curvelet parametrization of a single VSG. (a) Coherent events extracted from the selected windows. The roman numbers refer to the position of the VSG along receiver line 6. For each VSG, we choose a single window including a coherent event (referred to as the target window and marked by a bold line) and two windows from the left and right adjacent VSG positions. Note that for the configuration shown in this figure, the target window is WIN XII and the remaining windows are input to the denominator in eq. (B2). (b) Inverse curvelet transform of wedges associated with windows in (a). (c) Maximum values of the forward curvelet transforms narrowed to part of the data falling in the chosen window. (d) Sketch illustrating parameters used to obtain W_{P2}. (e) Sketch illustrating parameters used to obtain W_{P1}. For parameter definition, see the description of eqs (B3) and (B4).

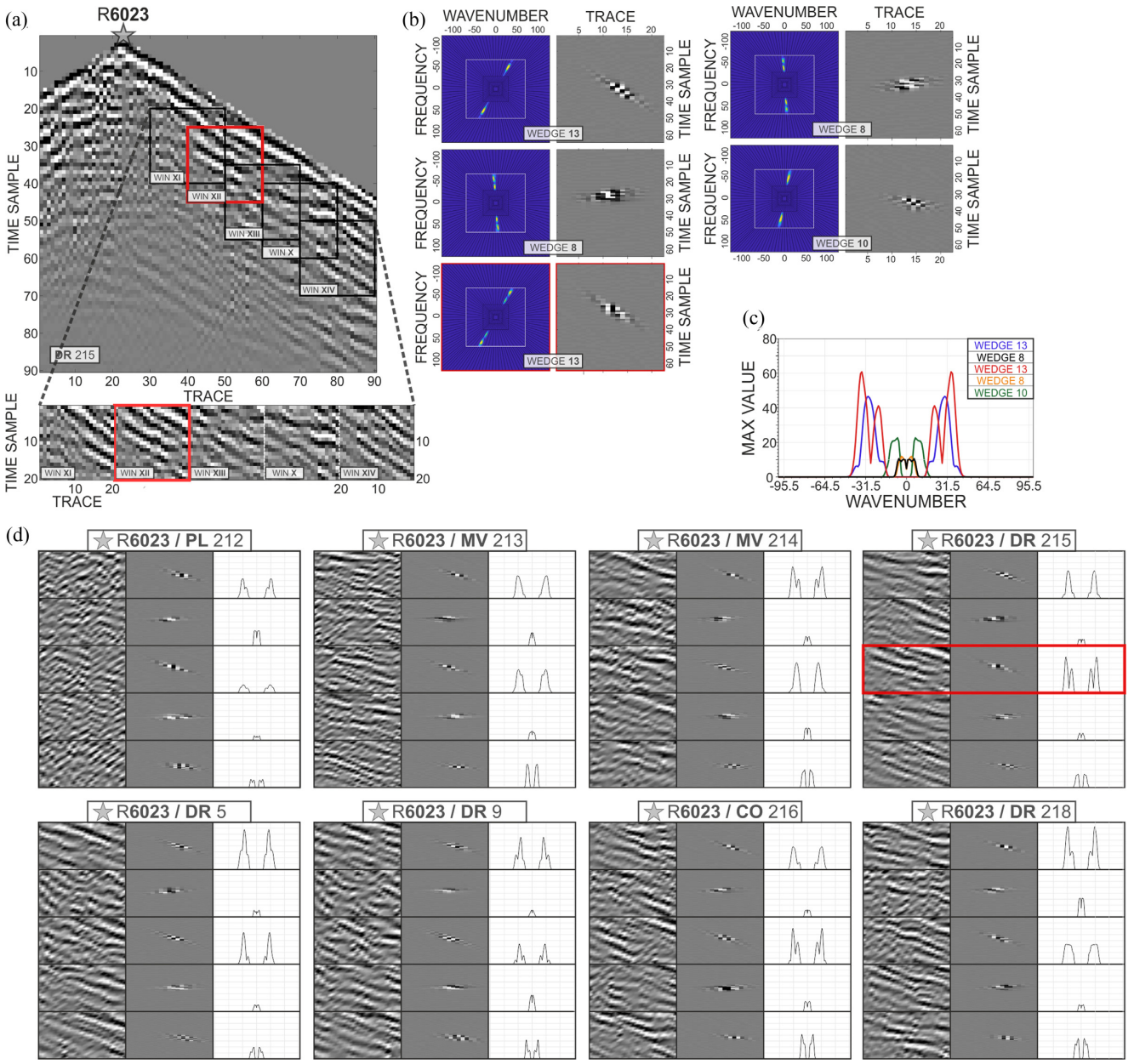

Figure B2. Curvelet representation of coherent events retrieved in VSGs. (a) Analysis of a VSG manually selected for receiver position 6023 (VSG 12) and used to illustrate domains, from which wedge parameters are derived. The numbering of spatio-temporal windows (roman numbers) refers to the master-trace position of the adjacent VSGs, from which the windows are used for deriving the wedge parameters (see Fig. B1 for details of window selection). The *F*–*k* and time-domain single-wedge representations shown in (b) are computed using data inside the windows marked by rectangles in (a). The numbering of wedges refers to their position in the 4th scale and starts from 1 in the top left edge indicated with a grey shaded polygon in Fig. 12b. Wedge parameters are obtained using two computational domains—*f*–*k* and time domain (left- and right-hand column in b, respectively). (c) Maximum values of the *f*–*k* transform shown in (b). Note that the highest amplitudes and peak separation are obtained for the coherent event falling in window XII and associated with the 13th wedge. (d) A similar analysis for the same receiver position using the eight VSGs shown in Figs 13(b) and (c). The five windows correspond to the five boxes for each VSG in Figs 13(b) and (d).

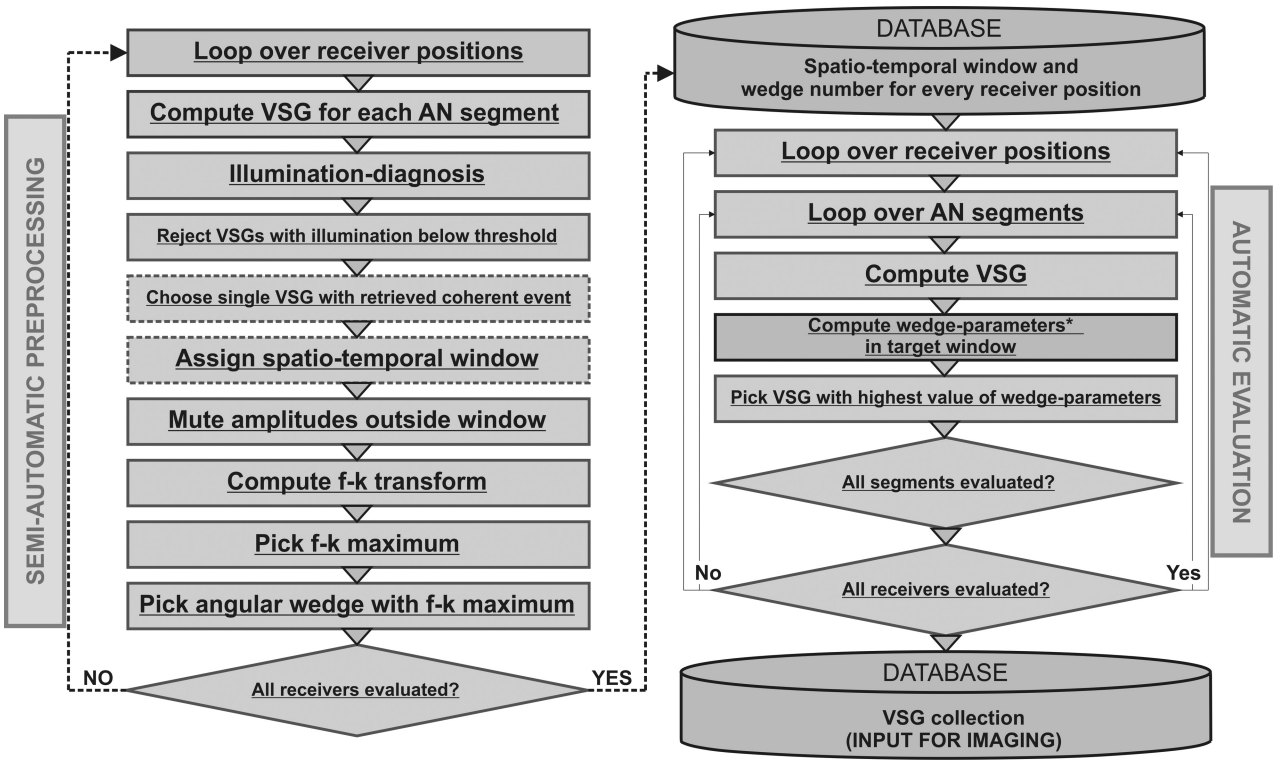

Figure C1. The workflow of semi-automatic curvelet-based evaluation of reflectivity present in VSGs.