

# Denoising of Distributed Acoustic Sensing Seismic Data Using an Integrated Framework

Electronic Seismologist

Yangkang Chen<sup>\*1</sup>, Alexandros Savvaidis<sup>1</sup> , Sergey Fomel<sup>1</sup> , Yunfeng Chen<sup>2</sup>,  
Omar M. Saad<sup>3</sup>, Hang Wang<sup>2</sup>, Yapo Abolé Serge Innocent Oboué<sup>2</sup> ,  
Liuqing Yang<sup>4</sup> , and Wei Chen<sup>5</sup> 

## Abstract

Distributed acoustic sensing (DAS) is an emerging technology that offers great potential in the high-resolution multi-scale seismic investigation due to its dense spatial coverage and cost-effectiveness. However, DAS data notoriously suffer from the low signal-to-noise ratio (SNR) due to various types of strong noise, for example, high-frequency noise, high-amplitude erratic noise, vertical or horizontal noise. Here, we propose a novel denoising framework by cascading several individual denoising methods that are designed for suppressing specific types of noise. First, to suppress the high-frequency noise, we apply a bandpass filter, which is implemented by recursive infinite impulse response filtering in the time domain. Second, to suppress the erratic noise, we apply a structure-oriented median filter that arises from the reflection seismology field. Finally, to suppress the vertical or horizontal noise, we apply a carefully designed dip filter in the frequency–wavenumber domain. The overall effect of these cascaded denoising steps is that the DAS data can be dramatically improved in terms of SNR. We introduce in detail the implementations of each step in the proposed denoising framework and analyze their respective contribution toward the final improvement. We demonstrate the effectiveness of the proposed denoising framework through the open-access Frontier Observatory for Research in Geothermal Energy (FORGE) geothermal DAS dataset and provide the reproducible processing workflows for all the DAS subsets containing the catalog earthquake and microseismic events.

Cite this article as Chen, Y., A. Savvaidis, S. Fomel, Y. Chen, O. M. Saad, H. Wang, Y. A. S. I. Oboué, L. Yang, and W. Chen (2022). Denoising of Distributed Acoustic Sensing Seismic Data Using an Integrated Framework, *Seismol. Res. Lett.* **XX**, 1–16, doi: 10.1785/0220220117.

## Introduction

Fiber-optic sensing technology has emerged over the past decade, and brings a new paradigm for recording seismic data, distributed acoustic sensing (DAS) using the fiber-optic cables (Johny *et al.*, 2021; Wamriew *et al.*, 2021). DAS is a revolutionary development in seismic acquisition because it allows unprecedentedly high spatial and temporal resolutions compared with conventional geophones, which facilitates high-resolution and high-fidelity subsurface imaging based on either active or passive sources (Wamriew *et al.*, 2021). In addition, the spatial sensing range is significantly improved thanks to the long deployment distance of fibers. From the engineering point of view, the DAS technique overcomes several limitations of traditional geophones, for example, troubles for working in hostile habitats, due to their small size, corrosion resistance, resistance to electromagnetic interference, and ability to work effectively in extreme conditions (e.g., high pressure and high temperature) (Johny *et al.*, 2021).

The DAS technique measures the strain over the fiber-optic cable, which is interrogated with a laser and timing instrument

(Spikes *et al.*, 2019). During the operation, the interrogator sends a laser pulse and records the backscattered light. The phase variation of the backscattered light is transformed into strain and averaged over a gauge length. DAS has been applied in seismic monitoring extensively using existing telecommunication fibers. Some noteworthy applications are earthquake recording (Lindsey *et al.*, 2017; Martin *et al.*, 2017; Lellouch *et al.*, 2019; Sladen *et al.*, 2019; Walter *et al.*, 2020; Lellouch

1. Bureau of Economic Geology, John A. and Katherine G. Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas, U.S.A.,  <https://orcid.org/0000-0001-6373-5256> (AS);  <https://orcid.org/0000-0002-9024-5137> (SF);
2. School of Earth Sciences, Zhejiang University, Hangzhou, Zhejiang, China,  <https://orcid.org/0000-0002-9956-7669> (YASIO); 3. ENSN Lab, Seismology Department, National Research Institute of Astronomy and Geophysics (NRIG), Helwan, Egypt; 4. State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing, China,  <https://orcid.org/0000-0002-4064-501X> (LY); 5. Key Laboratory of Exploration Technology for Oil and Gas Resources of Ministry of Education, Yangtze University, Wuhan, China,  <https://orcid.org/0000-0003-4222-964X> (WC)

\*Corresponding author: chenyk2016@gmail.com

© Seismological Society of America

*et al.*, 2021), oil and gas reservoir monitoring (Mateeva *et al.*, 2013), carbon dioxide storage characterization (Harris *et al.*, 2016), urban seismic site characterization (Spica *et al.*, 2020), near-surface characterization (Dou *et al.*, 2017; Ajo-Franklin *et al.*, 2019), illuminating seafloor faults (Lindsey *et al.*, 2019; Cheng *et al.*, 2021), traffic monitoring (Lindsey *et al.*, 2020; Yuan *et al.*, 2020), and so on. DAS measurements have been made in surface (Spikes *et al.*, 2019), downhole (Lellouch *et al.*, 2021), or marine seafloor environments (Lindsey *et al.*, 2019). The downhole deployment of DAS usually is of significantly better recording quality than the surface deployment because it is less affected by the near-surface complexity and viscoelastic attenuation.

Despite the low cost and other aforementioned advantages of the DAS technique, the signal quality of an individual channel can be much lower than that of a conventional geophone. Hull *et al.* (2017) conduct a case study on comparing the signal quality, in terms of signal detectability, between the DAS and conventional geophone recordings. They find that 31 events are detected using DAS when monitoring the hydraulic stimulation, whereas 781 events are detected using conventional geophones. More quantitatively, Webster *et al.* (2016) find that the event detectability using DAS is only 10% of that using conventional geophones. Li *et al.* (2021) apply template matching to DAS data and detect six times more aftershocks of the 2019 Ridgecrest earthquake than conventional detection. However, strong traffic noise still affects the event detectability based on DAS data especially during daytime (less than half of the number of detected events at night). These aforementioned issues are caused by relatively low signal-to-noise ratio (SNR) due to a variety of factors including instrumental, coupling, cultural, and environmental effects. Noise in DAS can be categorized into non-optical and optical noise. Nonoptical noise includes cultural and environmental noise. Cultural noise usually arises from footsteps, vehicles, airplanes, ships, construction, and electrical 60 Hz power lines. Environment noise is usually caused by winds, ocean waves, earth tides, and so on. Common mode noise, also known as horizontal noise, is a type of optical noise. It is caused by sonic waves that hit the interferometer and deform the two input optical I and Q signals. The I signal denotes the light stream directly coming out from the interferometer, and the Q signal denotes its Hilbert transform. The common mode noise is horizontal in a vertical seismic profile (VSP) setup (with a vertical time axis and a lateral channel axis). The fading noise is another type of optical noise, which commonly arises from the interaction between laser light and the intrinsic property of the optical glass. The fading noise usually has a strong amplitude and thus appears as the erratic noise in most DAS recordings. Coupling noise is a unique type of noise in VSP DAS recordings. It is caused by the imperfect attaching of fiber-optic cable on the borehole wall. It appears as the zigzag noise in many VSP datasets (Willis, 2022).

Denoising is a common way to attenuate noise in DAS data so as to improve the SNR (Huang *et al.*, 2018; Chen *et al.*, 2019;

Wamriew *et al.*, 2021). However, most researchers apply relatively simple signal analysis approaches to remove the noise. Lellouch *et al.* (2021) and Wamriew *et al.* (2021) both use a low-pass filter and a 2D median filter (MF) to remove the strong noise in the DAS datasets. For those very noisy traces, Lellouch *et al.* (2021) simply remove them to avoid affecting the subsequent analysis of the waveforms. More advanced methods come from in reflection seismology, considering the similarities between the DAS datasets and reflection seismic datasets (Dong and Li, 2020), for example, the wavelet method (Qin *et al.*, 2012), curvelet method (Qin *et al.*, 2017; Atterholt *et al.*, 2022), principal component analysis (PCA) (Ibrahim *et al.*, 2020). The wavelet method decomposes a signal into a group of dilated and translated mother wavelet functions so as to separate more coherent signal from more random noise (Qin *et al.*, 2012). This method highly depends on the choice of the mother wavelet, which is mostly chosen as a standard wavelet function derived from the general signal processing field. The seismic signal, however, is more complex and is difficult to be simply modeled as a combination of fixed-basis mother wavelets. The curvelet method considers the spatial coherency of the seismic wavefield and can perform much better than the wavelet method (Atterholt *et al.*, 2022). However, the curvelet method is still a fixed-basis sparse transform (Chen *et al.*, 2016), meaning that the basis functions used to sparsify seismic data are fixed and tend to fail when structures become challenging. The PCA method relies on the repetitive patterns hidden in the DAS data which can be represented by a principal component using methods like singular value decomposition. The DAS data, however, are highly nonstationary along the spatial dimension and could suffer from serious signal leakage using a PCA method.

Considering the special features of the noise in DAS gathers, we propose an integrated framework for denoising various types of noise in DAS data. Here, we consider the three most common types of noise, that is, high-frequency noise, high-amplitude erratic noise, vertical and horizontal (morphologically) noise. Similar to a common earthquake processing workflow, we apply a bandpass (BP) filter to remove the high-frequency noise. Different from a typical 2D MF, we propose a structure-oriented median filter (SOMF) that aims to apply a 1D MF following the spatial waveform pattern and thus preserves the signal energy. To remove the vertical/horizontal noise, we design a dip filter in the frequency-wavenumber (*f-k*) domain, with easy control of the filter strength. The application of the proposed cascaded filtering schemes on challenging DAS datasets demonstrates a significant improvement of data quality in terms of SNR, visual inspection, and event detectability.

## Method

### BP filter

The simplest BP filtering method is to apply a box function in the frequency domain. However, a box function in the frequency domain corresponds to an oscillating and slowly

decaying sinc function in the time domain. To avoid this problem, we apply a Butterworth filter (Butterworth, 1930; Proakis and Manolakis, 2006; Oppenheim and Schafer, 2009). The Butterworth filtering applies a frequency-domain function that behaves like a smoothed box. The frequency-domain function is expressed as

$$B_N^2(w) = \frac{1}{1 + (\frac{w}{w_0})^{2N}}. \quad (1)$$

The function  $B_N^2(w)$  is flat at the zero frequency, in which it takes the value of 1, and takes the value of 0 at infinity. The transition behavior is controlled by the number  $N$ .

The function  $B_N^2(w)$  has  $2N$  poles in the complex plane that are distributed around the circle with the radius  $w_0$  and come in conjugate pairs. Therefore, we can write

$$B_N^2(w) = \prod_{k=0}^{N-1} \frac{1}{\left(\frac{w}{w_k} - 1\right)\left(\frac{w}{w_k} - 1\right)} = \prod_{k=0}^{N-1} \frac{1}{\left(\frac{w^2}{w_0^2} - 2 \cos \theta_k \frac{w}{w_0} + 1\right)}, \quad (2)$$

in which  $w_k = w_0 e^{i\pi(2k+1)/(2N)}$  is the  $k$ th complex root of the equation  $w^{2N} = -w_0^{2N}$ , and  $\theta_k = \pi(2k+1)/(2N)$  is its phase angle in the complex plane.

A suitable digital approximation of the derivative is (when  $Z = e^{iw\Delta t}$ ):

$$iw \approx \frac{2}{\Delta t} \frac{1 - 1/Z}{1 + 1/Z} = i \frac{2}{\Delta t} \tan\left(\frac{w\Delta t}{2}\right). \quad (3)$$

Replacing  $w$  and  $w_0$  in equation (2) with their approximations from equation (3), the Butterworth function will map into the periodic domain and turn into the digital filter:

$$B_N^2(Z) = \prod_{k=0}^{N-1} \frac{1}{\left[\frac{1}{\tan(w_k \Delta t / 2)} \left(\frac{1-1/Z}{1+1/Z}\right) - 1\right] \left[\frac{1}{\tan(w_k \Delta t / 2)} \left(\frac{1-Z}{1+Z}\right) - 1\right]}. \quad (4)$$

$B_N^2(Z)$  is a cascade of recursive filters. It is suitable for implementation in the time domain.

## SOMF

Unlike the traditional MF that is applied directly along the horizontal direction, the SOMF takes the structural direction (as indicated in Fig. 1a) into consideration when applying an MF. Applying the 1D MF along the structural direction enables protecting more useful signals than the horizontal direction.

To facilitate the structure-oriented filtering, a critical step is to locally flatten the seismic gather, for example, the left panel shown in Figure 1a. We flatten the 2D seismic gather to a 3D cube like the right panel in Figure 1a via a recursive trace prediction strategy based on the pre-calculated local slope, as defined in Figure 1a. The general principles of the prediction

and the flattening process are plotted in Figure 1b. In each seismic gather, each trace (e.g., indicated by the red line in Fig. 1b) is predicted by shifting its left (green area in Fig. 1b) and right neighboring traces (orange area in Fig. 1b) in a recursive way. For example, the left prediction from  $i + j$ th trace to  $i$ th trace can be expressed as

$$\mathbf{d}_i = \mathbf{P}_{i+1,i} \cdots \mathbf{P}_{i+j-1,i+j-2} \mathbf{P}_{i+j,i+j-1} \mathbf{d}_{i+j}, \quad (5)$$

in which  $\mathbf{d}_i$  denotes  $i$ th seismic trace, and  $\mathbf{P}_{i+1,i}$  denotes the prediction operator from  $i + 1$ th trace to  $i$ th trace. The prediction operator  $\mathbf{P}$  is based on the plane-wave assumption:

$$\frac{\partial u}{\partial x} + \sigma \frac{\partial u}{\partial t} = 0, \quad (6)$$

in which  $\sigma$  is the slope of the local plane waves. Equation (6) is also known as the plane-wave destruction (PWD) equation.  $\sigma$  is a key parameter in the SOMF, which has been covered by many references in the literature (Fomel, 2002; Wang et al., 2022). Based on the PWD equation, neighbor traces are predictable, and the operator  $\mathbf{P}$  symbolize the prediction process.

Similarly, the right prediction from  $i - j$ th trace to  $i$ th trace can be expressed as

$$\mathbf{d}_i = \mathbf{P}_{i-1,i} \cdots \mathbf{P}_{i-j+1,i-j+2} \mathbf{P}_{i-j,i-j+1} \mathbf{d}_{i-j}. \quad (7)$$

In this way, all seismic traces obtain their corresponding flattened gathers, see right part of Figure 1b, in which a 1D MF is applied. This MF is used to remove erratic noise without harming the signal energy. The prediction distance  $N$  is equal to the half-width of the filter in SOMF.

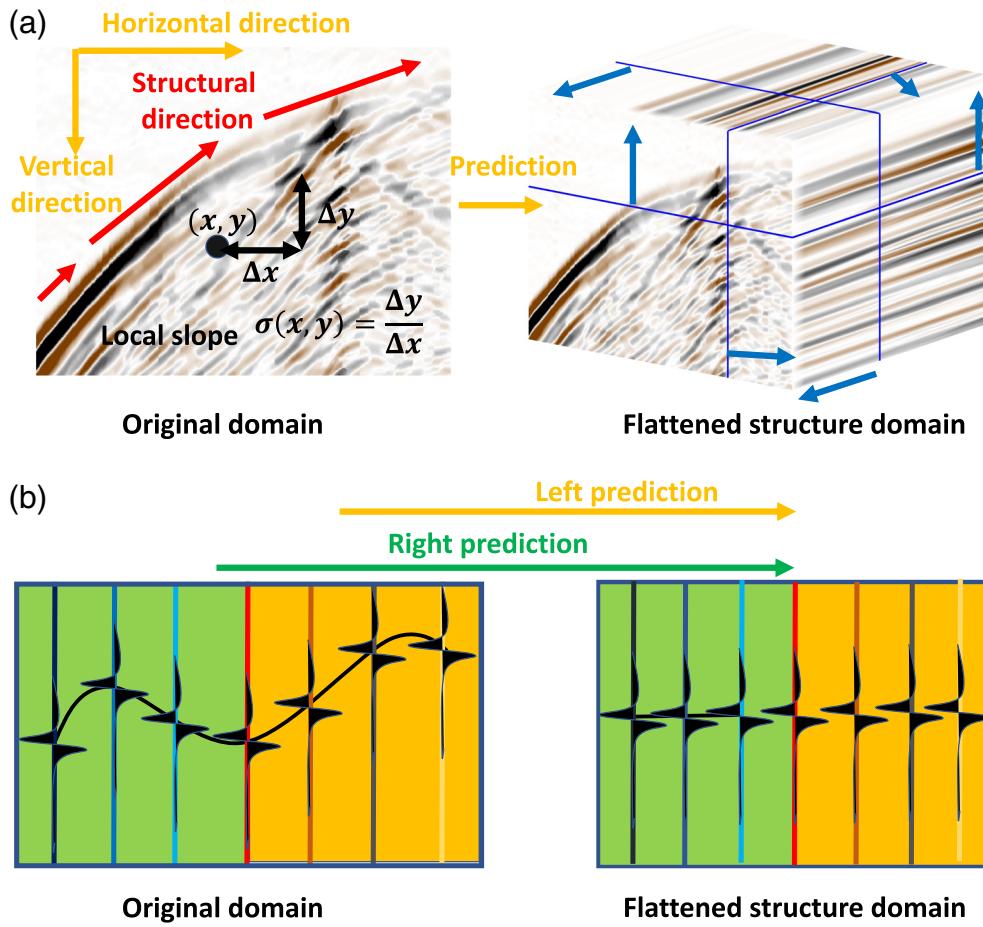
## Dip filter

The dip filter in the  $f$ - $k$  domain aims to remove or preserve some components that are below or above a certain dip angle threshold. The dip filter  $\mathcal{D}$  can be expressed as

$$\mathcal{D} = \mathcal{F}^{-1} \mathcal{M} \mathcal{F}, \quad (8)$$

in which  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  denote the forward and inverse 2D Fourier transform, respectively.  $\mathcal{M}$  denotes a mask operator, with ones in locations where the  $f$ - $k$  spectra are preserved, and zeros in locations where the  $f$ - $k$  spectra are muted. The mask operator in a dip filter usually has a triangular shape.

The conventional dip filters usually require an estimate of the threshold dip angle (or slope) one wants to apply, which is sometimes inconvenient to choose because this threshold is usually data-dependent and requires some calculations. Considering that, a user-friendly method should be more automatic and need less human interference. So, a smaller number of parameters are more convenient for large-scale data processing. In the proposed dip filter, we minimize the number of



**Figure 1.** Demonstration of the principle of structure-oriented median filter (SOMF). (a) Definition of the local slope in a distributed acoustic sensing (DAS) seismic data and an example of the flattened 3D cube. (b) Flattening along the structural direction. The median filter (MF) is applied long the flattened structure domain. The blue lines indicate which slice in each dimension is plotted as a side of the cube. The front side of the cube is the data domain ( $x-y$ ), the same as the left figure in panel (a). The right side shows the flattened domain. The top side shows the horizontal slice of this 3D cube. The color version of this figure is available only in the electronic edition.

input parameters to one, that is, the half-width of the base of the triangle mask filter in percentage  $w$ .  $w$  controls the strength of the filter in terms of removing the horizontal noise.

In Figure 2, we illustrate the meaning of the input parameter  $w$  and the principle of the dip filter. Here, we consider two cases: horizontal noise and vertical noise. Figure 2a shows a synthetic surface-recorded DAS data with horizontal noise. Figure 2b shows its  $f-k$  spectrum. The spectrum corresponding to the horizontal noise is the narrow white band around the zero wavenumber in the  $f-k$  spectrum. Figure 2b clearly illustrates the range of  $w$ . According to the extent of the horizontal noise in the  $f-k$  spectrum,  $w$  varies from minimum 0 (no filtering effect) to the maximum of 0.5 (the strongest filtering effect). Figure 2c shows the denoised result using  $w = 0.2$ . It is clear that most horizontal noise is removed. Figure 2d shows the spectrum corresponding to the denoised data, in which the

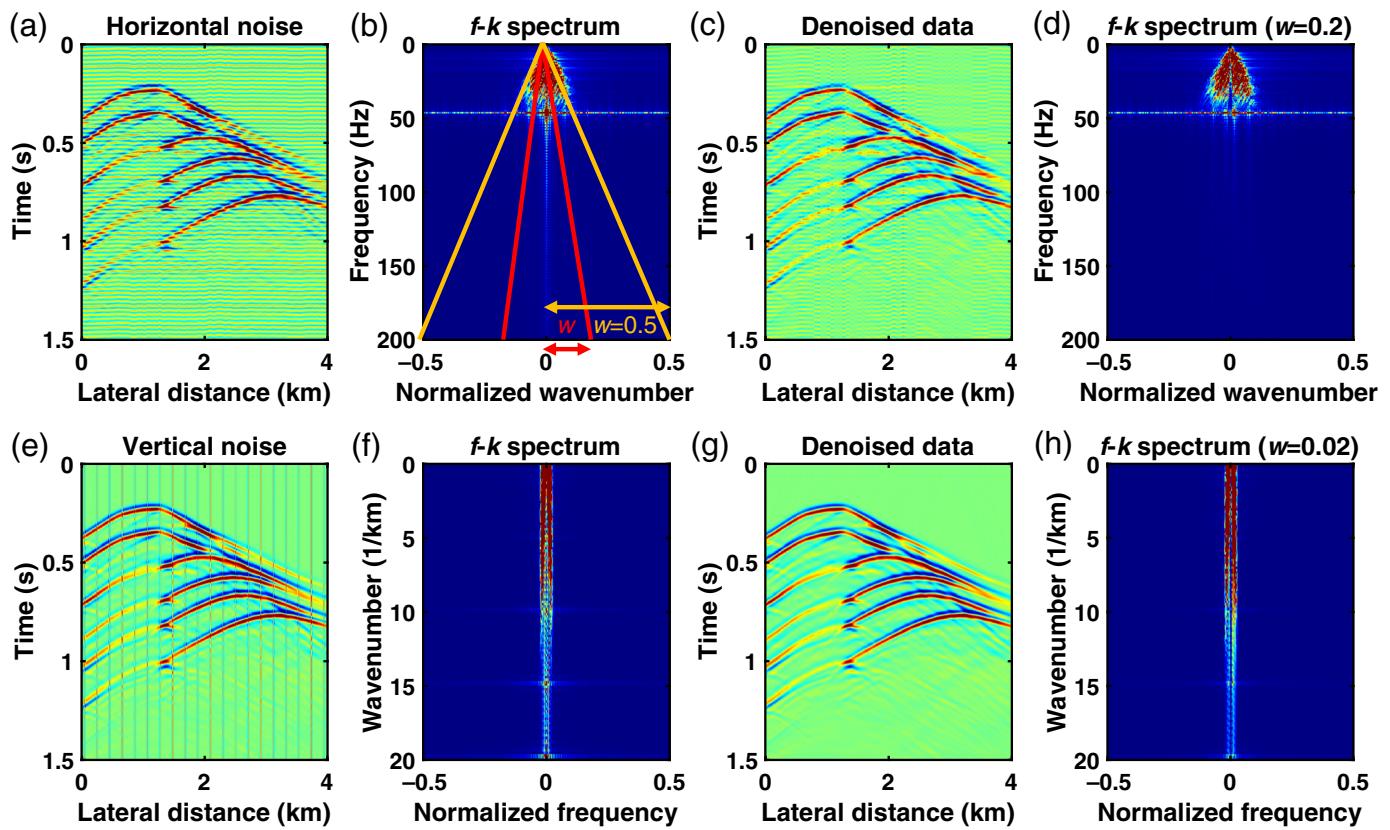
narrow white noise band is muted. When there is vertical noise in the data (e.g., in Fig. 2e), the same dip filter still works. One needs to first transpose the 2D matrix shown in Figure 2e to make the vertical noise horizontal as in the former case, then apply the same dip filter onto the transposed matrix, finally transpose the filtered matrix back to output the denoised data. The  $f-k$  spectrum of the transposed 2D matrix is plotted in Figure 2f, in which a narrow white band around the zero wavenumber is clearly indicating the existence of horizontal noise. The denoised data of this case (with  $w = 0.02$ ) are plotted in Figure 2g, with the  $f-k$  spectrum of its transpose shown in Figure 2h, in which the noise spectrum has been muted.

Considering the seismic data is nonstationary, meaning that the dip (velocity) content varies in the spatial domain, it is inconvenient to define a velocity threshold by measuring  $dt$  and  $dx$  that perfectly separates the dominant signals and unwanted noise. In fact, considering that the signal dip (velocity) ranges from 0 to a certain level, there is

always signal energy in the removed noise. The use of the  $w$  parameter, however, bypasses the selection of the perfect dip (velocity) threshold, and the default  $w$  is normally chosen as a relatively small but effective value, for example, 0.02. Despite the effectiveness of the  $w$ -parameter-based implementation, there are inevitable signals in the removed horizontal noise component. According to the examples shown in this article, the lost signal energy is acceptable even if a constant  $w$  parameter is used for all datasets. Simply speaking, the  $w$  parameter defines maximum and minimum velocities above and below which one easily find it acceptable to remove signal in the interest of getting rid of vertical and horizontal noise.

### Parameterization

There are three components in the proposed denoising framework, that is, BP, SOMF, and  $f-k$  filters. The parameters for



each component are straightforward in terms of physical interpretation and thus convenient to tune. The main parameters of the BP filter are the cutoff frequencies (low-cut or high-cut), which are familiar to almost all seismologists. In this article, we only consider the high-cut frequency and choose 200 Hz for all datasets. The other parameter, that is, the number of poles for the high-cut and low-cut filterings is chosen in default as 6. When the number of poles exceed 6, the performance does not change much. There are two subcomponents in the SOMF, that is, local slope estimation and median filtering in the flattened dimension. There are four main parameters in calculating the local slope, that is, smoothing radii in both vertical and lateral directions, number of conjugate-gradient iterations, and number of nonlinear Gauss–Newton iterations. The smoothing radii are chosen according to the noise levels of the input data. They are normally chosen in the range of 20–40 samples. The number of linear iterations is normally chosen as a large value (e.g., 50 or 100), but actually it does not take so many iterations because the convergence usually occurs within 20 iterations. The number of nonlinear iterations is chosen as 5 in default. Because the proposed framework is relatively stable and does not vary greatly in terms of denoising performance when a different parameter is chosen, we can conveniently set most parameters in default. There is only one parameter for flattening and median filtering the data according to the local slope: filter length. The filter length is chosen according to the strength of the erratic noise. The stronger the erratic noise, the longer the filter. Here, the

**Figure 2.** Demonstration of the principle of the dip filter in the frequency-wavenumber ( $f$ - $k$ ) domain. (a) Data with horizontal noise. (b)  $f$ - $k$  spectrum of panel (a). (c) Denoised data using  $w = 0.2$ . (d)  $f$ - $k$  spectrum of panel (c). (e) Data with vertical noise. (f)  $f$ - $k$  spectrum of panel (e). (g) Denoised data using  $w = 0.02$ . (h)  $f$ - $k$  spectrum of panel (g). The color version of this figure is available only in the electronic edition.

strength of erratic noise mainly refers to the spatial density in the gather, for example, more frequently existing erratic noise is considered as stronger. This parameter needs some tuning in practice, but the overall process is convenient based on a trial-and-error strategy. The  $f$ - $k$  filter, as introduced previously, is designed to be convenient to tune. Only a strength indicator  $w$  needs to be tuned and its effect in the  $f$ - $k$  domain is easy to understand.

### Estimation of local slope

The success of the SOMF highly depends on the accuracy and reliability of local slope estimation. The slope estimation method used in the proposed denoising framework is based on the PWD method (Fomel, 2002). As documented in many published papers (Zu *et al.*, 2017; Chen *et al.*, 2022), the accuracy of the local slope is mainly controlled by the noise level and the structural complexity of the input seismic data. The strong noise (especially the high-frequency random noise) will greatly affect the stability when calculating the local slope during the nonlinear inversion. Thus, some simple preconditioning steps to

mitigate the influence of random noise are helpful to the slope estimation. In this regard, the BP filtering step behaves as a pre-conditioning step for more stable slope estimation. On the other hand, when the structure of the seismic data becomes more complicated, the slope estimation could fail due to either the existence of multiple slopes (crossing events) or highly nonstationary features of seismic data (e.g., geological structures with different scales). In most cases, for example, all DAS datasets of the Frontier Observatory for Research in Geothermal Energy (FORGE) project, we are able to obtain robust slope estimations. In extreme cases, more advanced slope estimation strategies need to be applied (Wang *et al.*, 2021; Chen *et al.*, 2022).

## Dataset

The DAS dataset used to test the proposed integrated denoising framework is from the phase 2C hydraulic fracture stimulation of the FORGE site in Utah. The DAS data are acquired from the fiber-optic cable installed in the monitoring well 78-32 at the FORGE geothermal site. The Silixa iDAS v3 interrogator was used to measure the strain rate as the DAS signal. The fiber-optic cable continuously recorded the data for 10.5 days during the initial stimulation (in the testing well 58-32) of an enhanced geothermal system at the FORGE site from late April to early May 2019. All the DAS datasets can be freely downloaded from the Center for High Performance Computing at The University of Utah based on a shell script provided in the geothermal data repository. In this dataset, the gauge length is 10 m and channel spacing is 1 m.

The whole continuous data are split into many segments stored as SEGY files, with each lasting 15 s, with a 0.0005 s sampling interval and 30,000 samples for each file. Lellouch *et al.* (2021) detected 82 earthquakes from the DAS dataset. We use the catalog provided by Lellouch *et al.* (2021) to selectively download and segment the subsets of the whole FORGE DAS dataset to test our denoising framework. First, we download all the SEGY files that are detected to contain earthquakes, and then we segment each SEGY file based on the *P*-arrival sample number in the catalog from Lellouch *et al.* (2021). The beginning sample of each segment is 50 samples before the detected *P*-arrival sample number. The segment of each event has 2000 samples. We remove the first several channels and the last several channels that are obviously unusable and preserve 960 channels for each segment. In this way, we obtain 194 2D DAS seismic gathers (82 earthquake and 112 microseismic events), with each having a size of  $2000 \times 960$ . We apply the proposed denoising framework to all these testing DAS datasets, and demonstrate some representative examples. All denoising results are available from the online repository.

## Results

We analyze the result from one earthquake event in detail and briefly demonstrate the performance of several other representative datasets.

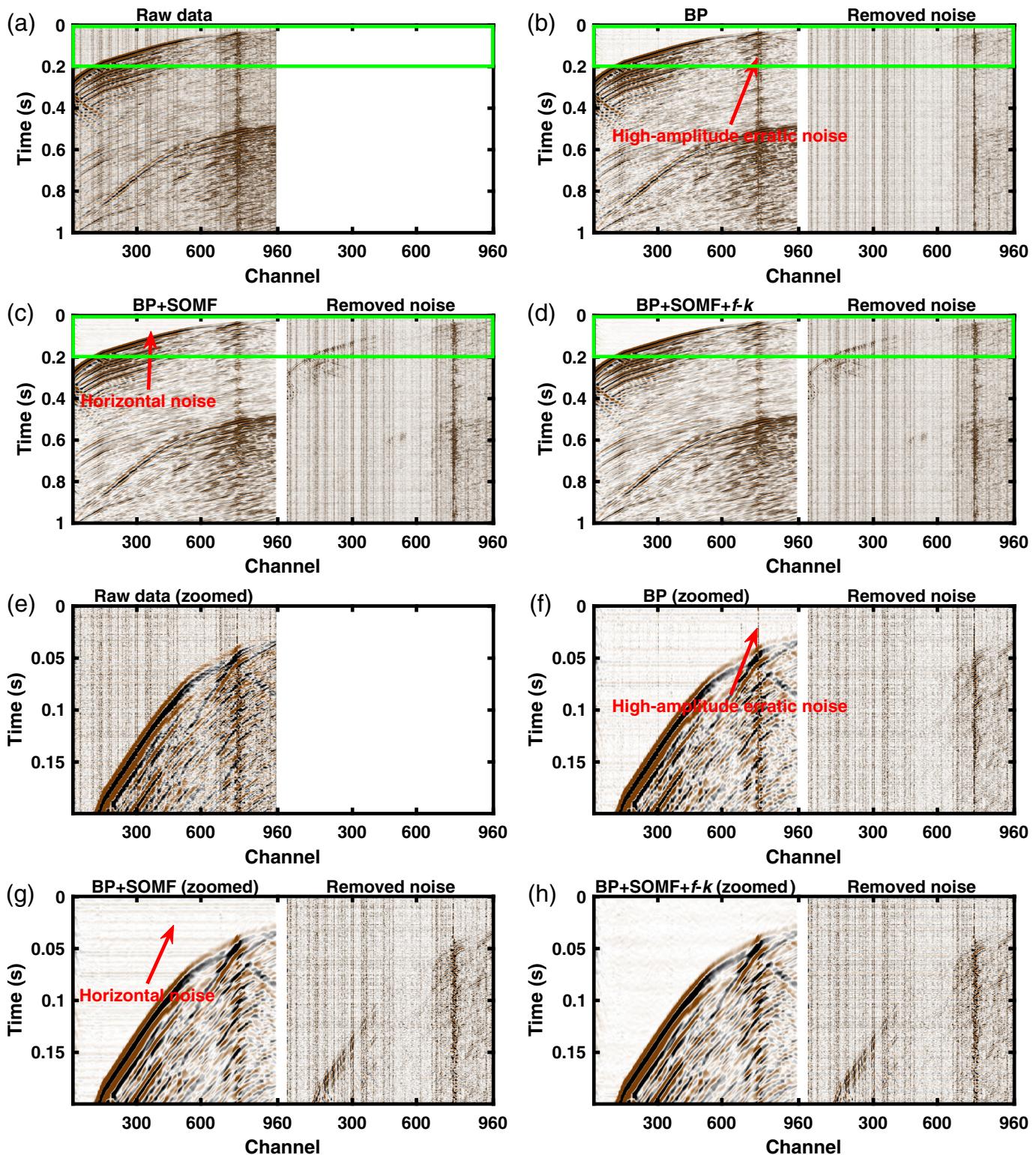
Figure 3 demonstrates the denoising processes for the event included in the SEGY file named “FORGE\_78-32\_iDASv3-P11\_UTC190423213209.sgy.” Figure 3a shows the raw DAS data. We show the results using the proposed denoising framework in different stages in Figure 3b–3d, respectively. Compared with the raw data, the BP filter (Fig. 3b) removes most high-frequency random noise while leaving the high-amplitude erratic noise contaminating many traces unchanged. After further applying the SOMF (Fig. 3c), the erratic noise is completely removed. The remaining horizontal noise is further removed by applying the dip filter in the *f-k* domain (hereafter referred to as the *f-k* filter). To visualize different types of noise more clearly and better compare the results, we show zoomed sections (indicated by the green frame boxes) from Figure 3a–3d in Figure 3e–3h, respectively. To analyze the denoising performance for single traces, we show a comparison of denoising performance in different stages for nine channels of the same dataset in Figure 4. In each panel of Figure 4, we can observe clearly that the data quality increases consistently from top to bottom, indicating the stronger denoising capability as more filters are involved in the denoising. In each panel, the bottom trace becomes very beneficial to event detection and *P*-wave arrival picking. The SNR improves from -4.72 dB in the raw data to 0.03 dB in the denoised data. We use the definition in Chen and Fomel (2015) to calculate the SNR:

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{s}\|_2^2}{\|\mathbf{n}\|_2^2}, \quad (9)$$

in which  $\mathbf{s}$  and  $\mathbf{n}$  denote signal and noise, respectively. The signal is chosen as the time window of 80 samples following the *P*-wave arrival after denoising. The noise is chosen as the time window of the same length before the *P*-wave arrival.

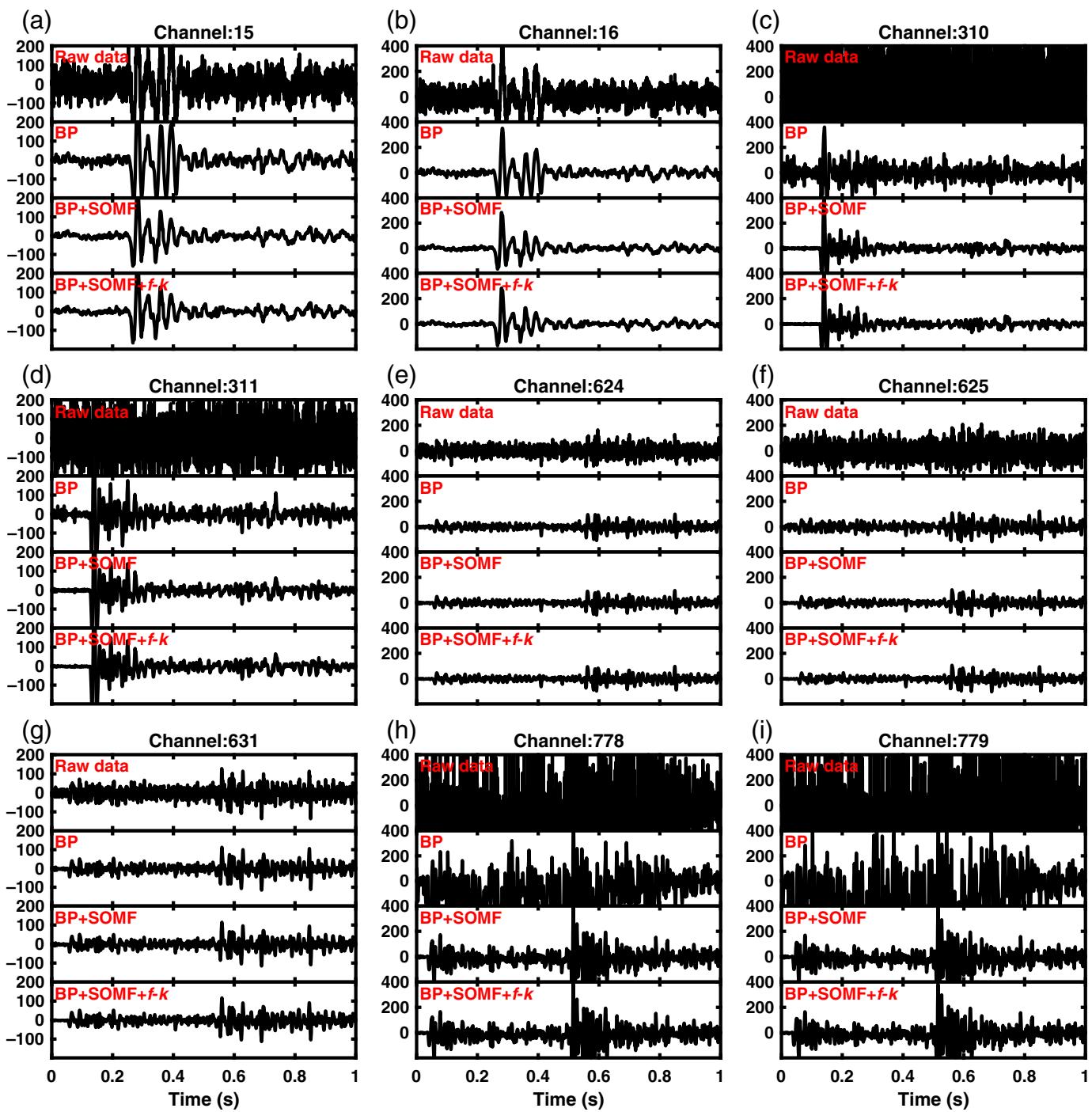
Figure 5 demonstrates more denoising results for different earthquake events. The SEGY file names including these earthquakes are marked on the bottom of each panel. All results demonstrate that the proposed denoising framework can help recover the earthquake signals effectively from very noisy DAS recordings with barely visible signals. The three rows in Figure 5 represent three different cases of recovered signal qualities, that is, top for strong recovered signals, middle for moderately strong recovered signals, and bottom for weak recovered signals.

One of the main differences between the proposed denoising framework and some published methods in the literature (Lellouch *et al.*, 2021; Wamrrew *et al.*, 2021) is the inclusion of the SOMF instead of the traditional MF. As illustrated in Figure 1, the traditional MF is simply applied along the horizontal direction to remove the high-amplitude erratic noise, which tends to damage many dipping signals. The SOMF, however, can better protect the signal energy by considering the spatial morphological trend of the signal waveforms. To compare the difference between MF and SOMF, we conduct denoising



**Figure 3.** Denoising results of the event recorded in FORGE\_78-32\_iDASv3-P11\_UTC190423213209.sgy with  $P$ -arrival sample number 1484. (a) Raw data. (b-d) Denoised results using the proposed framework in different stages. (e-h) Zoomed comparison from panels (a-d) (indicated the green frame boxes),

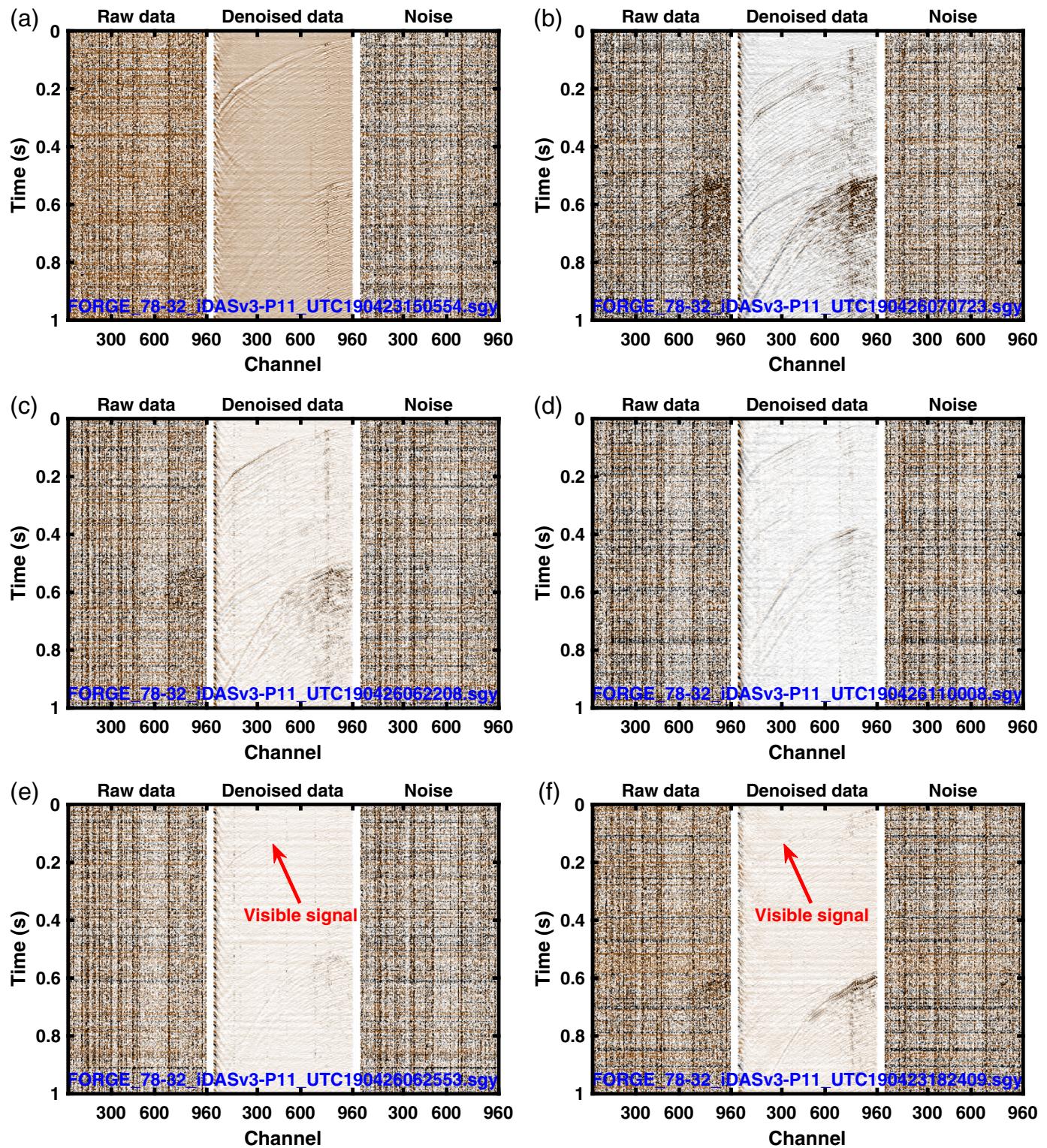
respectively. The apparently coherent energy in noise is referred to as the signal leakage (Chen and Fomel, 2015), which is inevitable in the denoising process. The color version of this figure is available only in the electronic edition.



tests by replacing the SOMF with the MF in the whole workflow. The filter lengths of MF and SOMF are both 17 traces. The results of two events (included in FORGE\_78-32\_iDASv3-P11\_UTC190423213209.sgy and FORGE\_78-32\_iDASv3-P11\_UTC190426070723.sgy) are plotted in Figure 6. In Figure 6a, we observe that for steeply dipping signals, the denoising combination of BP, MF, and  $f\text{-}k$  filters results in significant damages to the earthquake signals. Denoising by SOMF, however, much better preserves the signal energy and causes significantly less signal leakage. It is worth observing that the signal leakage issue can be further addressed by the local

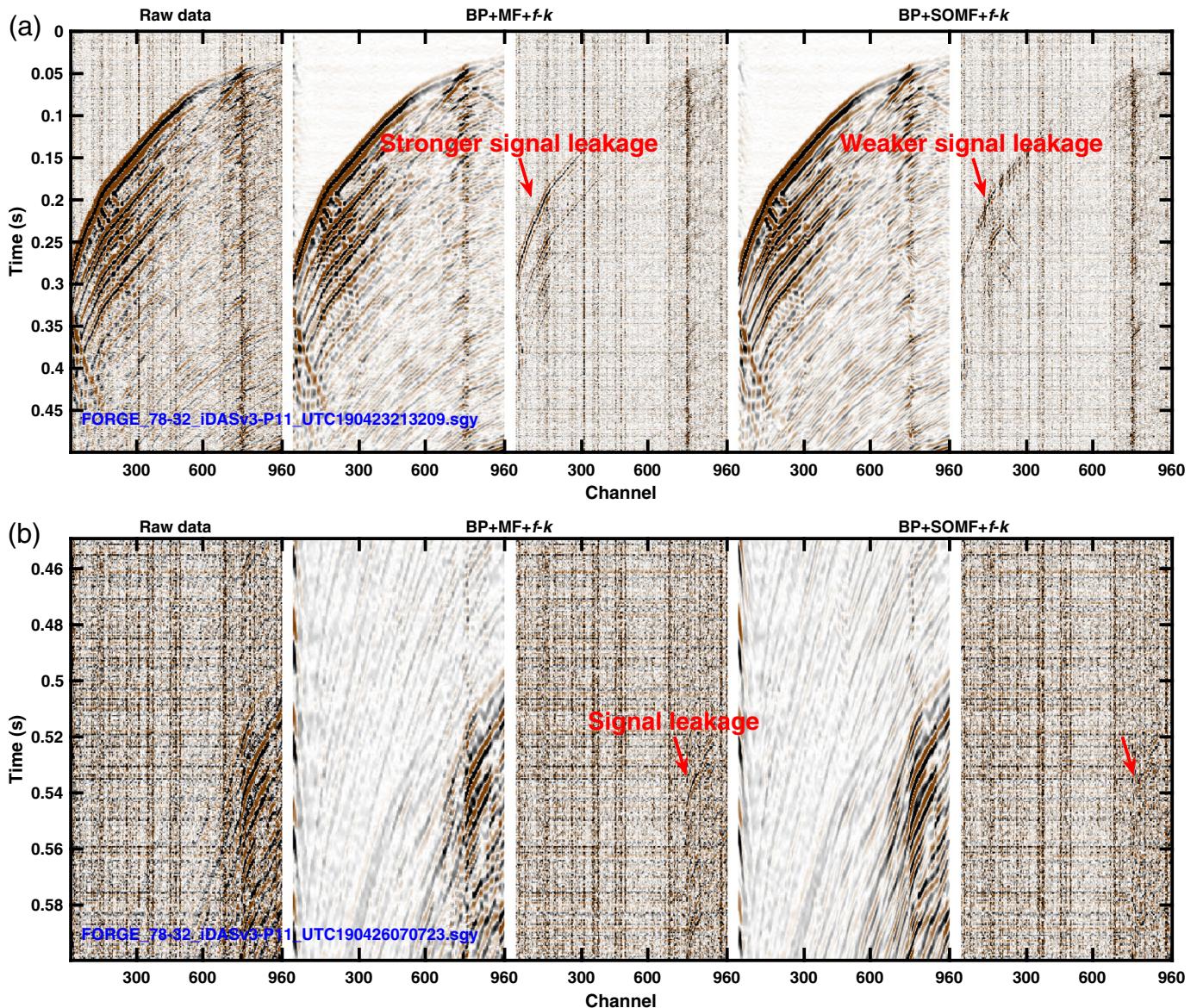
**Figure 4.** Trace-by-trace comparison of the denoising performance of the event recorded in FORGE\_78-32\_iDASv3-P11\_UTC190423213209.sgy with  $P$ -arrival sample number 1484. (a–i) Comparison of the single-traces in different denoising stages for different channels. The raw data in panels (c), (d), (h), and (i) contain high levels of erratic noise. The color version of this figure is available only in the electronic edition.

orthogonalization method (Chen and Fomel, 2015). The difference between MF and SOMF in terms of the final denoised effect is highlighted by the labeled red arrows. In Figure 6b, for a



**Figure 5.** More denoising examples using the proposed denoising framework in the case of very strong noise with barely observable signals. The raw data, denoised data, and removed noise are displayed side by side for six different events in different scenarios. (a,b) Two examples with strong recovered signals. (c,d) Two examples with moderately strong recovered signals.

(e,f) Two examples with weak recovered signals. All scenarios show much clear signals after denoising. The SEGY file names including these earthquake events are indicated on the bottom left of each pannel. The color version of this figure is available only in the electronic edition.



different event, we find that the MF causes visible signal leakage, whereas the signal leakage using the SOMF is negligible.

## Discussion

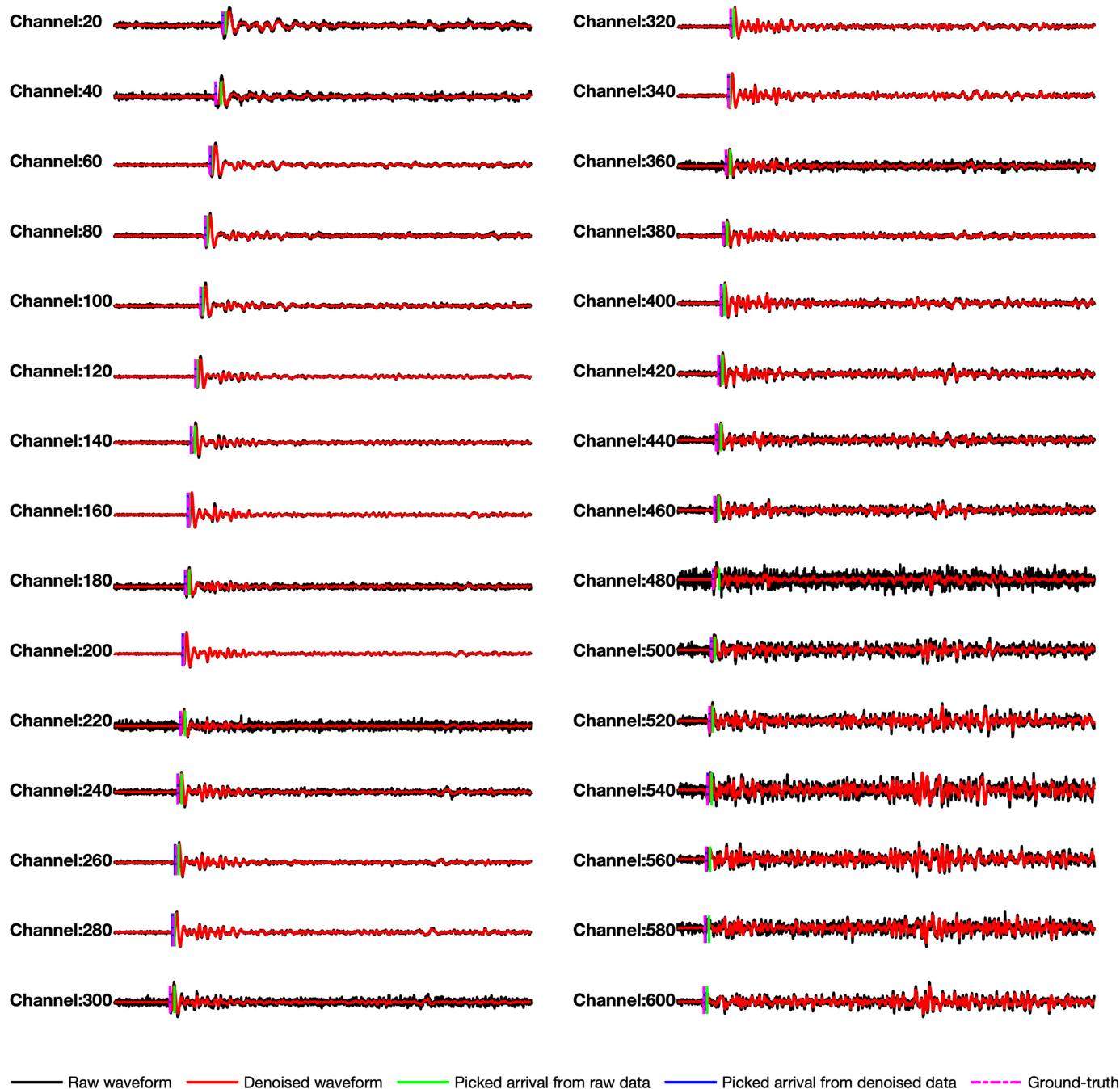
### Improved event detectability

The proposed denoising framework can significantly improve the signal quality of an input multichannel DAS dataset in cases of different noise levels. The improvement of event detectability due to the enhanced SNR of earthquake signals can also be analyzed in different cases. When the noise level of the raw data is not high compared with the amplitude of earthquake signals, the improvement of signal quality using the proposed method is limited. Correspondingly, the event detectability in terms of the accuracy of  $P$ -wave arrival picking can be slightly improved. Figure 7 shows the performance of the  $P$ -wave arrival picking using the short-term average/long-term average method (Trnkoczy, 2009) for 30 different

**Figure 6.** Denoising comparison between MF and SOMF. (a,b) The denoising comparison for different earthquake events. The color version of this figure is available only in the electronic edition.

channels of the DAS dataset analyzed previously in Figures 3 and 4. We show the detection results for both raw data (black waveforms) and denoised data (red waveforms) using green and blue trigger symbols, respectively. For a qualitative evaluation, we perform manual picking on the denoised data and treat the picks as the ground-truth solution. It is obvious that the picked arrivals (blue) from the denoised waveforms are more accurate than those (green) picked directly from the raw data, compared with the ground-truth solution (dashed magenta). This result represents the case of the mild noise level. In a different case in which the noise level is high and the earthquake signals are barely visible, we show the

## Earthquake detection of FORGE\_78-32\_iDASv3-P11\_UTC190423213209.sgy

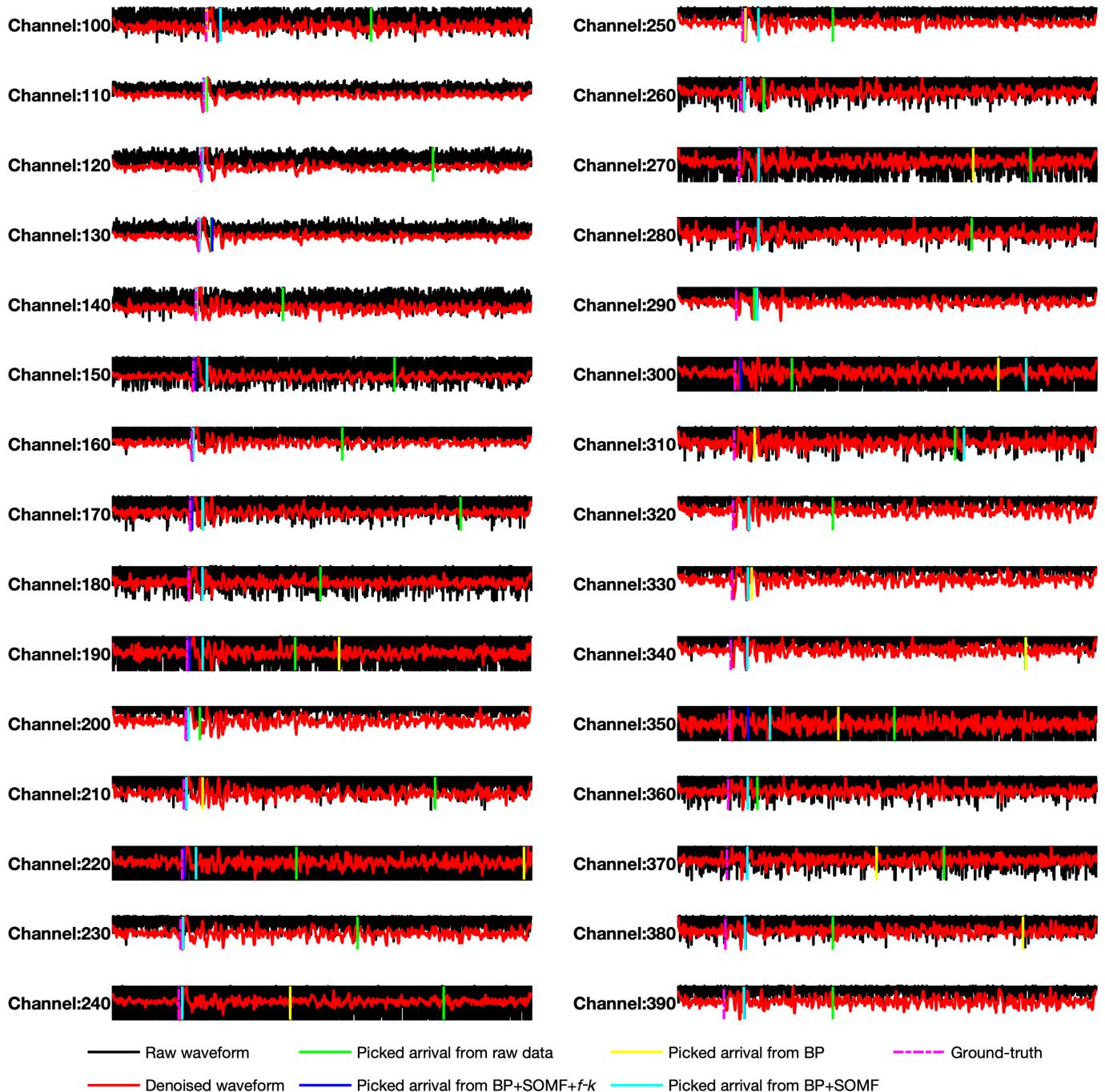


detection results in Figure 8, which corresponds to the same event that is demonstrated in Figure 5a. In Figure 8, the red denoised waveforms are plotted on top of the black background. The background corresponds to the waveforms of the raw data, which looks extremely noisy. The arrival picker fails (green arrivals) on most channels before denoising, whereas it succeeds in almost all channels after denoising. To evaluate the contribution of each individual filter to the detection performance, we also perform the picking on denoised data using BP only and using BP + SOMF. Their results are shown as yellow and cyan lines, respectively. It is

**Figure 7.** Comparison of earthquake detection and arrival picking performance before and after denoising for 30 channels in the case of mild noise. It is clear that denoising helps improve the accuracy of arrival picking slightly. The color version of this figure is available only in the electronic edition.

clear that the detection accuracy improves as more individual filters are involved in the framework. When only BP is used, there are many inaccurately picked arrivals. When BP and SOMF are used, detection becomes more accurate, but there

## Earthquake detection of FORGE\_78-32\_iDASv3-P11\_UTC190423150554.sgy

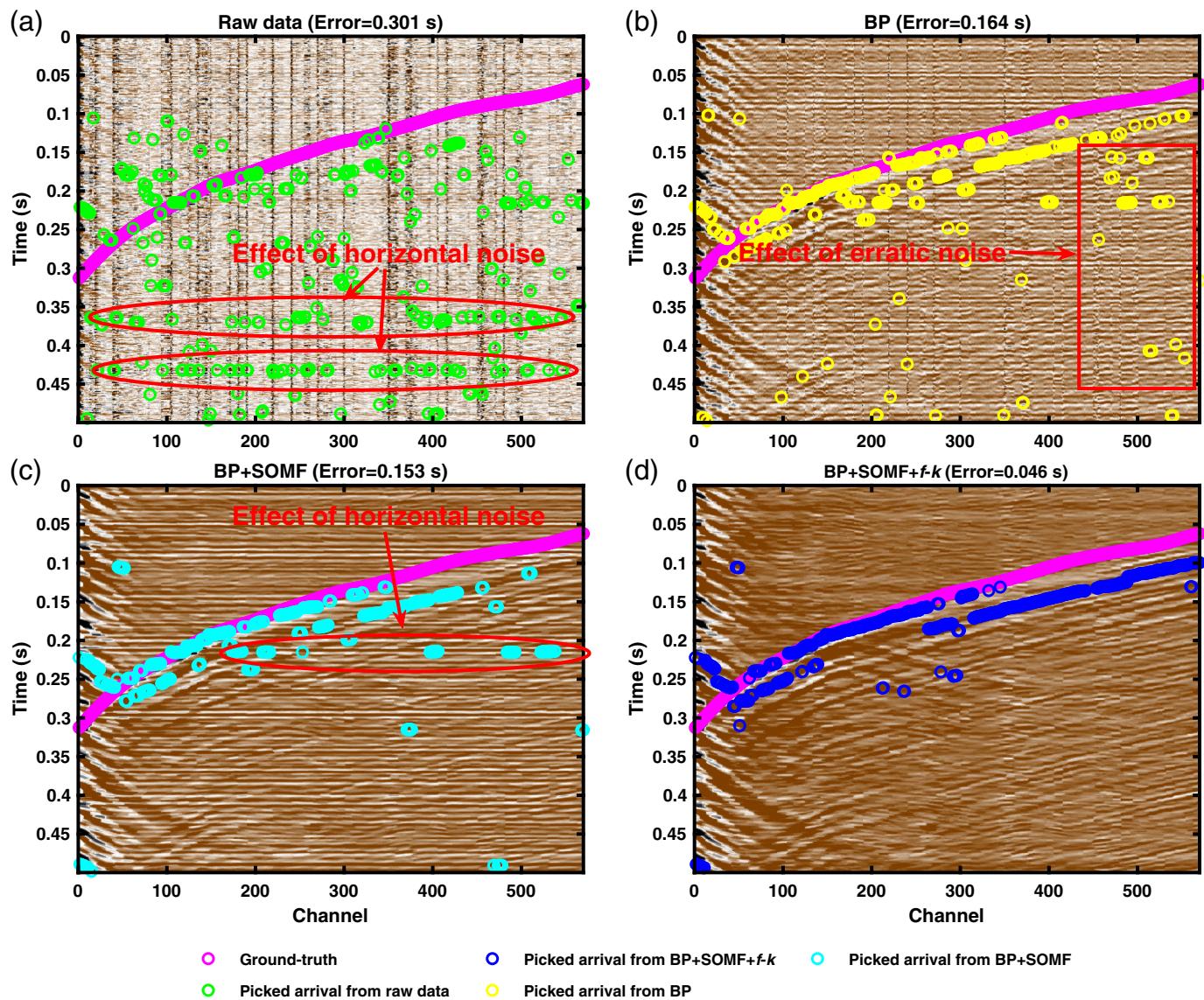


are obvious discrepancies between the ground-truth arrivals and the picked arrivals. The proposed integrated framework, however, helps obtain accurate arrival picking for most channels and has only small misfits between the picked arrivals and the ground-truth arrivals. We conclude that the proposed denoising framework can enhance the event detectability in different noise environments by improving the arrival picking and detecting missed events.

To better evaluate the influence of different noise components on the event detection and arrival picking, we plot different picked arrivals on the 2D gathers with decreasing noise

**Figure 8.** Comparison of earthquake detection and arrival picking performance before and after denoising for 30 channels in the case of strong noise. The amplitude of raw waveforms have been downscaled by a factor of 4 for display reasons. It is clear that the recovered signals after denoising are detectable, whereas the signals in the raw data can be missed. The color version of this figure is available only in the electronic edition.

levels in Figure 9. Figure 9a plots the raw data, ground-truth arrivals (magenta), and automatically picked arrivals (green) from the noisy raw data. Figure 9b plots the denoised

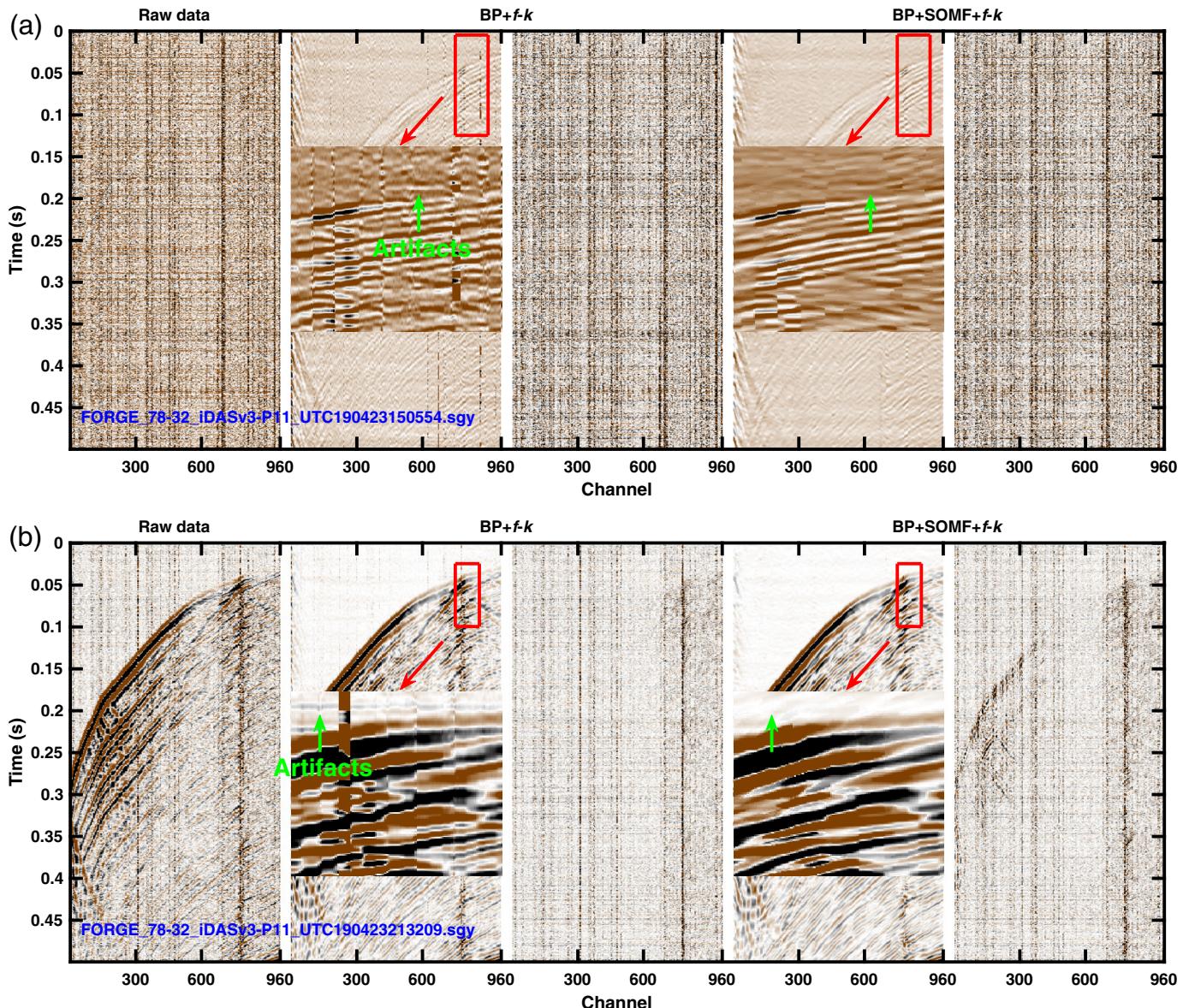


data using BP, ground-truth arrivals (magenta), and automatically picked arrivals (green) from the BP denoised data. Correspondingly, Figure 9c and 9d plots the denoised data and their picked arrivals using BP + SOMF and BP + SOMF +  $f\text{-}k$ , respectively. It is obvious that as more individual filters are included, arrivals are more accurately picked. Although there are some small discrepancies in some channels in Figure 9d, those arrivals are well aligned spatially. By observing the spatial distribution of those inaccurately picked arrivals in Figure 9a–c, we can understand how different noise components affect the arrival picking. For example, in Figure 9a, some green arrivals are clearly aligned along some horizontal noise components (see the ellipses and red arrows). In Figure 9b, after BP filtering, the wrong arrivals are mostly affected by the large-amplitude erratic noise (see the rectangle and red arrow). In Figure 9c, the arrivals are more accurate and better aligned, but there are still some picked arrivals aligned with the horizontal noise (see the ellipse and red arrow). To quantitatively compare the picking accuracy, we define the picking error as follows:

**Figure 9.** Comparison of earthquake detection and arrival picking performance before and after denoising using different combinations of individual filters in the case of strong noise. (a) Raw data. (b) bandpass (BP). (c) BP + SOMF. (d) BP + SOMF +  $f\text{-}k$ . It is clear that as more individual filters are included in the denoising framework, the arrival picking error becomes smaller. The color version of this figure is available only in the electronic edition.

$$\text{Error} = \sum_{i=1}^N |A_i - G_i|/N, \quad (10)$$

in which  $A_i$  and  $G_i$  denote the picked arrival time and the ground-truth arrival time of  $i$ th channel.  $N$  denotes the number of channels. Error measures the average picking error (in s) based on the STA/LTA method. Here, in the STA/LTA method, we choose the long term and short term as 200 and 30 samples, respectively. We use a triggering threshold of 0.98. As a result, the picking errors of the raw data, denoised data using BP, denoised data using BP + SOMF, and denoised data using



BP + SOMF +  $f\text{-}k$  are 0.301, 0.164, 0.153, and 0.046 s, respectively, which indicate a much improved arrival picking accuracy after applying the integrated denoising framework.

### Sequence of denoising steps

Although the proposed denoising method is a cascaded framework, the sequence of applying each component should not be changed. As explained in the last part, the BP filter acts as a preconditioning step for applying the SOMF; otherwise, the SOMF will be strongly affected by the strong high-frequency random noise. Therefore, the BP filter should be applied before the SOMF. The SOMF should be applied before the  $f\text{-}k$  dip filter because the high-amplitude erratic noise will strongly affect the performance of the  $f\text{-}k$  filter by causing locally horizontal artifacts. The time-domain response of the  $f\text{-}k$  filter is a horizontal component, which makes it able to suppress the horizontal and vertical (by transposing the data matrix) noise.

**Figure 10.** Denoising comparison between results using and without using an SOMF. (a,b) Two different earthquake events. It is clear that not using an SOMF before the  $f\text{-}k$  filtering will cause obvious horizontal artifacts, thereby affecting the arrival picking performance. It also demonstrates that the SOMF should be applied before the  $f\text{-}k$  filter. The color version of this figure is available only in the electronic edition.

However, the very strong erratic noise will have an imprint on the whole frequency band. The  $f\text{-}k$  filter will cause a discontinuous spectrum of the erratic noise, which is reflected as horizontal artifacts in the time domain. Figure 10 shows an example of such horizontal artifacts. Figure 10a shows a comparison between denoised results without and with the SOMF step based on an earthquake event recorded in the SEGY file named “FORGE\_78-32\_iDASv3-P11\_UTC190423150554.sgy.” When

an  $f$ - $k$  filter is applied without an SOMF, it creates obvious horizontal artifacts, as indicated by the green-labeled arrow in the zoomed red frame box. These horizontal artifacts are misleading when picking the arrivals. Figure 10b shows a more obvious example of horizontal artifacts caused by the erratic noise on the test data of a different event.

## Conclusion

The DAS seismic data have a much higher spatial sampling rate than the passive data acquired conventionally but with a much lower acquisition cost. The significantly lower SNR in an individual channel of DAS data, however, still limits its wide applications in the seismological community. Because the noise types are very diverse, no such single denoising method can deal with all types of noise in the DAS data. We propose an integrated framework that cascades several individual methods for denoising the DAS seismic data and demonstrate it to be very effective even in the case of very low SNR. The integrated denoising framework is mainly divided into three steps. The BP filtering method could remove all high-frequency noise but remains those noise components sharing a frequency band with the useful seismic signals. The SOMF can be further applied to remove large-amplitude erratic noise. Compared with a common MF, the SOMF can better preserve the signal energy following the spatial pattern of seismic waveforms. The dip filter in the  $f$ - $k$  domain can further attenuate regular vertical or horizontal noise by easily tuning the filter strength. Concatenation of the proposed framework with other well-developed filters for enhancing the noise removal in specific situations is straightforward. The proposed framework is easy to parameterize and obtain stable and robust results on a given dataset, for example, the FORGE geothermal DAS dataset.

## Data and Resources

Because the distributed acoustic sensing (DAS) seismic dataset at the Frontier Observatory for Research in Geothermal Energy (FORGE) geothermal site is open access, it is possible to make all the results of this article fully reproducible. Here, reproducibility not only means that the general denoising effect or performance of the proposed framework can be reproduced and validated (from the algorithmic point of view) but also means that all the numerical examples in this article can be fully reproduced with the exactly same plotting style. The same plotting style enables the peers to illustrate results when doing similar research conveniently. The subroutines that are used in this article either for processing the raw DAS seismic data or plotting the results are all included in an online repository (<https://github.com/chenyk1990/dasdenoising>, last accessed October 2022). The processing package only depends on one external package, that is, the slope estimation and structural filtering package, *seistr*, published recently by Wang *et al.* (2022). In addition to the results presented here, all other results, including denoising performance of all the 82 earthquake events and 112 microseismic events, as cataloged in Lellouch *et al.* (2021), are included in the online repository. All results are produced using the same set of parameters, that is, the cutoff

frequency in the bandpass (BP) filter is 200 Hz, the lateral and vertical smoothing radii in the slope estimation of the structure-oriented median filter (SOMF) are both 40 samples, the numbers of linear and nonlinear iterations in the slope estimation are 10 and 3, respectively, the filter length of the SOMF is 17 samples, the strength of the  $f$ - $k$  filter is 0.02. The proposed denoising framework appears to be very stable considering the good denoising performance of all the testing datasets regardless of the data quality or noise levels of the input data.

## Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

## Acknowledgments

The authors are grateful for the constructive comments from the editors and anonymous reviewers that significantly improved this work. The research is partially supported by sponsors of the Texas Consortium of Computational Seismology (TCCS), research funding granted to the Texas Seismological Network Project (TexNet), and the starting funds from Zhejiang University.

## References

- Ajo-Franklin, J. B., S. Dou, N. J. Lindsey, I. Monga, C. Tracy, M. Robertson, V. R. Tribaldos, C. Ulrich, B. Freifeld, T. Daley, *et al.* (2019). Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection, *Sci. Rep.* **9**, no. 1, 1–14.
- Atterholt, J., Z. Zhan, Z. Shen, and Z. Li (2022). A unified wavefield-partitioning approach for distributed acoustic sensing, *Geophys. J. Int.* **228**, no. 2, 1410–1418.
- Butterworth, S. (1930). On the theory of filter amplifiers, *Wirel. Eng.* **7**, 536–541.
- Chen, J., J. Ning, W. Chen, X. Wang, W. Wang, and G. Zhang (2019). Distributed acoustic sensing coupling noise removal based on sparse optimization, *Interpretation* **7**, no. 2, T373–T382.
- Chen, W., H. Wang, L. Yang, X. Liu, and Y. Chen (2022). Non-stationary local slope estimation via forward-backward space derivative calculation, *Geophysics* **87**, no. 1, O1–O11, doi: [10.1190/geo2021.0255.1](https://doi.org/10.1190/geo2021.0255.1).
- Chen, Y., and S. Fomel (2015). Random noise attenuation using local signal-and-noise orthogonalization, *Geophysics* **80**, no. 6, WD1–WD9.
- Chen, Y., J. Ma, and S. Fomel (2016). Double-sparsity dictionary for seismic noise attenuation, *Geophysics* **81**, no. 2, V17–V30.
- Cheng, F., B. Chi, N. J. Lindsey, T. C. Dawe, and J. B. Ajo-Franklin (2021). Utilizing distributed acoustic sensing and ocean bottom fiber optic cables for submarine structural characterization, *Sci. Rep.* **11**, no. 1, 1–14.
- Dong, X., and Y. Li (2020). Denoising the optical fiber seismic data by using convolutional adversarial network based on loss balance, *IEEE Trans. Geosci. Remote Sens.* **589**, no. 12, 10,544–10,554.
- Dou, S., N. Lindsey, A. M. Wagner, T. M. Daley, B. Freifeld, M. Robertson, J. Peterson, C. Ulrich, E. R. Martin, and J. B. Ajo-Franklin (2017). Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study, *Sci. Rep.* **7**, no. 1, 1–12.

- Fomel, S. (2002). Application of plane-wave destruction filters, *Geophysics* **67**, no. 6, 1946–1960.
- Harris, K., D. White, D. Melanson, C. Samson, and T. M. Daley (2016). Feasibility of time-lapse VSP monitoring at the aquistore CO<sub>2</sub> storage site using a distributed acoustic sensing system, *Int. J. Greenh. Gas Control* **50**, 248–260.
- Huang, W., R. Wang, and Y. Chen (2018). Regularized non-stationary morphological reconstruction algorithm for weak signal detection in micro-seismic monitoring: Methodology, *Geophys. J. Int.* **213**, 1189–1211.
- Hull, R., R. Meek, H. Bello, and D. Miller (2017). Case history of DAS fiber-based microseismic and strain data, monitoring horizontal hydraulic stimulations using various tools to highlight physical deformation processes (Part A), *Unconventional Resources Technology Conference*, Austin, Texas, 24–26 July 2017, 3050–3062, Society of Exploration Geophysicists, American Association of Petroleum.
- Ibrahim, A. D. A., S. Lin, J. Xiong, J. Jiang, Y. Fu, and Z. Wang (2020). Integrated principal component analysis denoising technique for phase-sensitive optical time domain reflectometry vibration detection, *Appl. Opt.* **59**, no. 3, 669–675.
- Johny, J., S. Amos, and R. Prabhu (2021). Optical fibre-based sensors for oil and gas applications, *Sensors* **21**, no. 18, 6047.
- Lellouch, A., R. Schultz, N. J. Lindsey, B. Biondi, and W. L. Ellsworth (2021). Low-magnitude seismicity with a downhole distributed acoustic sensing array—examples from the forge geothermal experiment, *J. Geophys. Res.* **126**, no. 1, e2020JB020462, doi: [10.1029/2020JB020462](https://doi.org/10.1029/2020JB020462).
- Lellouch, A., S. Yuan, W. L. Ellsworth, and B. Biondi (2019). Velocity-based earthquake detection using downhole distributed acoustic sensing—examples from the San Andreas fault observatory at depth, *Bull. Seismol. Soc. Am.* **109**, no. 6, 2491–2500.
- Li, Z., Z. Shen, Y. Yang, E. Williams, X. Wang, and Z. Zhan (2021). Rapid response to the 2019 Ridgecrest earthquake with distributed acoustic sensing, *AGU Adv.* **2**, no. 2, doi: [10.1029/2021AV000395](https://doi.org/10.1029/2021AV000395).
- Lindsey, N. J., T. C. Dawe, and J. B. Ajo-Franklin (2019). Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing, *Science* **366**, no. 6469, 1103–1107.
- Lindsey, N. J., E. R. Martin, D. S. Dreger, B. Freifeld, S. Cole, S. R. James, B. L. Biondi, and J. B. Ajo-Franklin (2017). Fiber-optic network observations of earthquake wavefields, *Geophys. Res. Lett.* **44**, no. 23, 11–792.
- Lindsey, N. J., S. Yuan, A. Lellouch, L. Gaultier, T. Lecocq, and B. Biondi (2020). City-scale dark fiber DAS measurements of infrastructure use during the covid-19 pandemic, *Geophys. Res. Lett.* **47**, no. 16, e2020GL089931, doi: [10.1029/2020GL089931](https://doi.org/10.1029/2020GL089931).
- Martin, E. R., C. M. Castillo, S. Cole, P. S. Sawasdee, S. Yuan, R. Clapp, M. Karrenbach, and B. L. Biondi (2017). Seismic monitoring leveraging existing telecom infrastructure at the SDASA: Active, passive, and ambient-noise analysis, *Lead. Edge* **36**, no. 12, 1025–1031.
- Mateeva, A., J. Lopez, J. Mestayer, P. Wills, B. Cox, D. Kiyashchenko, Z. Yang, W. Berlang, R. Detomo, and S. Grandi (2013). Distributed acoustic sensing for reservoir monitoring with VSP, *Lead. Edge* **32**, no. 10, 1278–1283.
- Oppenheim, A. V., and R. W. Schafer (2009). *Discrete-Time Signal Processing*, Third Ed., Prentice Hall, Hoboken, New Jersey.
- Proakis, J. G., and D. K. Manolakis (2006). *Digital Signal Processing*, Fourth Ed., Prentice Hall, Hoboken, New Jersey.
- Qin, Z., H. Chen, and J. Chang (2017). Detection performance improvement of distributed vibration sensor based on curvelet denoising method, *Sensors* **17**, no. 6, 1380.
- Qin, Z., L. Chen, and X. Bao (2012). Wavelet denoising method for improving detection performance of distributed vibration sensor, *IEEE Photonics Technol. Lett.* **24**, no. 7, 542–544.
- Sladen, A., D. Rivet, J.-P. Ampuero, L. De Barros, Y. Hello, G. Calbris, and P. Lamare (2019). Distributed sensing of earthquakes and ocean-solid earth interactions on seafloor telecom cables, *Nat. Commun.* **10**, no. 1, 1–8.
- Spica, Z. J., M. Perton, E. R. Martin, G. C. Beroza, and B. Biondi (2020). Urban seismic site characterization by fiber-optic seismology, *J. Geophys.* **125**, no. 3, e2019JB018656, doi: [10.1029/2019JB018656](https://doi.org/10.1029/2019JB018656).
- Spikes, K. T., N. Tisato, T. E. Hess, and J. W. Holt (2019). Comparison of geophone and surface-deployed distributed acoustic sensing seismic data, *Geophysics* **84**, no. 2, A25–A29.
- Trnkoczy, A. (2009). Understanding and parameter setting of STA/LTA trigger algorithm, *New Manual of Seismological Observatory Practice (NMSOP)*, Deutsches GeoForschungsZentrum GFZ, 1–20.
- Walter, F., D. Gräff, F. Lindner, P. Paitz, M. Köpfli, M. Chmiel, and A. Fichtner (2020). Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain, *Nat. Commun.* **11**, no. 1, 1–10.
- Wamrrew, D., R. Pevzner, E. Maltsev, and D. Pisarenko (2021). Deep neural networks for detection and location of microseismic events and velocity model inversion from microseismic data acquired by distributed acoustic sensing array, *Sensors* **21**, no. 19, 6627.
- Wang, H., Y. Chen, O. M. Saad, Y. A. S. I. Oboué, L. Yang, S. Fomel, and Y. Chen (2022). A matlab code package for 2D/3D local slope estimation and structural filtering, *Geophysics* **87**, no. 3, F1–F14, doi: [10.1190/geo2021.0266.1](https://doi.org/10.1190/geo2021.0266.1).
- Wang, H., G. Huang, and Y. Chen (2021). Robust nonstationary local slope estimation, *IEEE Trans. Geosci. Remote Sens.* **59**, no. 7, 6225–6233, doi: [10.1109/TGRS.2020.3021375](https://doi.org/10.1109/TGRS.2020.3021375).
- Webster, P., M. Molenaar, and C. Perkins (2016). DAS microseismic fiber-optic locating DAS microseismic events and errors, *CSEG Rec.* **41**, 38–39.
- Willis, M. E. (2022). Distributed acoustic sensing for seismic measurements – What geophysicists and engineers need to know, doi: [10.1190/1.9781560803850](https://doi.org/10.1190/1.9781560803850).
- Yuan, S., A. Lellouch, R. G. Clapp, and B. Biondi (2020). Near-surface characterization using a roadside distributed acoustic sensing array, *Lead. Edge* **39**, no. 9, 646–653.
- Zu, S., H. Zhou, Q. Li, H. Chen, Q. Zhang, W. Mao, and Y. Chen (2017). Shot-domain deblending using least-squares inversion, *Geophysics* **82**, no. 4, V241–V256.

Manuscript received 13 April 2022

Published online 10 November 2022