

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Infrasound single-channel noise reduction: application to detection and localization of explosive volcanism in Alaska using backprojection and array processing

Permalink

<https://escholarship.org/uc/item/50s924rz>

Journal

Geophysical Journal International, 232(3)

ISSN

0956-540X

Authors

Sanderson, Richard W
Matoza, Robin S
Fee, David
[et al.](#)

Publication Date

2022-11-14

DOI

10.1093/gji/ggac182

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

Title: Infrasound single-channel noise reduction: application to detection and localization of explosive volcanism in Alaska using backprojection and array processing

Authors: Richard W. Sanderson¹, Robin S. Matoza¹, David Fee², Matthew M. Haney³, John J. Lyons³

Affiliations: ¹Department of Earth Science and Earth Research Institute, University of California, Santa Barbara, CA, USA, ²Wilson Alaska Technical Center, Alaska Volcano Observatory, Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA, ³U.S. Geological Survey Volcano Science Center, Alaska Volcano Observatory, Anchorage, AK, USA

Initial submission: 12 October 2021

Final files submitted: 6 May 2022

Abbreviated title: Infrasound noise-reduction processing

Corresponding author:

Richard W. Sanderson

Dept. of Earth Science, 1006 Webb Hall, University of California, Santa Barbara, California, 93106-9630

rws@ucsb.edu

<https://orcid.org/https://orcid.org/0000-0003-3772-4144>

Co-author contact information:

Robin S. Matoza, rmatoza@ucsb.edu, <https://orcid.org/0000-0003-4001-061X>

David Fee, dfee1@alaska.edu, <https://orcid.org/https://orcid.org/0000-0002-0936-9977>

Matthew M. Haney, mhaney@usgs.gov, <https://orcid.org/https://orcid.org/0000-0003-3317-7884>

John. J. Lyons, jlyons@usgs.gov, <https://orcid.org/https://orcid.org/0000-0001-5409-1698>

Summary

Infrasound sensors are deployed in a variety of spatial configurations and scales for geophysical monitoring, including networks of single sensors and networks of multi-sensor infrasound arrays. Infrasound signal detection strategies exploiting these data commonly make use of inter-sensor correlation and coherence (array processing, multi-channel correlation); network-based tracking of signal features (e.g., reverse time migration); or a combination of these such as backazimuth cross-bearings for multiple arrays. Single-sensor trace-based denoising techniques offer significant potential to improve all of these various infrasound data processing strategies, but have not previously been investigated in detail. Single-sensor denoising represents a preprocessing step that could reduce the effects of ambient infrasound and wind noise in infrasound signal association and location workflows. We systematically investigate the utility of a range of single-sensor denoising methods for infrasound data processing, including noise gating, non-negative matrix factorisation, and data-adaptive Wiener filtering. For the data testbed, we use the relatively dense regional infrasound network in Alaska, which records a high rate of volcanic eruptions with signals varying in power, duration, and waveform and spectral character. We primarily use data from the 2016–2017 Bogoslof volcanic eruption, which included multiple explosions, and synthetics. The Bogoslof volcanic sequence provides an opportunity to investigate regional infrasound detection, association, and location for a set of real sources with varying source spectra subject to anisotropic atmospheric propagation and varying noise levels (both incoherent wind noise and coherent ambient infrasound, primarily microbaroms). We illustrate the advantages and disadvantages of the different denoising methods in categories such as event detection, waveform distortion, the need for manual data labelling, and computational cost. For all approaches, denoising generally performs better for signals with higher SNR and with less spectral and temporal overlap between signals and noise. Microbaroms are the most globally pervasive and repetitive coherent ambient infrasound noise source, with such noise often referred to as clutter or interference. We find that denoising offers significant potential for microbarom clutter reduction. Single-channel denoising of microbaroms prior to standard array processing enhances both the quantity and

bandwidth of detectable volcanic events. We find that reduction of incoherent wind noise is more challenging using the denoising methods we investigate; thus, station hardware (wind noise reduction systems) and site selection remain critical and cannot be replaced by currently available digital denoising methodologies. Overall, we find that adding single-channel denoising as a component in the processing workflow can benefit a variety of infrasound signal detection, association, and location schemes. The denoising methods can also isolate the noise itself, with utility in statistically characterizing ambient infrasound noise.

Key words: Infrasound, explosive volcanism, volcano monitoring, time-series analysis, statistical methods, volcano seismology

1. Introduction

Between 2016 and 2021, seven different volcanoes erupted in Alaska, spanning both a wide geographical area and a range of explosive and effusive behaviours (Fig. 1). Notable among these was the December 2016 to August 2017 eruption of Bogoslof volcano, which produced seventy explosive eruptions (e.g., Coombs *et al.*, 2019). With some plumes reaching over 10 km above sea level at this isolated island, hazards to passing aircraft were a pressing concern. An absence of on-ground instrumentation meant that all observations were made remotely, including using seismicity, satellite imagery, lightning detection, and infrasound. Infrasound refers to acoustic waves with frequencies below 20 Hz, the typical lower threshold of human hearing. Volcanic explosions radiate infrasound through the atmosphere, where it can propagate for thousands of kilometres (e.g., Matoza *et al.*, 2011a). This feature makes infrasound a useful tool for remote detection and location of activity at volcanoes such as Bogoslof, and others in Alaska. The arrival of the EarthScope Transportable Array (TA) in Alaska, beginning in 2014 (Fig. 1), was unprecedented in terms of having a large, dense, regional seismo-acoustic network close to a volcanic arc. The TA complemented sparse infrasound stations already in place, operated by the Alaska Volcano Observatory (AVO). From 2019–2021, as the TA’s mission wound down, many sites were transferred to other operators

such as the Alaska Earthquake Center (AEC), with the remaining sites removed entirely. In this paper we continue to refer to the TA, as that was the network in place during the eruption being assessed.

1.1. Background and motivation

In previous work we explored using the TA and regional infrasound stations for detection and location of explosive eruptions in the Aleutian Arc and Kamchatka, Russia (Sanderson *et al.*, 2020a). We found that a simple envelope-based backprojection technique, and automated event identification process, were able to capture up to 85% of the infrasound generating events from Bogoslof volcano that were catalogued by AVO (Coombs *et al.*, 2019). Notable limitations to the scheme come from anisotropic atmospheric propagation and low signal-to-noise (SNR) conditions, with wind being the dominant noise source. Here we focus on improving waveform data processing to reduce noise, thereby enhancing signal detection and source location accuracy. Signals from Bogoslof volcano often have similar character to microbaroms (ocean noise) and wind, providing a strong challenge for signal separation. In part this is due to attenuation and remote detection, but also due to the often submarine nature of the eruptions (e.g., Lyons *et al.*, 2019).

In this study, we extend our focus from a backprojection location technique (reverse time migration, RTM) to include array processing. Both strategies work on the principle of beamforming (stacking) signals from different sensors according to an acoustic velocity model, such that constructive and destructive interference improves the amplitude of acoustic arrivals and decreases non-acoustic noise sources, such as wind. Wind noise is a common problem that can make beamforming ineffective for three principal reasons: (1) in array processing, local wind turbulence (among other factors) can reduce the coherence of signals of interest between the array elements; (2) in RTM, wind affects the amplitude of traces, which are typically normalized prior to stacking; and (3) again in RTM, a common approach is to take envelopes of waveforms before stacking, with this smoothing effect reducing the impact of destructive interference. One of our aims, therefore, is to explore how wind noise can be removed from our data prior to beamforming, and whether this removal provides benefits to detection and location strategies.

A second natural phenomenon that impacts infrasound records are microbaroms. These are acoustic arrivals generated by non-linear wave-wave, and wave-shore interaction, with a spectral peak between 0.1–0.4 Hz (e.g., Bowman *et al.*, 2005). This frequency range overlaps with the majority of the Bogoslof volcanic events (e.g., Fee *et al.*, 2020; Lyons *et al.*, 2020), more noticeably at remote stations where attenuation has reduced higher frequency components. Microbaroms are detectable worldwide due to their low frequencies, and there are many potential source areas (e.g., Landès *et al.*, 2012; Matoza *et al.*, 2013). In Alaska, the primary origin and strength of microbaroms varies over the course of the year, such that backazimuth of the microbaroms and volcanoes may periodically coincide. Being acoustic arrivals, microbaroms often constructively add during beamforming (depending on the algorithm and spatial extent of the source region). Such directional interference can cause false alarms and missed detections in monitoring applications (e.g., Matoza *et al.*, 2011b). This pervasiveness leads to microbaroms being the main source of so-called clutter (Ceranna *et al.*, 2019; Mialle *et al.*, 2019). Consequently, we also aim to see if reducing microbaroms improves array processing and RTM performance.

1.2. Single-channel noise removal approaches

To our knowledge, there is limited prior literature on denoising individual infrasound traces, as SNR enhancement usually comes through beamforming. These denoising methods enable a waveform to be separated into signal and noise components, enabling manipulation of either. This contrasts with classification approaches that identify time periods of signals and noise, but do not modify the signal. One recent development is a study by Carniel *et al.* (2014), with the supervised approach employing statistical foreground activity detection, non-negative matrix factorisation (NMF), and Wiener filtering to reduce wind noise affecting volcanic tremor. Another recent study, by Williams *et al.* (2020), showed how denoising infrasound traces could improve the number of volcanic eruption detections made during array processing by reducing microbaroms. This approach is based on semi-supervised adaptive spectral subtraction using non-local means. Finally, Cook *et al.* (2021) demonstrated a wind noise detection and compensation technique based on the characteristic spectral shape of wind noise in the presence of a

windscreen, using synthetic sources in the low-audio range. These approaches do not currently exist as open-source tools, and are limited in terms of test cases, factors we aim to improve on. Many contemporary signal separation techniques stem from developments more commonly seen in the audio, music, and image processing industries going back some decades (e.g., Boll, 1979). Seismology has also had some recent developments in denoising single traces, using statistical threshold and machine learning based techniques (e.g., Cabras *et al.*, 2012; Mousavi and Langston, 2016a, 2016b, 2017; Langston and Mousavi, 2019; Zhu *et al.*, 2019; Jiang *et al.*, 2020). A range of other seismic-based techniques are noted by Mousavi and Langston (2016a). Approaches such as seismo-acoustic coherence and correlation (e.g., Matoza and Fee, 2014; Fee *et al.*, 2017; Ichihara *et al.*, 2021) take advantage of co-located sensors and air-to-ground coupled signals. Median filters can be effective for removing (or keeping) impulsive signals and white Gaussian noise (e.g., Bednar, 1983).

1.3. Objectives and paper organisation

To remove unwanted wind noise and microbaroms from our data, we examine several existing open-source data decomposition tools. The approaches span fully supervised (training with signals and noise), semi-supervised (training with noise only), and unsupervised classes. The methods can be applied to a single trace, however, without need for large training datasets. Our analysis covers several kinds of volcanic events, station types, and source-station distances. We aim to identify how well these kinds of trace denoising approaches are beneficial in terms of detecting and locating events under these different but common circumstances. As with any realistic implementation of a processing workflow, there is no one optimal method or set of parameter choices, particularly with such a large number of free parameters (e.g., Withers *et al.*, 1998; Mialle *et al.*, 2019). For evaluation criteria here, we use metrics like SNR, distortion, and phase changes, which are particularly relevant for meeting the goals of the project.

The organization of the paper is as follows: First, we describe the typical characteristics of Bogoslof volcano infrasound, wind noise, and microbarom clutter, as well as how hardware and other site characteristics can affect these observations. Next, we review the data decomposition methods used for

isolating signals of interest, and then the array processing and RTM schemes for detecting and locating events. Finally, we present our results, which show that microbarom reduction primarily improves array processing by not only reducing clutter, but also by enhancing event detections, particularly in the band previously dominated by microbaroms. Some methods only reduce microbarom amplitude, whereas others also reduce residual microbarom inter-sensor similarity and thus detectability. Wind noise reduction does not strongly affect array processing performance, but can improve SNR and event identification for backprojected data.

2. Data and station characteristics

The geological, geochemical, and geophysical nature of the Bogoslof volcanic eruption sequence is presented in a special journal issue (Waythomas *et al.*, 2019), following several prior publications (e.g., Coombs *et al.* 2018, and references therein). Of particular relevance here are those studies which address the infrasound character of the eruptions, including Haney *et al.* (2018), Wech *et al.* (2018), Lyons *et al.* (2019), Haney *et al.* (2020), and Schwaiger *et al.* (2020). Infrasound records of the eruptions persist from minutes to hours, with the frequency range of the eruptions spanning from below 0.1 Hz, to more than 10 Hz (as recorded at the nearby Okmok array), depending on whether the eruption was at a submarine or subaerial stage (Fee *et al.*, 2020; Lyons *et al.*, 2020). Higher frequency components are attenuated with distance. For eruption catalogue information, we refer to Coombs *et al.* (2019), who included which infrasound arrays made observations, as well as detection start and stop times. Lyons *et al.* (2020) had a differing record of which arrays recorded events in some instances. Differences in detection rates between sources may be due to alternative array processing methods and thresholds (Section 3.4). We include notes in the Supporting Information on some of these variations, as well as other potential improvements to catalogues. All dates here are in Coordinated Universal Time (UTC).

Wind noise nominally has a spectrum with amplitude inversely proportional to frequency. This slope largely reflects the inertial subrange of the turbulence spectra, produced by atmospheric eddies of varying wavenumber and kinetic energy (e.g., Raspet *et al.*, 2006). How wind noise presents at, and affects

the stations in Alaska, can vary drastically due to hardware differences, location, and time of year. In this study, we evaluate data from all infrasound sensors within 2,500 km of Bogoslof Island, including the TA, an array which is part of the International Monitoring System (IMS IS53), and six arrays operated by AVO (Fig. 1). The AVO operated array locations and station codes are as follows: Akutan (AKS), Adak (ADKI), Cleveland (CLCO), Dillingham (DLL), Okmok (OKIF), and Sand Point (SDPI). Infrasound sensors generally have a flat response from 0.01 or 0.1 Hz to 50 Hz, and four to 8 elements per array. During the eruption, ADKI, DLL, IS53, and OKIF had either domes or pipe systems acting to reduce higher frequency wind noise through spatial averaging (Dixon *et al.*, 2019; Lyons *et al.*, 2020). As the majority of remaining stations have limited wind noise reduction hardware in the form of basic diffusers (Busby and Aderhold, 2020), detection capability is particularly affected, and efforts to reduce wind noise must be done post-recording. Snow cover can also act as a lowpass filter (Adam *et al.*, 1998). Stations away from coasts and at forested sites are typically less windy. Sources closer to stations will have limited path attenuation, and so broader frequency ranges will be recorded. Proximal stations without wind noise reduction hardware will often have relatively similar eruption and wind noise characteristics.

3. Methods

Here we introduce the techniques employed to prepare data and to separate the events of interest from microbaroms and wind noise. We also describe the array processing and RTM strategies used to detect and locate events. Fig. 2 provides an overview of these stages, with parameters and program links given in the Supporting Information. The processing workflow is similar to those of, e.g., Cabras *et al.* (2010), and Carniel *et al.* (2014), who began with removal of persistent background sources, and then subdivided the remaining data into signals and noise.

3.1. Preprocessing

We preprocess all traces via the following steps: (1) taper the waveform; (2) filter above 0.01 Hz; (3) remove instrument responses using a water level at 35 dB; (4) down-sample to 20 Hz after applying an

anti-aliasing filter. These parameters are used to limit amplification of very low frequency noise, and ensure workable computational processing times by limiting the sample rate, while retaining dominant event and noise frequencies (up to 10 Hz). This preprocessing is termed stage 1 of the methodology sequence (Fig. 2). For local networks, where event spectra deviate from background noise at higher frequencies than the examples here, use of the full sample rate may be beneficial.

3.2. Microbarom removal

Stage 2 of the workflow is to remove microbarom signals from the remaining data, which comprises potential events of interest, as well as high-amplitude intermittent wind noise. In this study we compare five different methods to evaluate some of the advantages and disadvantages of each. For all methods except for nearest-neighbour filtering, a section of microbarom must be labelled as a reference. It is also possible to use a microbarom profile from an archive or use a synthetic spectral representation. This can be useful in instances where a trace does not have a viable microbarom-only section. Brief descriptions of the five methods follow, and we refer the reader to the references cited for further details. Several methods overlap in procedure, and thus are not wholly independent concepts. Parameter choices are those that result in similarly reduced microbarom spectral amplitudes. For each method, it is possible to reduce the microbarom amplitude further, but this can be at the expense of the event waveform.

Even when local winds are minimal, infrasound spectra will still typically present with a turbulence-induced negative slope, superposed with a microbarom peak. Stage 2 methods using a noise reference window will inevitably encompass both aspects of the ambient infrasound, unless some prior compensation is applied (e.g., Cook *et al.*, 2021). Consequently, while this stage removes the microbarom as primarily intended, persistent atmospheric effects are also reduced. In this paper we will refer to both components of this lower-noise ambient state as the microbarom to avoid confusion with more prominent wind noise. Acronyms for the following methods are included in Table 1 for reference.

3.2.1. Block choice (BC)

We use the BCseis (Block Choice Seismic Analysis) program to adjust the power in time-frequency blocks formed via spectral transformation of data (Langston and Mousavi, 2019). This method incorporates continuous wavelet transforms, making it unique among the tools described here, with the others based on Fourier transformations. The transformed data are plotted as wavelet scale (akin to frequency) versus time lag, with the block amplitude representing the wavelet coefficients. Based on the statistics of the noise sample specified, a threshold value is determined for each wavelet scale. For hard thresholding, blocks with amplitudes below this threshold are zeroed out. Soft thresholding takes this further by also deducting the scale-dependent threshold value from the amplitudes of all remaining blocks. The latter method reportedly performs better at minimizing noise outliers (Langston and Mousavi, 2019). In this study we implemented the hard thresholding option, however. This was a result of evaluating the performance of each thresholding type under the aforementioned condition of keeping the post-denoising spectral amplitudes of the microbaroms similar between methods (for effective comparison). In Section 4.1, we revisit differences between the two approaches when discussing the impact of parameter choices.

3.2.2. Harmonic regeneration (HR)

Harmonic regeneration noise reduction (HRNR; Plapous *et al.*, 2006) is the final part of this data-adaptive method. For each time step, the current a posteriori SNR (ratio of noisy signal power and noise power), and a priori SNR (ratio of clean signal power and noise power) from the previous step are used to determine the new a priori SNR and spectral gain required for a clean signal. Noise is removed using a Wiener-type filter, and following additional gain control, HRNR reduces harmonic distortion in the result.

3.2.3. Non-negative matrix factorisation (NMF)

NMF is the process of factorising a matrix \mathbf{X} into two other matrices, \mathbf{W} and \mathbf{H} . In our case, \mathbf{X} is a spectrogram. \mathbf{W} is composed of basis vectors, also known as dictionary elements, representing the spectral amplitude at each frequency. \mathbf{H} is a matrix of activations, also known as weights or gains, representing

when the basis vectors occur. There are many popular NMF algorithms that perform this task, and here we use the Kullback-Leibler application by Bryan *et al.* (2013). To ensure the preprocessed trace is split into a foreground (components of interest) and background (undesired elements, typically noise), the basis vectors in \mathbf{W} are pre-grouped into two sets, one set based on the frequency characteristics of the microbarom sample, and the other set based on everything else in the trace. The initial state of \mathbf{H} is randomized. The individual time series are then reconstructed by using a masking filter to downweigh the undesired spectral elements. We did not have success when keeping \mathbf{W} randomized, i.e., exploring whether the data naturally falls into two groups. Without pre-grouping, the resulting components of \mathbf{W} and \mathbf{H} have no ordering, necessitating some type of rule-based clustering mechanism. Given the large frequency overlap between Bogoslof volcanic events, wind noise, and microbaroms, effective clustering rules would be difficult to design.

3.2.4. Nearest-neighbour filtering (NN)

For this technique, no noise region needs to be specified, making the method fully unsupervised (McFee *et al.*, 2015). The process first filters a time-frequency representation of the data by replacing blocks with the median properties of those nearby (i.e., neighbours), thereby discarding variable elements, and keeping repetitive ones such as the microbarom. This resulting matrix is then used as a background reference in a square-root-based Wiener filter, retaining only the event and wind noise elements. One of the key parameters is to specify the number of nearest neighbours being aggregated. Given that some of the Bogoslof eruptive events have durations of several hours, it is necessary in these cases to make this parameter quite large in order to ignore such events. The consequence, however, is that processing slows down considerably (Section 5.1.1). Another limitation is that for periods of persistently high winds, this method may attenuate some of the wind noise, not just the microbarom.

3.2.5. Spectral gating (SG)

Noise gating is the practice of only passing signals through a system that are above a particular amplitude. For this technique, we use the implementation of Sainburg and Amr (2019). Similar to BCseis, this method operates by masking trace elements that have lower power frequency components than thresholds based on the statistics of the designated noise-only section. A filter smooths and scales the mask before being applied to the data. Such a mask is termed soft, as values can range from 0 to 1. In a hard/binary mask, values are 0 or 1, such as the BCseis case. The aforementioned Wiener filters also create soft masks. The largely Gaussian distribution of the microbarom amplitudes allow for Gaussian statistical thresholds.

In prior work (Sanderson *et al.*, 2020b), we explored the impact of several unsupervised algorithms on microbarom and wind noise for RTM applications. This included spectral gating, but by treating the entire trace as the reference from which to determine the average noise from, rather than a manually identified region as here. Other unsupervised techniques assessed by Sanderson *et al.* (2020b) included spectral subtraction, adaptive foreground detection, and frequency-sorted NMF. These techniques performed inconsistently for the wide range of source-station distances and station hardware, with limited benefit to RTM detection and location. Consequently, we have focused more here on semi-supervised approaches, which can benefit from the specific nature of each station.

3.3. Wind noise removal

After the preprocessed trace has had the microbarom removed by any of the five methods in Section 3.2, a second decomposition step separates out the wind noise in stage 3 (Fig. 2). Stage 3 uses NMF, regardless of the stage 2 technique. This selection follows from finding that the non-NMF stage 2 methods are relatively ineffective at removing wind noise while leaving the events intact. In part this is due to limited training options, but also that threshold-based methods that use high wind noise as a reference, will inevitably remove relatively low amplitude events. Here we take a fully supervised approach where we label an example wind section, as well as the event section using eruption catalogues. This approach is appropriate for effective separation as volcanic jetting (i.e., explosions like jet engines) and wind noise are

often superficially similar in terms of waveforms and spectral content (e.g., Matoza *et al.*, 2011a). The effectiveness of the wind noise removal also depends on how similar wind characteristics are at different times of day, as the amplitude and frequency range can change. Selecting a region of high amplitude wind noise generally ensures higher noise reduction levels. Scaling the amplitude of the basis vectors has the effect of adjusting the degree to which wind noise is removed. For simplicity, here we do not perform any such scaling. Notably, we find that a semi-supervised approach of only labelling the event can slightly improve wind noise reduction when compared to when both the event and wind are selected, but sometimes this is at the expense of event fidelity. Also, by not including a wind label, the option of basis vector scaling is unavailable. A semi-supervised approach of only labelling wind performs less well than the fully supervised case, doubling the wind noise amplitude in the foreground of stage 3 (results not shown). This latter approach may be useful in automated classification systems, where the characteristics of the microbarom and wind noise are generally well established, but the nature of events are not. Here, we are focusing on the benefits of denoising using as much prior knowledge as possible, leaving integration into a classification system to future work.

The full signal separation sequence is illustrated in Fig. 3. The 31 January 2017 Bogoslof eruptive event was chosen as an example given the low amplitude activity hidden by the microbarom, transition from subaqueous explosions to subaerial jetting behaviour (Fee *et al.*, 2020), and varying spectral overlap between the event and multiple wind bursts. In stage 2, the stage 1 preprocessed trace (Fig. 3, first row) is operated on using NMF to separate out the microbarom background from any remaining arrivals in the foreground (Fig. 3, second and third rows). The blue-line bracketed region in the top plot is used as a relatively clean microbarom sample. Next, we separate the stage 2 foreground into a new foreground-background pair (stage 3), containing predominantly event and wind noise components respectively (Fig. 3, fourth and fifth rows). The yellow and green bracket regions on the third row of Fig. 3 are those designated as representative of event and wind noise characteristics for this second NMF process. The decomposition strategies are not perfect, however, and desired or undesired components may or may not pass between stages, depending on how similar they are to each other. For instance, the example in Fig. 3

shows effective microbarom isolation and removal in stage 2, whereas some wind noise remains in the foreground by stage 3. In part this is due to the frequency content of the wind at the start, middle, and end of the day varying considerably. Consequently, additional processing may be able to remove the very low frequency wind noise components that coincide with the event. However, this could be challenging given the similarity between the non-jetting parts of the event, which include very low frequencies, and this midday wind noise. What we consider wind here is in fact somewhat speculative given the timing relative to the eruption, and dissimilar frequencies to wind later in the day. However, the semi-coherent arrivals have apparent velocities on the order of 10 m/s, with beamforming suggesting arrivals are from the northwest and southeast, and so an eruption related phenomenon seems unlikely. Although gravity waves may have such velocities, we do not further speculate on such a source here. Variable wind noise spectra are not unprecedented, given that different kinds of turbulence produce different spectral characteristics (e.g., Walker and Hedlin, 2010; Cook *et al.*, 2021). Spectral analysis of the infrasound suggests that a volcanic jetting episode occurs during the final midday wind burst, ending at ~13:09. This would extend the catalogued infrasound activity by an hour, aligning more with the seismic end time (13:30) provided by Fee *et al.* (2020). For reference, minimal processing of the original trace (e.g., filtering > 0.1 Hz), produces a largely unchanged result, except for removal of the midday wind, and some impulses earlier in the day.

In Fig. 4(a), we compare the first-order performance of each stage of the signal separation methods, using infrasound records of the 31 January 2017 Bogoslof eruption, recorded at OKIF. Each individual stage 2 method effectively removes the majority of the microbarom, and the common stage 3 (NMF) method significantly reduces the wind noise to low levels. The degree of wind noise reduction during stage 2, which aims to only remove microbaroms, varies slightly between methods. Minor deviations in the microbarom-dominated portion of the signal are primarily shown by the grey (stage 2) BC trace, which retains some impulsive components. For the event portion, NN performs relatively poorly on preserving the initial impulses. To indicate the processing combinations in this paper, where only stage 2 is performed, we use the relevant acronym, and if stage 3 is also performed, we use '+ NMF' as a suffix.

As it is not possible to know exactly what this event looks like noise-free, it is challenging to say which of the signal separation methods shown in Fig. 4(a) are best at preserving the original waveform from this illustration alone. As noted in the section introduction, the noisy/original data beam in Fig. 4(b) cannot be used as an indicator of the true signal, given that the microbaroms remain after beamforming. For the beam effective velocity and backazimuth, we use mean values from array processing results (Section 3.4) during the eruption (333 m/s and 347°). This backazimuth differs from the true direction (352°), likely due to deflection of the arrivals by crosswinds. Such deviations from true are common for many Bogoslof volcanic explosions, at OKIF, as well as at the other arrays. Fig. 4(b) does illustrate, however, that wind noise is reduced by beamforming without any additional denoising steps, but much more so when beamforming denoised traces. In Section 5, however, for a more impulsive and discrete Bogoslof explosion, we use metrics such as cross-correlation, phase changes, and the signal-to-distortion ratio to more systematically measure waveform alteration under different noise conditions.

Another useful processing evaluation metric is comparing the power spectral density (PSD) of each segment of the labelled data section at each stage of the processing. By doing this, we are able to establish whether signal separation is occurring as intended under different conditions. For example, Fig. 5 compares PSDs for two of the stage 2 and 3 combinations from Fig. 4, where we choose two widely contrasting examples for illustration. For the microbarom labelled time section (Figs 5a and d), both methods similarly reduce the microbarom peak from their original levels (stage 1, blue trace) to those in the stage 2 foreground (purple), when compared at ~0.2 Hz. This is by design to keep the methods generally comparable. However, the shape of the resulting spectra is quite different, with the sharper NN peak demonstrating higher attenuation away from the centre. Further, after applying NMF in stage 3, Fig. 5(a) shows a frequency split in the peak of the residual microbarom between the foregrounds (green) and background (yellow), whereas those in Fig. 5(d) are more similar. During the wind segment (Figs 5b and e), many profiles are similar, with differences resulting from the initial stage 2 separation. Here BC has a relatively sharp stage 2 background peak (red) compared to NN, meaning that the original wind noise (blue) is relatively unaltered in the stage 2 foreground (purple), except around 0.2 Hz. An ideal case is consistent removal of the

microbarom (and other persistent sources) through the trace, which neither method does a good job of here. In terms of the event portion we see that after stage 2, the BC background (Fig. 5c, red) has the desired shape (i.e., like that in Fig. 5a), indicating that the foreground (purple) has been adjusted appropriately. In contrast, for NN (Fig. 5f), overcompensation has occurred due to the broader profile microbarom extracted (red), particularly at frequencies < 0.1 Hz. Despite having removed most of the microbarom in stage 2, stage 3 inadvertently reduces the microbarom further, particularly in Fig. 5(c). To a first-order for this event, the BC, HR, and NMF methods have similar PSDs, with NN contrasting the most to these, and SG sharing characteristics of these endmembers. For other events and parameter combinations, however, results can vary. PSD analysis showed minimal differences between performing a single NMF operation to simultaneously separate microbaroms, wind noise, and events, and the two-step NMF + NMF approach described thus far. For consistency and ease of comparison with the other stage 2 approaches, we retain use of the two-step process.

3.4. Array processing

For array processing, we use the progressive multi-channel correlation software, PMCC (Cansi, 1995; Le Pichon *et al.*, 2010). PMCC searches in time-frequency space for coherent plane-wave arrivals that have similar wavefield parameters on multiple three element subarrays. As the three element combinations are of varying size, the array is sensitive to a wide range of frequencies, which reduces spatial aliasing. Where four element arrays such as OKIF have four subarrays and an aperture of ~ 100 m, the eight-element array, IS53, has 56 subarrays, and an aperture of ~ 2 km. PMCC groups detections of similar arrivals into families for which the corresponding time-frequency blocks are called pixels. Output wavefield parameters from the processing include backazimuth, apparent velocity, mean frequency, and maximum amplitude (\sim waveform peak to peak). PMCC configuration parameters are supplied in the Supporting Information, which includes a threshold to only detect acoustic arrivals (apparent velocities 250–500 m/s). Trace denoising occurs prior to running PMCC.

AVO infrasound array processing for the Bogoslof eruption was performed using a least-squares beamforming algorithm (Olson and Szuberla, 2005), incorporating the mean of the cross-correlation maximum (MCCM; e.g., Haney *et al.*, 2018; Lyons *et al.*, 2020) as a detection metric. Declaration of a detection is based on specific thresholds, which cannot be directly reproduced with PMCC. As mentioned in Section 2, use of variable parameters over time have led to slightly different published catalogues on the Bogoslof eruption (e.g., Coombs *et al.*, 2019; Lyons *et al.*, 2020). As we show in Section 4, given the range of acoustic sources around Alaska and often great source-receiver distances, it is possible that some detections are coincidental rather than eruption related, and equally that some eruptions can be hidden.

3.5. Backprojection

Backprojection, also known as RTM, describes a delay and stack grid-search approach for locating events. For each trial source location on a grid, appropriate sections of waveform envelopes from the surrounding stations are extracted according to a velocity model, then aligned and stacked. To take some account of the different atmospheric paths, our simple backprojection scheme performs the spatial grid search for a range of fixed celerities (range/time, 250–350 m/s) and then retains the highest amplitude from all stacks at each time step to form a detector function (DF). The grid node which gives the highest amplitude in the DF is typically assumed to be the event source. Details of the basic data preprocessing, detection, and location methodologies can be found in Sanderson *et al.* (2020a). Here we use the ‘time-summed DF’ approach from that study, which involves a running summation window applied to each stack before the DF is calculated. This tends to improve location accuracy and event classification rates, at the expense of event detection sensitivity.

For regional source location, backprojection of direct infrasound is an uncommonly applied tool when compared to backazimuth-yielding methods. In part this is due to the rarity until recently of sufficiently dense infrasound networks with single stations, such as the TA. Backprojection using arrays is also possible, however (e.g., Shani-Kadmiel *et al.*, 2018). Air-to-ground coupling recorded on seismometers has also been used as a proxy for infrasound in backprojection schemes (e.g., Walker *et al.*, 2010).

For RTM, we will address two topics. Firstly, whether denoising traces prior to stacking ('pre-stack') is beneficial to event detection and location. Secondly, we will compare linear stacking to three advanced stacking techniques from seismology, namely: (1) phase-weighted stacking (Schimmel and Paulssen, 1997), (2) robust stacking (Pavlis and Vernon, 2010), and (3) selective stacking (Yang *et al.*, 2020). We refer to these methods as 'co-stack' signal enhancement strategies. Each of these methods, though differing in implementation, ultimately penalize traces that do not share common waveform elements, and thus noise is reduced. We explored some further beamform/stacking-based methods such as pure-state filtering (e.g., Samson and Olson, 1981), auto-covariance filtering (Nakata *et al.*, 2015), and a technique using singular value decomposition-based Wiener filters (Moreau *et al.*, 2017). The computational time for these approaches, however, was not balanced by the performance level for the Bogoslof volcano dataset. Nth-root stacking, semblance, and F-statistic function stacking were previously found to not be advantageous versus linear stacking for Bogoslof volcanic data (Sanderson *et al.*, 2020a).

Data are processed in the same way as for array processing, i.e., the workflow in Fig. 2. For each trace, we label event sections according to catalogue onsets and expected delay times to each station, based upon a mean celerity derived from record section inspection. These steps keep with the intention of focusing on the degree to which trace-based denoising improves RTM results, rather than incorporating an automated feature classification scheme. Sample microbarom and wind noise sections are picked manually.

4. Results

To illustrate some of the capabilities of the denoising processes, we use both array processing and RTM detection and location approaches. We focus on the 15 January, 31 January, and 8 March 2017 Bogoslof volcanic eruptions, as well as on a longer period covering a series of explosions.

4.1. Array processing

4.1.1. 31 January 2017 Bogoslof volcano eruption

In Fig. 6 we show PMCC processing results for the 31 January event, and compare the results of the original (noisy) data with denoised data using BC + NMF. Fig. 6(a) shows 24 hours of beamed data from OKIF and a summary of PMCC detections, plotted as backazimuth versus time versus frequency. In the original data, the microbarom presents as largely continuous detections to the southeast. The BC processing removes the large majority of the microbarom detections, particularly the earlier section, which had a slightly lower amplitude than the reference section (Fig. 3). Scaling the plotted microbarom detections by amplitude would reduce the clutter further.

Another persistent signal migrates from $\sim 67^\circ$ to $\sim 62^\circ$ over the course of the day (the ~ 5 – 8 Hz detections in Fig. 6a). Amplitudes of this component are very small (< 0.01 Pa) as well as broadband (up to 20 Hz when processed with a 100 Hz sampling rate). Apparent velocities are in the acoustic range. High-frequency arrivals often point to anthropogenic sources. However, at OKIF, through January, such arrivals are often seen on the edges of the azimuthally dispersed microbarom detection clusters. At Dillingham, there is a more gradual grading to higher frequencies away from the microbarom centre, with these marginal detections usually having lower quality detection statistics. Post-processing thresholds can help filter these components. Other arrays have few anomalies of this type. Consequently, this feature is likely an artefact, rather than a natural phenomenon, although the cause is uncertain. Changing the PMCC configuration, number of sensors, sensor positions, or the preprocessing approach do not noticeably affect the result. Frequency-wavenumber analysis (Capon, 1969), an alternate array processing method, shows no such detections, suggesting a PMCC-based origin.

Figure 6(b) shows a close-up of only those detections likely associated with the eruption. The denoised data has more PMCC detections, as well as lowering the minimum frequency range of those detections into that previously occupied by the microbarom. The mean frequency of the families is also lower as a consequence. It should be noted that the lower corner frequency of the OKIF instruments was

0.1 Hz, limiting correct representation of < 0.1 Hz arrivals despite deconvolution of the instrument response. The azimuthal distribution of the denoised arrivals is slightly more dispersed than in the original data, potentially due to distortion during processing rather than reflecting improved accuracy. Notable features from results for the other methods include NN + NMF further extending the detection frequency range (at times to 0.01 Hz), whereas NMF + NMF doesn't appear to affect the frequency range.

The summary of the PMCC results for the different denoising techniques in Fig. 6(c) shows some significant variation in terms of detection of the different activity. As noted for Fig. 6(a), BC effectively removes much of the microbarom (as does SG) and related artefacts (which SG preserves). For the other methods, the number of microbaroms detected is not greatly reduced, even though the amplitude of the microbaroms are reduced. As the PMCC threshold parameters do not include amplitude, the lack of microbaroms detections for BC and SG indicate that such arrivals are not simply made sufficiently small, but rather the waveforms are made incoherent via denoising. Compared to the results for the noisy data, all denoising methods increase the numbers of families for the eruption. NN increases eruption detections the most, by $\sim 24.5\%$ versus the original data.

In Fig. 6(a), gaps occurring in the microbarom align with increases in wind noise. These gaps include the period $\sim 12:00$ – $13:00$, speculated as wind in Section 3.3. We find that if stage 3 is performed without stage 2, there is a minimal (4%) uptick in detected microbaroms. This is an important result as it suggests limited value for array processing by removing wind noise in this way, i.e., few new detections will be found. The picture is more complex when combining stage 2 and 3, however, with microbarom families decreasing by an additional 20% during stage 3 for BC + NMF, for example, and increasing by 7% for SG + NMF (Fig. 6c). Equally Fig. 6c shows some minor ($< 5\%$) variation for event detections between stage 2 and 3.

4.1.2. Source triangulation

To assess denoising performance on event location using multiple arrays, we apply an automated cross-bearings approach with the IMS-vASC algorithm (Matoza *et al.*, 2017). IMS-vASC uses a grid search to tally the number of intersecting backazimuths over time. There are two output layers which we use here. Firstly, the during layer, which is the map of total intersecting pixels for the day of the event in question. Secondly, the cleaned layer, which retains only those areas of the during layer that have a specific ratio (here, > 1.5) when compared to the average pixel counts in those areas for the days preceding the eruption. The idea of the cleaned layer is to remove persistent background sources like microbaroms, thereby retaining prominent transient events. In Fig. 7, we compare two events using original and BC-denoised data. BC provides a compromise between strong microbarom reduction and moderate increases in event detection (Fig. 6). The 31 January event uses two prior days for the cleaned grid calculation, and the 8 March event uses one prior day. These short prior durations are not ideal to provide background averages but are constrained by the rapid movement of the dominant microbarom sources. During these prior periods, gaps in array detections are common due to wind noise. Consequently, the estimates of the background noise sources are not particularly robust here. As the prior days do not contain any events, it is not possible to apply stage 3 processing to the data, and so this step isn't applied to the during-layer results either.

For the 31 January event (Figs 7a–f), all five AVO arrays have clear PMCC detections, with OKIF detecting the most (Fig. 6). Consequently, the event location is constrained accurately, but a nearby microbarom source dominates the mapped results for the unprocessed data in the during layer (Fig. 7a), as well as the cleaned layer (Fig. 7b). As the event and microbarom do not share a common backazimuth at any of the arrays, separating these sources in backazimuth-frequency space is straightforward (Fig. 7c). For the BC data, however, the microbarom is much reduced for the during layer (Fig. 7d), and further still in the cleaned layer (Fig. 7e). Now the Bogoslof eruption is the only prominent source. This provides an example where the denoising process provides an improvement over layer cleaning, useful in cases where there are limited prior data. Results for the noisy data can be improved significantly if the minimum number

of pixels to count a station in intersection is set to 100 instead of 0 (as used in Fig. 7), while also requiring each station to contribute (instead of at least two). Detections by all five arrays is a rare occurrence, however, and so such a constrained strategy cannot generally be relied upon. As crosswinds variously deviate arrivals from their true backazimuth, we use a 5° tolerance. For reference, Coombs *et al.* (2019) report only OKIF and SDPI detecting this event, whereas Lyons *et al.* (2020) list AKS, CLCO, and OKIF.

For the 8 March event (Fig. 7g–l), only three AVO arrays have clear PMCC detections (AKS, DLL, OKIF). SDPI has some detections coincident with Bogoslof volcano during the event window, but these are removed during denoising, so it's uncertain from this analysis alone whether these arrivals are event related. As the primary microbarom source in this example shares a backazimuth with Bogoslof volcano for most arrays, there is typically an overlap in backazimuths between event and microbarom detections. Coupled with less detecting arrays, the source location is less precise (e.g., Fig. 7g versus 7a). Following layer cleaning, the noisy data provides a narrowed-down source location (Fig. 7h). The scatter plot in Fig. 7(i) shows the backazimuth overlap for the microbarom and event at the AKS array. For the BC data, the microbarom source in the during layer is slightly reduced (Fig. 7j), but otherwise there is not a great change apparent. For the cleaned layer, however (Fig. 7k), there is only a modest reduction in background elements. This implies that many areas from the during layer had ratios > 1.5 compared to the prior layer for the BC data. Incidentally, if the ratio is changed from 1.5 to 3, then no event is detected at Bogoslof volcano for the noisy-data clean layer, but still is for BC, suggesting the denoised data are a more robust result. This is also reflected by limiting the during period to the event itself, rather than the whole day, where for the denoised data, the maximum overlapping pixel count around Bogoslof Island is $\sim 8\%$ higher than without denoising (when comparing cleaned layers). In terms of backazimuth versus frequency, the event is clearer following denoising, and with an extended lower frequency range (Figs 7i and l). For this event Coombs *et al.* (2019) noted the same detecting arrays as above (including SDPI), while Lyons *et al.* (2020) listed AKS and OKIF.

4.1.3. Multi-event analysis

A longer-term example using DLL array data illustrates some of the benefits of denoising more explicitly. Fig. 8 shows how several Bogoslof volcanic eruptions become much clearer in the array processing results once microbaroms are reduced. This example uses microbarom removal processing only, without any explicit wind noise reduction. Three of the eruptions in Fig. 8, one on 14 December 2016, and two on 12 January 2017, did not meet the published cataloguing criteria for infrasound detection (Coombs *et al.*, 2019; Lyons *et al.*, 2020). The first two of these events are detectable using PMCC, however, with less confidence for the third. Some eruptions, however, become less detectable with processing, such as on 16 and 29 December 2016. In terms of the methods compared in Fig. 8, SG shows the events most clearly by reducing microbarom detections and amplitudes of those that remain. BC also reduces such detections, but the amplitudes are less affected. This is due to BC leaving residual microbarom spikes in the foreground, despite having similar time-average PSD peak values as the other methods. Although using NMF in stage 3 would reduce these spikes (Fig. 3), it's not possible to apply stage 3 consistently to those days where events didn't occur (~40% of Fig. 8). If soft rather than hard thresholding is used (Section 3.2.1), then such spikes are also reduced. However, there are more microbarom detections at the thresholding levels needed to maintain spectral amplitude equivalence between methods. As such, soft thresholding was not pursued.

Notably, a group of detections at ~18:22 15 December 2016 has a backazimuth similar to that of an eruption on the day prior. The SDPI array also detected these arrivals ~25 minutes earlier, with a backazimuth also coincident with Bogoslof volcano. This time delay is reasonable if Bogoslof volcano was the source. Nearer arrays to Bogoslof Island such as OKIF and AKS (Fig. 1) do not show corresponding detections, although shadow zones with high transmission loss in the 300 km around Bogoslof Island were typical of the time (Schwaiger *et al.*, 2020). There is, however, no other known evidence that this was an eruption.

To more explicitly quantify the performance of denoising methods, we develop a simple ranking system based on PMCC analysis of 15 events from the Bogoslof eruption sequence. These events were

those found from inspection of Okmok data that had at least 25 PMCC detections, with little to no azimuthal overlap from microbaroms. For each method and event, a rank from one to six is assigned based on the relative method performance in 10 statistical categories listed below, with a rank of one being optimal. Rank values are totalled to provide a singular value for the method-event combination (Table 2). An overall score is calculated for each method by then summing across the metrics. For the following categories, each is applicable to detections within $\pm 16^\circ$ of the true backazimuth during the catalogued eruption period, except for 9 and 10, which apply to clutter at all other backazimuths. Numbers 3-9 represent median values. Categories include: (1) number of detections, (2) proportion of eruption duration that contained detections, (3) inter-sensor correlation, (4) F-statistic, (5) event amplitude, (6) standard deviation of backazimuths, (7) minimum frequency, (8) number of detections with a mean frequency < 0.3 Hz, (9) clutter amplitude, and (10) number of clutter detections. These categories broadly cover outcomes from denoising processing that may be of specific interest depending on the use-case, and generally would indicate effectiveness of the denoising. Statistics for unprocessed data are also provided for comparison.

The results in Table 2 show that although the HR method demonstrates the best overall performance using this analysis system, HR is not the strongest in all categories. For instance, NN excels at enhancing low frequency detections and reducing the amplitude of clutter. On the other hand, NN has poor scores for preserving the event amplitude, or detecting arrivals throughout the eruptions, and has more dispersed backazimuth estimates. As illustrated in Fig. 6, BC removes the greatest amount of clutter, but the amplitude of that which remains is still high. The NMF and SG methods have the lowest overall scores from the denoising methods, but still outperform the noisy data in most areas. NMF is the strongest method for detecting arrivals throughout the catalogued period, but the total number of detections is relatively low. For reference, the values in the main rows of Table 2 have a possible range of 15–90 (e.g., a rank of three for each of the 15 events will give a total of 45). Whereas there is considerable variety in performance within and between methods when comparing statistical categories, the overall totals for the methods are somewhat similar given the possible summed range across the 10 categories of 150–900. This reflects the relative

strengths and weaknesses of the methods balancing out. Use of median values, rather than summations, when compiling Table 2, does not meaningfully impact these results.

4.2. Backprojection

We now turn our attention to backprojection, and whether removing the microbarom and wind noise elements from traces can improve detection and location capabilities. We focus on the 15 January 2017 Bogoslof eruption due to its complex nature with four main pulses, and traces typically having poor signal to noise ratios, with clear waveforms only at AV.MSW (located 82 km east of Bogoslof Island) and DLL. This event was entirely submarine, as indicated by the relatively low mean frequency content compared to that of 31 January 2017 (Fee *et al.*, 2020).

For the base denoising method, we choose the HR approach as it one of the fastest to compute (Section 5.1) and has high overall performance (e.g., Table 2). As each technique produces similar envelopes to a first-order (e.g., Fig. 4), we anticipate that the following results are representative of the other systems (verified with the SG method). We also assess the effects of changing the lower frequency limit from 0.01 Hz to 0.1, 0.2, and 0.35 Hz. For this event, we focus on discussing results from the linear and robust stacking methods, as phase-weighted stacking and selective stacking performed relatively poorly regardless of several parameter and processing combinations. For other events, relative performance may differ (Sanderson *et al.*, 2020b).

Generally, we find that denoising helps improve the SNR of the 15 January 2017 event (Fig. 9). Removing the microbarom only provides a modest improvement (results not shown), whereas incorporating wind noise removal provides a more noticeable, if sometimes subtle effect. For this latter case, denoising tends to become less impactful as the lower frequency limit increases, indicating that high-pass filtering out wind-noise components can have a net benefit despite removal of some of the event elements. We also tend to find a greater SNR improvement from high-pass filtering (i.e., 0.01 Hz noisy versus 0.35 Hz noisy) than by denoising (0.01 Hz noisy versus 0.01 Hz denoised), as Fig. 9 illustrates. Similar results are found for events on 22 December 2016 and 8 March 2017, using stations within 2,000 km of Bogoslof Island. These

assessments are based on overall trends when also evaluating three different RTM preprocessing strategies, two of which involve different kinds of automatic gain control (referred to as AGC1, and AGC2; Sanderson *et al.*, 2020a). AGC has the effect of regularizing signal amplitudes, and so can enhance weak event arrivals, but can also increase weak windy sections instead. Consequently, DF shapes may not directly reflect those of the original waveforms recorded. There are some parameter combinations that go against the overall trends, such as AGC2 performing best when the lower frequency limit is 0.1 Hz. AGC2 is also the only method that resolves the second event pulse on 15 January, although the overall SNR is poorer than Fig. 9 shows.

Evaluating the sharpness of the DF peaks is also important, as location estimates are performed on the samples making up the tops of each peak. If the samples happen to represent grid nodes far from the source, then the mislocation value (source distance error) will be high. In this respect, the ≥ 0.35 Hz results are better than those for ≥ 0.01 Hz data, as are denoised compared to noisy data. Location accuracy and peak sharpness is also affected by the azimuthal coverage of detecting stations. Map slices of the stack results in Fig. 9(c) show the distribution of the stations within a 1,250 km radius around Bogoslof Island, which are biased to the east. Only a subset of these stations contributes positively to the stacks, as indicated by the shading of each station (fewer arrivals occur beyond 1,250 km). These map slices correspond to the first pulse in the eruption, marked by arrows in Fig. 9(b). This pulse varies in amplitude and shape between parameter configurations, similarly reflected in map form (Fig. 9c). The smallest and least prominent pulse does locate in the Bogoslof Island area, but the energy is more distributed than the other cases. For the noisy versus denoised maps at ≥ 0.35 Hz, the former case is less sharp overall, but there is less smearing of the highest values. Ring features in these maps indicate a dominant impact by the SDPI array at the ring centre.

Our evaluation of the 15 January event suggests that the eruption onset occurred 13 minutes prior to that listed in the catalogue, i.e., at 06:27, rather than 06:40. This stems from aligning SG-denoised DLL infrasound waveforms and PMCC results with travel-time corrected seismic data (Fig. 10). The seismic stations, AV.OKER and AV.MAPS are ~ 55 km southwest and ~ 75 km southeast of Bogoslof Island respectively. Although the catalogue eruption start time is based on multiple types of data, this time matches

that given by the catalogue infrasound origin, which itself assumes a 340 m/s celerity (used for all events). Adjusting this celerity to 320 m/s, a best-fit value from the RTM results, would change the eruption onset by less than three minutes when also assuming a surface travel path, meaning that celerity alone cannot explain the onset time discrepancy. Our suggested revised eruption onset of 06:27 matches the reported explosion seismic onset at AV.OKER of ~06:27 (Searcy and Power, 2020), and slightly prior to that at AV.MAPS of ~06:33. Our assessment of the AV.MAPS data indicates an earlier onset similar to that of AV.OKER (Fig. 10). Prior to the eruption, precursory seismicity occurs (from 05:41), with discrete events merging into low amplitude tremor. Tepp and Haney (2019) marked the end of this precursory activity at 06:32, the same time used as the eruption seismic onset by Coombs *et al.* (2019) for the AVO catalogue. Thus, there is some uncertainty regarding the transition from precursory to eruptive behaviour using the seismicity alone, but integration of the infrasound analysed here points toward an eruption nearer to 06:27 than 06:40. Separately, as the catalogued infrasound duration is based on OKIF, which didn't record the eruption well, the listed duration is only 76 minutes. At DLL, however, PMCC shows the eruption continued until at least 12:11 (accounting for travel time), approaching six hours from the onset (but still prior to the catalogued seismic end time of 12:48). Given the low amplitude of the start and end of the eruption, as well as azimuthally coincident microbaroms, denoising helps with accurately identifying these features in the data.

5. Discussion

Infrasound is becoming an increasingly established method for volcano monitoring at local, regional, and global scales. Microbaroms are largely an inescapable aspect of such recordings (e.g., Le Pichon *et al.*, 2005; Matoza *et al.*, 2013; Tailpied *et al.*, 2013; Matoza *et al.*, 2019), and such clutter may inhibit clear detection of eruptions when viewed as waveforms, spectrograms, or in array processing results. For chemical blasts with yields of ~0.1–10 kT, expected frequency spectra overlap with the microbarom band (e.g., Green and Bowers, 2010). This is a potential issue for the IMS infrasound component, which is designed with such activity in mind. Wind noise also manifests in infrasound recordings, and may disrupt

coherent records of signals of interest, depending on the wind speed. In this paper we present a preliminary investigation of several open source denoising tools and techniques as applied to the Bogoslof volcano shallow-submarine eruption sequence, whose predominantly low infrasound frequencies overlapped both wind noise and the microbarom. Although these styles of eruption are relatively uncommon, remote monitoring involves greater attenuation of higher frequencies with distance, such that denoising techniques have wide applicability. In the following, we review our main findings, and explore some subtleties which may affect denoising performance.

5.1. Microbarom clutter

We find that all the stage 2 methods increase the number of PMCC detections associated with eruptions, particularly where the mean family frequency overlaps the microbarom (Fig. 6). PMCC results also show that even where the amplitude of the microbarom is reduced to similar levels by the different methods (Fig. 4), there can be along-trace variability during the event and wind sections (Fig. 5). Additionally, the microbarom detectability in PMCC can be quite variable. For instance, whereas all methods affect the amplitude of detections, BC and SB perform best in terms of reducing microbarom detections entirely (Fig. 6c), allowing events of interest to be more easily distinguished for arrays where the microbarom shares a backazimuth with the volcano (Fig. 8). Backazimuth estimates typically disperse by a few degrees from the noisy data estimates, but given crosswind advection is typical, it is difficult to establish the relative accuracy. During the nine-month Bogoslof volcanic eruption, AKS, DLL, and IS53 were the only arrays with significant microbarom backazimuth conflicts. With more infrasound stations added since 2017, such conflicts have increased. Incidentally, PMCC does not always resolve similar concurrent signals arriving from different directions. Thus, for large, low frequency events like that of 8 March 2017, the microbarom has the appearance of stopping during the event, such that the event appears relatively clearly in time-series at the aforementioned arrays without any denoising. Our results also highlight several instances where the current AVO catalogue for Bogoslof volcano can be improved by: (1) extending the reported infrasound duration of some events, with that of 15 January 2017 for example

appearing to start ~13 minutes earlier than reported, and lasting several hours longer; (2) increasing the number of arrays that made detections, with, e.g., those on 14 December 2016 and 12 January 2017 detected at DLL despite no catalogued infrasound at any array; (3) potentially detecting additional events (e.g., 15 December 2016), though these currently lack supporting evidence. Section S2 of the Supporting Information provides additional examples. Care should be taken with interpretation, however, as the array processing of microbarom signals can result in artefacts that may be of similar frequency and backazimuth to eruptions. Such findings and event versus microbarom versus artefact discrimination are typically facilitated by denoising processing.

We also find that the denoising methods can sometimes decrease the detectability of events, along with the microbarom, thus potentially negatively impacting any event catalogues and/or duration estimates. Some examples of these events are shown in Fig. 8 for SG-processed data (16 and 29 December). Although these events may appear dubious as detectable at all given the dispersed backazimuths, there are distinct groups of detections that are separate from the microbarom (centred off figure at $\sim 270^\circ$) at the exact travel-time adjusted catalogue range. From closer inspection, the SG PMCC results show about 1/3 less detections versus the unprocessed data for these events, but more notably, the amplitudes are much smaller. These two events are very weak, and largely indistinguishable from background even after denoising. At least here, the BC approach has an advantage. Similar behaviour may occur for traces used during RTM processing. Other stage 2 approaches may also have particular weaknesses, such as those outlined for NN in Section 3.2.4. The nature of NN makes it more suitable for situations where events are short and wind is sporadic. Otherwise, the untrained nature of the technique makes it challenging to effectively remove only the microbarom in an effective and computationally efficient way.

To assess alteration to denoised signals under different noise conditions, we begin by taking a high SNR Bogoslof volcanic signal from OKIF at $\sim 02:38$ 20 February 2017 as a reference, and bury it in progressively higher amplitude microbaroms, then assess how well that contamination is removed. The microbarom sample is taken from $\sim 06:12$ the same day and scaled between 1 and 100 times its original amplitude. This type of event has its peak frequency in the microbarom range, presenting a strong denoising

challenge. The results in Fig. 11(a) are based on comparing the reference and the denoised result during the marked 30 s event section. For methods which require a noise reference, we use a section 15–45 s prior to the event. For SNR calculations, we use a section 15–45 s after the event. For some metrics such as the correlation coefficient, each method performs relatively similarly, with values only dropping below 0.8 where the original SNR was < 4 dB. In terms of maximum amplitude, there is more variety, with most deviations occurring where the original SNR is < 8 dB. In terms of SNR for denoised traces, NMF provides little advantage for this event (unlike that in Fig. 4), with NN improving SNR the most. Finally, we present scale-invariant signal-to-distortion ratio (SI-SDR) values (Le Roux *et al.*, 2018), where the distortion is essentially the residual between the original signal and the denoised version. Here, results are generally similar, except for SG, and are affected by the aforementioned amplitude change. We find that for this short event, particularly at higher noise levels, the denoising can be sensitive to spectral windowing parameters, the amount of data before and after the event, the microbarom section used as an overlay, and how that overlay aligns with the reference waveform. Higher denoised-SNR values are seen for HR in particular if the duration of the part labelled as noise is increased. Such factors may help explain differences in post-denoising SNR between the results in Fig. 11, and those in Fig. 4, where the SNR values are similar. The Fig. 4 event is one of the 15 day-long traces from Table 2, where post-denoising SNR is very similar for most methods. In terms of the other metrics from Fig. 11a, for the Table 2 events, the BC and HR methods generally perform best. However, the assessment is more challenging given there is no clean reference trace available. In Appendix A, we repeat the same tests for three synthetic event signals that have both similar, and dissimilar frequency content to the microbarom. Denoising performance is enhanced in the latter case, particularly at lower initial SNR. Again, there are post-denoising SNR differences between methods. Certainly, inconsistent or unpredictable denoising performance at different timescales is not a desirable trait, with NMF suffering the most in this aspect.

To get an initial insight into why the BC and SG approaches reduce microbarom clutter to a much higher degree compared to HR, NMF, and NN (Fig. 6), we examine the distortion of the microbarom in the stage 2 foreground (Figs 11b and c). This is relevant as simply reducing the amplitude of the microbarom

will not significantly change the number of detections. There appear to be three primary factors: (1) SG processed traces have relatively high changes to the inter-channel delay times when compared to the noisy data—these several sample adjustments can mean that the apparent velocity PMCC calculates is no longer in the acceptable acoustic range, and so a pixel is not generated; (2) BC and SG processed channels show relatively low correlation and more distorted components versus the noisy data equivalents, suggesting that traces are less likely to meet PMCC wavefield parameter thresholds; (3) BC data often has a low correlation between its own traces. Not all microbarom packets exhibit this behaviour during processing, however, with yet to be determined factors dictating how microbaroms are split between foreground and background.

5.1.1. Parameter selection

An important aspect of choosing a stage 2 processing method is computational efficiency, particularly for real time applications and/or processing large datasets. The operation of BCseis requires choosing a spectral resolution parameter, which has a considerable effect on the computation time. We find little first-order waveform difference between settings (e.g., 4 and 64 voices), but that time-frequency representations of one voice data are noticeably coarse. Consequently, for analysis here, we use four-voice data, which remains practical for large data sets. The following run times and memory use values are those required to process 24 hours of 20 sample per second data with a single core 3.1 GHz processor. For the Fourier transform based methods, here we use 2,048 FFT points, and 75% overlap of 2,048 sample windows. BC: 213 s, 12.5 GB; HR: 7.5 s, 0.3 GB; NMF: 5.5 s, 0.55 GB; NN: 258 s, 0.9 GB; SG: 2.6 s, 0.1 GB. As mentioned in Section 3.2.4, for NN, speed can be increased by reducing the number of nearest-neighbours, appropriate for cases where events and noise are shorter in duration than the cases presented here. The five wavelet options included with BCseis can also make a noticeable difference to results. For instance, when using a Shannon, rather than the Morlet wavelet, ~9% more detections are found for the event in Fig. 6, including during some of the gaps between pulses. Also, for Fig. 8, events on 23 December 2016 and 22 January 2017 are noticeably clearer with the Shannon wavelet. However, waveform distortion was relatively high for several test events and so not pursued here. So far we have used the same spectral

resolution parameters between methods to allow for direct comparison, but this may not always be ideal. For example, NN has a more even along-trace foreground/background separation with 50% overlap. Generally, however, NN results appear somewhat insensitive to FFT-parameters, as do those for NMF. In comparison, HR and SG results can vary more drastically. In Fig. 11 for instance, we use 512 FFT points, and a 480 sample overlap of 512 sample windows, in order to suit these two methods, as well as the relatively short labelled noise sections. Such parameters provide clearly worse SNR when using SG for the event in Fig. 4. Further, if windows are made 128 samples for Fig. 11, the SNR of denoised traces generally improves for all applicable methods, but at the expense of higher distortion levels. Evidently, the time-frequency resolution in the Fourier-based techniques can be critical, with the controlling parameters challenging to apply consistently with predictable outcomes. NMF performance can be improved on a case by case basis by adjusting supervision options. However, this is another example of performance inconsistency, which is undesirable when trying to select a single method for general use.

5.2. Wind noise

We find that wind noise reduction has little effect on improving signal or clutter detectability in array processing results (Fig. 6). Whereas PMCC can detect signals with $\text{SNR} < 1$, at high wind speeds infrasound arrivals become very poorly correlated and hard to resolve. For RTM, wind noise removal similarly affects SNR and mislocation when using robust or linear stacking. For both stacking types, high-pass filtering provides a greater net benefit, however. Microbarom removal has a relatively small impact here but may be more important where most stations have very low initial SNR. The RTM processing parameters are primarily tuned for noisy linear stacking, whereas other parameters may work better depending on the configuration. Similarly, the effects of tuning parameters for the stacking methods could be explored. AGC to some extent undoes the wind noise reduction, but results are still better with AGC, than without. As with stage 2 processing, there is a wide variation on the computation times for stacking methods, with the python-based phase-weighted stacking implementation taking 230 times longer than linear for each stacking operation. Although the former approach performed poorly here, and could not

justify the computation time, improvements may be seen when using a realistic propagation model to improve the alignment of arrivals and thus SNR (e.g., Shani-Kadmiel *et al.*, 2018).

A potential factor in how well stage 3 denoising works is that as the SNR and frequency content of signals typically decrease with distance, event characteristics may change relative to noise, depending on propagation conditions, the local site environment (vegetation, topography, snow), and station design. We analyse the 8 March 2017 Bogoslof volcanic eruption, which was one of the best spatially recorded, with arrivals evident at 36 different stations within a 2,500 km radius. Of these, we use 25 stations which have relatively low winds coincident with the event. Fig. 12(a) shows some contrasting arrivals at stations at increasing distance from Bogoslof volcano, with progressive attenuation of higher frequencies with distance generally evident. OKIF, being the closest array, shows the event frequency range extending to at least 0.05 Hz. The wind noise at the start and end of the day is mostly < 0.1 Hz, but does extend higher. Wind noise at the DLL array, however, generally does not exhibit these higher frequencies, with the TA.TCOL station being similar in this regard. These two stations are some of the few sites in this study that are in dense tree cover, with the wind noise reduction pipes at DLL also playing a part. Nearby stations retain the higher wind noise frequencies seen at OKIF, and to a greater extent at TA.L27K, which is much more typical for TA stations. TA.L27K has a relatively poor wind noise reduction levels compared to the other stations shown, with relatively similar event and wind spectra.

As NMF depends on the spectral content of the training examples, we hypothesize that dissimilar wind noise and event characteristics will produce better wind noise reduction in stage 3. As metrics, first, we implement the Frequency Index (FI) concept of Buurman and West (2006) as follows: $FI = \log_{10}(A_{upper}/A_{lower})$, where A_{lower} is the mean spectral amplitude from 0.01–0.25 Hz, and A_{upper} is the same for 0.25–1.0 Hz. This division uses the microbarom centre as the division point. In the following, eFI refers to the FI value for the event sample, and wFI for the wind noise sample. Fig. 12(b) shows that eFI generally decreases with distance indicating attenuation of higher frequencies, with wFI displaying some of the variation seen in Fig. 12(a). To measure how well NMF affects SNR, we compare changes in event to wind noise amplitude ratios. For this we define a second metric, $RR = (e_3/w_3)/(e_2/w_2)$, where e and w are the root-

mean-square amplitudes of event and wind noise samples, and 2 and 3 are stage 2 and 3 of the processing. Higher RR values indicate better wind noise reduction. In terms of whether RR depends on eFI/wFI , i.e., how similar the event and wind noise spectra are, there isn't a clear picture. For instance, RR does vary for $eFI/wFI < 1$, but not above. RR values are generally higher for arrays than for non-array stations, likely due to the closer source proximity and presence of more sophisticated wind noise reduction systems. Some variation for metrics exists within arrays, reflecting slight differences between elements, even at 100 m apertures. A caveat to eFI -dependent results in Fig. 12(b), and more generally, is that when using 0.01 Hz as a lower filter bound, many stations will have some degree of wind noise during the event (e.g., three of four stations in Fig. 12a). This presence could make signal separation harder if events and wind noise are similar. For the culled events in Fig. 12, however, noise reduction is generally better if using 0.01 Hz rather than 0.05 or 0.1 Hz. For these high-passed results, there is much less RR differentiation between stations.

A caveat to the analysis of Fig. 12 is that the SNR of the unprocessed data varies between stations, which, as Fig. 11 shows, is a factor influencing the SNR of denoised signals. In Appendix A, we evaluate stage 3 noise reduction for the same synthetic event signals as for stage 2, using two sources of wind noise that each do, and do not overlap with the events. Briefly, these results indicate improved denoising performance for: (1) non-overlapping events and wind noise, (2) wind noise samples from an array rather than TA station, and (3) events that are relatively high frequency compared to the wind noise. Alternative frequency comparison metrics, such as the spectral angle, may prove useful as an alternative to the FI scheme used here. As for the stage 2 denoising, in future work we aim to expand the range of events and recording conditions analysed with such methods in order to understand their best method of application. Particularly, being able to use available information about path and site effects to help select a specific denoising scheme, or configuration, would increase the impact and portability of these kinds of tools.

6. Conclusions

By evaluating five different trace-denoising methods, we find that persistent background clutter such as the microbarom can be reduced effectively in infrasound data from Bogoslof volcano without any

data labelling, but more consistently with such training. Microbarom reduction primarily aids progressive multi-channel correlation (PMCC) array-processing by expanding the signal detection frequency range, and also limiting surrounding clutter in cases of a shared backazimuth. Although the harmonic regeneration (HR) method offers a leading overall combination of benefits in terms of event enhancement and noise suppression performance given the computation time, each of the five methods tested here has important strengths and weaknesses that make them more suitable for particular objectives. For instance, all methods can decrease the microbarom amplitude, but the block choice (BC) and spectral gating (SG) implementations also make the remaining microbaroms relatively undetectable by PMCC (desirable if inadvertent). The nearest neighbour (NN) algorithm excels at low frequency detections, with non-negative matrix factorisation (NMF) best at detecting arrivals throughout the eruption. Consequently, selection of one or more of the presented approaches can be tailored toward a specific analysis task, data property, and workflow. The range of analyses, impacts, and considerations presented here help to develop the limited existing literature on infrasound denoising.

Separating wind noise from events with NMF is a relatively challenging task. For this objective, good separation at most Transportable Array (TA) stations is only possible by labelling the event, and optionally wind noise data. Training solely on the wind noise typically provides only a moderate improvement over the original signal. At Alaska Volcano Observatory (AVO) and International Monitoring System arrays, the wind noise character typically differs from the events significantly due to hardware and/or source proximity, and as such, separating signals from wind noise is easier. It is likely that some training is a fundamental aspect of a general wind noise reduction scheme, as unsupervised foreground/background separation techniques are not well suited to this task. Wind noise reduction explored here does not appreciably affect array processing results, however. In contrast, reverse time migration (RTM) benefits mainly from wind noise reduction due to the greater trace amplitude modulation than comes from microbarom removal. The impact is more apparent when other gain control processing is avoided, but this may be less beneficial overall.

We also find that PMCC can be a useful tool for helping to refine existing Bogoslof volcanic eruption catalogues in that PMCC clearly detects infrasound from several catalogued events previously reported to have no infrasound records, or otherwise for specific arrays. The onsets and duration of some events also warrant re-evaluation based on our observations, which can be aided via denoising.

In this study we focus largely on the extent to which denoising can improve signal-to-noise ratios (SNR) when using semi-supervised and fully supervised processing. We implement open source denoising methods that can be applied to a single trace. This approach promotes accessibility and flexibility, as not all datasets will contain the thousands of records potentially necessary to produce reliable dictionaries. Infrasound noise reduction on individual waveforms can likely improve results from many analysis types aside from detection and location strategies, and should be an area of future study for the community. Even for simply viewing waveforms and spectrograms, microbarom (or microseism) removal is generally useful by increasing fidelity, but particularly benefits identification and presentation of weakly detected events at remote distances. More broadly, trace-based noise reduction can increase the value of isolated infrasound sensors in places where an array, or noise reducing hardware such as a wind dome, is impractical.

7. Future directions

Developing a ‘one scheme fits all data sets’ trace-level signal enhancement strategy is a challenging undertaking due to potentially high variability in signal and noise character, even at the same infrasound station over the space of a few hours. For a diverse region like Alaska, developing dictionaries accounting for the wide range of source, path, and site conditions across would be a central issue. Certainly, excluding the event aspect reduces the complexity of the objective, in that training only with noise samples permits the cleaned data to contain both volcanic and non-volcanic signals. Incorporating adaptive data processing (included with, e.g., HR) can be an important aspect of real-time monitoring, where the data may be in small chunks, aggregated, or otherwise not fit existing dictionaries. Such real-time use is affected by processing runtime, and travel-time-latency, which for infrasound can be, e.g., 15 mins at a range of 300 km. We suggest a combination of stage 2, 3, and 4 type approaches that sufficiently balance speed and

effectiveness in reducing noise while preserving signals. For RTM, it is computationally more efficient to enhance SNR during pre-stack processing than during stacking itself, as each denoising operation is only performed once per trace, rather than per grid node. There are more grid nodes than traces in a typical scenario.

Importantly, configurations of the denoising schemes as presented here are chosen with a bias toward preserving similar event waveform amplitude across methods, at the expense of further noise reduction. Above the applied ~78% microbarom reduction level, the NN method would be especially impacted. An initial assessment using the 31 January 2017 event suggests that from the categories in Table 2, that only the duration proportion, correlation, and F-statistic are largely unaffected by changing denoising level. Using a similar ranking system to Table 2, we find that above a 70% microbarom reduction, the average benefits become progressively limited. However, this figure disguises some important trends, which are outlined in the Supporting Information, along with the methodology. This information also shows that some methods have upper and lower limits on the amount of noise they can remove. In particular, the HR method is restricted to reducing the microbarom to between ~80 and 85%. A key future task would be to further explore how results change depending on denoising level using more events, as well as identify other influential factors and potential code modifications. Assessing how array processing algorithms other than PMCC can benefit from denoising individual traces prior to beamforming is also of interest.

Although we describe and present results from several denoising methods here, and have evaluated many more (e.g., Sanderson *et al.*, 2020b), this work only scratches the surface of the field. For instance, Müller (2015) explained, how in music, approaches should vary for cases of multiple instrument separation, harmonic/percussive separation, and voice separation. Though not volcanic-acoustic examples, they demonstrate that different techniques may be needed depending on the nature of the data, even without any noise present. At the geophysics cutting edge, diverse machine learning techniques are helping to rapidly advance the fields of signal classification and signal denoising, a combination of which will enhance automated volcano event detection and location practices. A popular core technique for both areas is the artificial neural network, with implementations using various degrees of supervision and training (e.g., Zhu

et al., 2019; Jiang *et al.*, 2020; Novoselov *et al.*, 2020; Son *et al.*, 2020). Fortunately, open source tools are becoming more available for the community to implement and further develop, with some applied to volcanic seismic data such as tsfresh (Christ *et al.*, 2018; Dempsey *et al.*, 2020), PICOSS (Bueno *et al.*, 2020), and those comprising the VULCAN.ears project (Moreno *et al.*, 2020).

Acknowledgments

We acknowledge and thank the Alaska Volcano Observatory for collecting and supplying detailed volcanic event information, as well as operators of the monitoring networks and data management centres this study benefited directly from. Data from the TA network were made freely available as part of the EarthScope USArray facility, operated by Incorporated Research Institutions for Seismology (IRIS) and supported by the National Science Foundation (NSF), under Cooperative Agreement EAR-1261681. Global Seismographic Network (GSN) is a cooperative scientific facility operated jointly by IRIS, the U.S. Geological Survey (USGS), and the NSF, under Cooperative Agreement EAR-1261681. The facilities of IRIS Data Services, and specifically the IRIS Data Management Center (DMC), were used for access to waveforms, related metadata, and/or derived products used in this study. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience and EarthScope (SAGE) Proposal of the NSF under Cooperative Agreement EAR-1261681. We thank C. Sens-Schönfelder, F. Storey, S.M. Mousavi, W. Thelen, C. Magirl, M. Coombs, T. Neal, T. Swift, A. Iezzi, and an anonymous reviewer, whose comments helped improve the manuscript. This work was funded by NSF grants EAR-1614855 and EAR-1614323. DF, MH, and JL acknowledge funding from the USGS Alaska Volcano Observatory. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data availability

For the study period, we downloaded TA, AEC, IMS, and GSN data from the IRIS DMC (<https://www.iris.edu/>, network codes TA, AK, IM, and II); AVO station data were obtained from the USGS

Winston Wave Server (<http://pubavol.wr.usgs.gov:16023/menu>, network code AV). Station location information, and instrument responses came from the IRIS DMC, but certain parameters have been adjusted, as detailed in the Supporting Information. The Supporting Information also includes comments on AVO catalogues for Bogoslof volcano, effects of adjusting the amount of microbarom clutter removed from the data, as well as tools used for data processing, analysis, and plotting.

References

- Adam, V., Chritin, V., Rossi, M. & Van Lancker, E., 1998. Infrasonic monitoring of snow-avalanche activity: What do we know and where do we go from here? *Ann. Glaciol.*, **26**, 324–328. <https://doi.org/10.3189/1998AoG26-1-324-328>
- Amante, C. & Eakins, B.W., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, data sources and analysis, *NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA*. <https://doi.org/10.7289/V5C8276M>
- Bednar, J.B., 1983. Applications of median filtering to deconvolution, pulse estimation, and statistical editing of seismic data, *Geophysics*, **48**, 1598–1610. <https://doi.org/10.1190/1.1441442>
- Boll, S., 1979. A spectral subtraction algorithm for suppression of acoustic noise in speech, *1979 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 200–203. <https://doi.org/10.1109/ICASSP.1979.1170696>
- Bowman, J.R., Baker, G.E. & Bahavar, M., 2005. Ambient infrasound noise, *Geophys. Res. Lett.*, **32**, L09803. <https://doi.org/10.1029/2005GL022486>
- Bryan, N.J., Sun, D. & Cho, E., 2013. See NMF software in the Supporting Information.
- Bueno, A., Zuccarello, L., Díaz-Moreno, A., Woollam, J., Titos, M., Benítez, C., Álvarez, I., Prudencio, J. & De Angelis, S., 2020. PICOSS: Python Interface for the Classification of Seismic Signals, *Comput. Geosci.*, **142**. <https://doi.org/10.1016/j.cageo.2020.104531>
- Busby, R.W. & Aderhold, K., 2020. The Alaska Transportable Array: As built, *Seismol. Res. Lett.*, **91**(6), 3017–3027. <https://doi.org/10.1785/0220200154>

- Buurman H. & West, M.E., 2006. Seismic precursors to volcanic explosions during the 2006 eruption of Augustine Volcano, in *The 2006 eruption of Augustine Volcano, Alaska*, pp. 41–57, ed. Power, J., Coombs, M. & Freymueller, J. <https://doi.org/10.3133/pp17692>
- Cabras, G., Canazza, S., Montessoro, P.L. & Rinaldo R., 2010. The restoration of single channel audio recordings based on non-negative matrix factorization and perceptual suppression rule, in *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, pp. 458–465, ed. Pomberger, H., Zotter, F., & Sontacchi, A.
- Cabras G., Carniel R. & Jones J., 2012. Non-negative Matrix Factorization: An application to Erta 'Ale volcano, Ethiopia, *Boll. Geofis. Teor. Appl.*, **53**(2), 231–242. <https://doi.org/10.4430/bgta0056>
- Cansi, Y., 1995. An automatic seismic event processing for detection and location: The P.M.C.C. method, *Geophys. Res. Lett.*, **22**(9), 1021–1024. <https://doi.org/10.1029/95GL00468>
- Capon, J., 1969. High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE.*, **57**(8), 1408–1418. <http://doi.org/10.1109/proc.1969.7278>
- Carniel, R., Cabras, G., Ichihara, M. & Takeo, M., 2014. Filtering wind in infrasound data by non-negative matrix factorization, *Seismol. Res. Lett.*, **85**(5), 1056–1062. <https://doi.org/10.1785/0220130142>
- Ceranna, L., Matoza, R., Hupe, P., Le Pichon, A. & Landès, M., 2019. Systematic array processing of a decade of global IMS infrasound data, in *Infrasound monitoring for atmospheric studies: challenges in middle-atmosphere dynamics and societal benefits*, pp. 471–482, ed. Le Pichon, A., Blanc, E. & Hauchecorne, A. Cham: Springer. https://doi.org/10.1007/978-3-319-75140-5_13
- Christ, M., Braun, N., Neuffer, J. & Kempa-Liehr, A.W., 2018. Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package), *Neurocomputing*, **307**, 72–77. <https://doi.org/10.1016/j.neucom.2018.03.067>
- Cook, M.R., Gee, K.L., Transtrum, M.K., Lympany, S.V. & Calton, M., 2021. Automatic classification and reduction of wind noise in spectral data, *JASA Express Letters*, **1**, 063602. <https://doi.org/10.1121/10.0005308>

- Coombs, M.L., Wallace, K., Cameron, C., Lyons, J.J., Wech, A., Angeli, K. & Cervelli, P., 2019. Overview, chronology, and impacts of the 2016–2017 eruption of Bogoslof volcano, Alaska, *Bull. Volc.*, **81**(11), 1–23. <https://doi.org/10.1007/s00445-019-1322-9>
- Coombs, M.L., Wech, A.G., Haney, M.M., Lyons, J.J., Schneider, D.J., Schwaiger, H.F., Wallace, K.L., Fee, D., Freymueller, J.T., Schaefer, J.R. & Tepp, G., 2018. Short-term forecasting and detection of explosions during the 2016–2017 eruption of Bogoslof volcano, Alaska, *Front. Earth Sci.*, **6**. <https://doi.org/10.3389/feart.2018.00122>
- Dempsey, D.E., Cronin, S.J., Mei, S. & Kempa-Liehr, A.W., 2020. Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand, *Nat. Commun.*, **11**(3562). <https://doi.org/10.1038/s41467-020-17375-2>
- Dixon, J.P., Stihler, S.D., Haney, M.M., Lyons, J.J., Ketner, D.M., Mulliken, K.M., Parker, T. & Power, J.A., 2019. Catalog of earthquake parameters and description of seismograph and infrasound stations at Alaskan volcanoes—January 1, 2013, through December 31, 2017, *U.S. Geological Survey Data Series*, 1115. <https://doi.org/10.3133/ds1115>
- Fee, D., Haney, M.M., Matoza, R.S., Van Eaton, A.R., Cervelli, P., Schneider, D.J. & Iezzi, A.M., 2017. Volcanic tremor and plume height hysteresis from Pavlof Volcano, Alaska, *Science*, **355**(6320), 45–48. <https://doi.org/10.1126/science.aah6108>
- Fee, D., Lyons, J., Haney, M., Wech, A., Waythomas, C.F., Diefenbach, A.K., *et al.*, 2020. Seismo-acoustic evidence for vent drying during shallow submarine eruptions at Bogoslof volcano, Alaska, *Bull. Volc.*, **82**(1), 1–14. <https://doi.org/10.1007/s00445-019-1326-5>
- Friedlander, F.G., 1946. The diffraction of sound pulses. I. Diffraction by a semi-infinite plane, *Proc. R. Soc. Lond. A*, **186**, 322–344. <http://doi.org/10.1098/rspa.1946.0046>
- Green, D.N. & Bowers, D., 2010. Estimating the detection capability of the International Monitoring System infrasound network. *J. Geophys. Res.*, **115**, D18116. <https://doi.org/10.1029/2010JD014017>

- Haney, M.M., Van Eaton, A.R., Lyons, J.J., Kramer, R.L., Fee, D. & Iezzi, A.M., 2018. Volcanic thunder from explosive eruptions at Bogoslof volcano, Alaska, *Geophys. Res. Lett.*, **45**, 3429–3435. <https://doi.org/10.1002/2017GL076911>
- Haney, M.M., Van Eaton, A.R., Lyons, J.J., Kramer, R.L., Fee, D., Iezzi, A.M., Dziak, R.P., Anderson, J.F., Johnson, J.B., Lapierre, J.L. & Stock, M., 2020. Characteristics of thunder and electromagnetic pulses from volcanic lightning at Bogoslof volcano, Alaska, *Bull. Volc.*, **82**(15). <https://doi.org/10.1007/s00445-019-1349-y>
- Ichihara, M., Yamakawa, K. & Muramatsu, D, 2021. A simple method to evaluate the air-to-ground coupling efficiency: a tool helping the assessment of seismic/infrasound energy partitioning during an eruption, *Earth Planets Space*, **73**(180). <https://doi.org/10.1186/s40623-021-01510-4>
- Jiang, J., Stankovic, L., Stankovic, V. & Pytharouli, S., 2020. *Automatic detection and classification of microseismic events from super-sauze landslide using convolutional neural networks*, abstract [S052_0009] presented at the 2020 Fall Meeting, AGU.
- Landès, M., Ceranna, L., Le Pichon, A. & Matoza, R.S., 2012. Localization of microbarom sources using the IMS infrasound network, *J. Geophys. Res.*, **117**, D06102. <https://doi.org/10.1029/2011JD016684>
- Langston, C.A. & Mousavi, S.M., 2019. Separating signal from noise and from other signal using nonlinear thresholding and scale-time windowing of continuous wavelet transforms, *Bull. Seism. Soc. Am.*, **109**(5), 1691–1700. <https://doi.org/10.1785/0120190073>
- Le Pichon, A., Blanc, E., Drob, D., Lambotte, S., Dessa, J.X., Lardy, M., *et al.*, 2005. Infrasound monitoring of volcanoes to probe high-altitude winds, *J. of Geophys. Res.: Atmos.*, **110**, D13106. <https://doi.org/10.1029/2004JD005587>
- Le Pichon, A., Matoza, R., Brachet, N. & Cansi, Y., 2010. Recent enhancements of the PMCC infrasound signal detector, *InfraMatics*, **26**.
- Le Roux, J., Wisdom, S., Erdogan, H. & Hershey, J., 2019. SDR – half-baked or well done? *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 626–630. <https://doi.org/10.1109/ICASSP.2019.8683855>

- Lyons, J.J., Haney, M.M., Fee, D., Wech, A.G. & Waythomas, C.F., 2019. Infrasound from giant bubbles during explosive submarine eruptions of Bogoslof volcano, Alaska, *Nat. Geosci.*, **12**(11), 952–958. <https://doi.org/10.1038/s41561-019-0461-0>
- Lyons, J.J., Iezzi, A.M., Fee, D., Schwaiger, H.F., Wech, A.G. & Haney, M.M., 2020. Infrasound generated by the 2016–17 shallow submarine eruption of Bogoslof volcano, Alaska, *Bull. Volc.*, **82**(19). <https://doi.org/10.1007/s00445-019-1355-0>
- Matoza, R.S. & Fee, D., 2014. Infrasonic component of volcano-seismic eruption tremor, *Geophys. Res. Lett.*, **41**, 1964–1970. <https://doi.org/10.1002/2014GL059301>
- Matoza, R., Fee, D., Green, D. & Mialle, P., 2019. Volcano infrasound and the International Monitoring System, in *Infrasound monitoring for atmospheric studies: challenges in middle-atmosphere dynamics and societal benefits*, pp 1023–1077, ed. Le Pichon, A., Blanc, E. & Hauchecorne, A. Cham: Springer. https://doi.org/10.1007/978-3-319-75140-5_33
- Matoza, R.S., Green, D.N., Le Pichon, A., Shearer, P.M., Fee, D., Mialle, P. & Ceranna, L., 2017. Automated detection and cataloguing of global explosive volcanism using the International Monitoring System infrasound network, *J. Geophys. Res.: Solid Earth*, **122**, 2946–2971. <https://doi.org/10.1002/2016JB013356>
- Matoza, R.S., Landès, M., Le Pichon, A., Ceranna, L. & Brown, D., 2013. Coherent ambient infrasound recorded by the International Monitoring System, *Geophys. Res. Lett.*, **40**. <https://doi.org/10.1029/2012GL054329>
- Matoza, R.S., Le Pichon, A., Vergoz, J., Herry, P., Lalande, J.-M., Lee, H.-I., *et al.*, 2011a. Infrasonic observations of the June 2009 Sarychev Peak eruption, Kuril Islands: Implications for infrasonic monitoring of remote explosive volcanism, *J. Volcanol. Geoth. Res.*, **200**, 35–48. <https://doi.org/10.1016/j.jvolgeores.2010.11.022>
- Matoza, R.S., Vergoz, J., Le Pichon, A., Ceranna, L., Green, D.N., Evers, L.G., *et al.*, 2011b. Long-range acoustic observations of the Eyjafjallajökull eruption, Iceland, April–May 2010, *Geophys. Res. Lett.*, **38**, L06308. <https://doi.org/10.1029/2011GL047019>

- McFee, B., Raffel, C., Liang, D., Ellis, D.P. W., McVicar, M., Battenberg, E. & Nieto, O., 2015. librosa: Audio and music signal analysis in python, *Proc. 14th Python in Science Conference*, 18–25. <https://doi.org/10.25080/Majora-7b98e3ed-003>
- Mialle, P., Brown, D., Arora, N. & colleagues from IDC, 2019. Advances in operational processing at the International Data Centre, in *Infrasound monitoring for atmospheric studies: challenges in middle-atmosphere dynamics and societal benefits*, pp. 209–248, ed. Le Pichon, A., Blanc, E. & Hauchecorne, A. Cham: Springer. https://doi.org/10.1007/978-3-319-75140-5_6
- Moreau, L., Stehly, L., Boué, P., Lu, Y., Larose, E. & Campillo, M., 2017. Improving ambient noise correlation functions with an SVD-based Wiener filter, *Geophys. J. Int.*, **211**(1), 418–426. <https://doi.org/10.1093/gji/ggx306>
- Moreno, G.C., Carniel, R., Lesage, P. & Mendoza Pérez, M.A., 2020. geoStudio & liveVSR - software supporting "VULCAN.ears" project (MSCA - IF - 20126, No.[749249]), *Zenodo*. <http://doi.org/10.5281/zenodo.4305100>
- Mousavi, S.M. & Langston, C.A., 2016a. Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding, *Bull. Seis. Soc. Am.*, **106**(4), 1380–1393. <https://doi.org/10.1785/0120150345>
- Mousavi, S.M. & Langston, C.A., 2016b. Adaptive noise estimation and suppression for improving microseismic event detection, *J. Appl. Geophys.*, **132**, 116–124. <http://dx.doi.org/10.1016/j.jappgeo.2016.06.008>
- Mousavi, S.M. & Langston, C.A., 2017. Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data, *Geophysics*, **82**(4), 211–227. <https://doi.org/10.1190/geo2016-0433.1>
- Müller, M., 2015. *Fundamentals of Music Processing – Audio, Analysis, Algorithms, Applications* (489 p.). Cham: Springer. <https://doi.org/10.1007/978-3-319-21945-5>

- Nakata, N., Chang, J.P., Lawrence, J.F. & Boué, P., 2015. Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *J. Geophys. Res.: Solid Earth*, **120**, 1159–1173. <https://doi.org/10.1029/2015JB011870>
- Novoselov, A., Balazs, P. & Bokelmann, G., 2020. *Separation and denoising of seismically- induced ground-motion signals with dual-path recurrent neural network architecture*, abstract [S052_0007] presented at the 2020 Fall Meeting, AGU.
- Olson J.V. & Szuberla, C.A.L., 2005. Distribution of wave packet sizes in microbarom wave trains observed in Alaska, *J. Acoust. Soc. Am.*, **117**, 1032–1037. <https://doi.org/10.1121/1.1854651>
- Pavlis, G.L. & Vernon F.L., 2010. Array processing of teleseismic body waves with the USArray, *Comput. Geosci.*, **36**, 910–920. <https://doi.org/10.1016/j.cageo.2009.10.008>
- Plapous, C., Marro, C. & Scalart, P., 2006. Improved signal-to-noise ratio estimation for speech enhancement, *IEEE Trans. Audio Speech Lang. Proc.*, **14**(6), 2098–2108. <https://doi.org/10.1109/TASL.2006.872621>
- Raspet, R., Webster, J. & Dillon, K., 2006. Framework for wind noise studies, *J. Acoust. Soc. Am.*, **199**, 834–843. <https://doi.org/10.1121/1.2146113>
- Sainburg, T. & Amr, K., 2019. See NoiseReduce software in the Supporting Information.
- Samson, J.C. & Olson, J.V., 1981. Data-adaptive polarization filters for multichannel geophysical data, *Geophysics*, **46**(10). <https://doi.org/10.1190/1.1441149>
- Sanderson, R.W., Matoza, R.S., Fee, D., Haney, M.M. & Lyons, J.J., 2020a. Remote detection and location of explosive volcanism in Alaska with the EarthScope Transportable Array, *J. Geophys. Res.: Solid Earth*, **125**, e2019JB018347. <https://doi.org/10.1029/2019JB018347>
- Sanderson, R.W., Matoza, R.S., Fee, D., Haney, M.M. & Lyons, J.J., 2020b. *Infrasonic backprojection with the EarthScope Transportable Array in Alaska: Improving detection and localization of explosive volcanism via noise reduction*, abstract [S001-0013] presented at the 2020 Fall Meeting, AGU.

- Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.*, **130**(2), 497–505. <https://doi.org/10.1111/j.1365-246X.1997.tb05664.x>
- Schwaiger, H.F., Lyons, J.J., Iezzi, A.M., Fee, D. & Haney, M.M., 2020. Evolving infrasound detections from Bogoslof volcano, Alaska: Insights from atmospheric propagation modeling, *Bull. Volc.*, **82**(27). <https://doi.org/10.1007/s00445-020-1360-3>
- Searcy, C.K. & Power, J.A., 2020. Seismic character and progression of explosive activity during the 2016–2017 eruption of Bogoslof volcano, Alaska, *Bull. Volc.*, **82**(1), 1–15. <https://doi.org/10.1007/s00445-019-1343-4>
- Shani-Kadmiel, S., Assink, J.D., Smets, P.S.M. & Evers, L.G., 2018. Seismoacoustic coupled signals from earthquakes in central Italy: Epicentral and secondary sources of infrasound, *Geophys. Res. Lett.*, **45**(1), 427–435. <https://doi.org/10.1002/2017GL076125>
- Son, S., Seydoux, L., de Hoop, M.V. & Campillo, M., 2020. *Detecting and clustering seismic signals with datasets from a single station located in southern California by unsupervised deep learning*, abstract [S052_0005] presented at the 2020 Fall Meeting, AGU.
- Tailpied, D., Le Pichon, A., Marchetti, E., Ripepe, M., Kallel, M., Ceranna, L. & Brachet, N., 2013. Remote Infrasound Monitoring of Mount Etna: Observed and predicted network detection capability, *InfraMatics*, **2**(1), 1–11. <https://doi.org/10.4236/inframatics.2013.21001>
- Tepp, G. & Haney, M.M., 2019. Comparison of short-term seismic precursors and explosion parameters during the 2016–2017 Bogoslof eruption, *Bull. Volc.*, **81**(63). <https://doi.org/10.1007/s00445-019-1323-8>
- Walker, K.T. & Hedlin, M.A.H., 2010. A review of wind-noise reduction methodologies, in *Infrasound monitoring for atmospheric studies*, pp. 141–182, ed. Le Pichon, A., Blanc, E. & Hauchecorne, A. Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-9508-5_5

- Walker, K.T., Hedlin, M.A.H., de Groot-Hedlin, C., Vergoz, J., Le Pichon, A. & Drob, D.P., 2010. Source location of the 19 February 2008 Oregon bolide using seismic networks and infrasound arrays, *J. Geophys. Res.: Solid Earth*, **115**, B12. <https://doi.org/10.1029/2010JB007863>
- Waythomas, C.F., Lyons, J.J., Fee, D. & Wallace, K.L., 2019. The 2016–2017 eruption of Bogoslof volcano, Alaska: preface to the special issue, *Bull. Volc.*, **81**(48). <https://doi.org/10.1007/s00445-019-1301-1>
- Wech, A., Tepp, G., Lyons, J. & Haney, M., 2018. Using earthquakes, T waves, and infrasound to investigate the eruption of Bogoslof volcano, Alaska, *Geophys. Res. Lett.*, **45**(14), 6918–6925. <https://doi.org/10.1029/2018GL078457>
- Williams, R.A., Perttu, A. & Taisne, B., 2020. Processing of volcano infrasound using film sound audio post-production techniques to improve signal detection via array processing, *Geosci. Lett.*, **7**(9). <https://doi.org/10.1186/s40562-020-00158-4>
- Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S. & Trujillo, J., 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. Seism. Soc. Am.*, **88**(1), 95–106. <https://doi.org/10.1785/BSSA0880010095>
- Yang, X., Bryan, J., Okubo, K., Jiang, C., Clements, T. & Denolle, M.A., 2020. Optimizing the stacking of noise correlation functions, submitted to *Geophys. J. Int.*
- Zhu, W., Mousavi, S.M. & Beroza, G.C., 2019. Seismic signal denoising and decomposition using deep neural networks, *IEEE Trans. Geosci. Remote Sens.*, **57**(11), 9476–9488. <https://doi.org/10.1109/TGRS.2019.2926772>

Appendix A: Synthetic noise reduction tests

This section follows from Fig. 11, which explores how well different denoising schemes perform depending on the signal to noise ratio of the original data. Details of the methodology are provided in Section 5.1. Here we combine synthetic signals with scaled microbaroms, as well as with scaled wind noise. The synthetic waveforms are: (1) a shocked pressure pulse (Friedlander, 1946), with a broad spectra that

peaks at 1 Hz; (2) a 0.2 Hz exponentially decaying signal; and (3) a 2.0 Hz exponentially decaying signal (Fig. A1). Amplitudes are scaled to ~ 7.9 Pa, similar to the event in Fig. 11. These signals are chosen to be similar to, and in contrast to, the peak frequency of the microbarom (~ 0.2 Hz). Similarly, these signals have varying degrees of spectral overlap with the two wind samples borrowed from Fig. 12. One wind sample is from AV.DLL, with the station using wind noise reduction pipes to reduce higher frequency noise. The other wind sample is from TA.L27K, which uses a simple diffuser, and retains more of the higher frequencies (Fig. A1). These wind samples are both scaled to have an initially similar amplitude to the microbarom sample, before all of the noise samples are scaled from 1 to 100 times in order to change the SNR. Two wind data formats are assessed: (1) Continuous wind samples that overlap with the synthetic events; and (2) wind samples that are zeroed in the middle so that there is no noise overlap with the event, where wind only occurs at the end of the traces. For the wind noise reduction, the event labelling window is set to start prior to the actual event onset (here by 13 s). This approach produces improved results when compared to having the window start right at the onset. This may relate to reduced tapering of the impulsive onsets of the events during processing. The expanded window does, however, enable calculations to incorporate any acausal distortion prior to the onset of the event. Due to the lower frequency of the wind compared to the microbarom, longer time windows are used for calculations.

For the microbarom removal evaluation in Fig. A2, we find generally similar results between the Friedlander pulse, and the 2 Hz signal. Both show decreasing statistical parameters as the original SNR decreases. The relative performances between methods are also similar for each event type. For the 0.2 Hz signal, which has a very similar frequency content to the microbarom, denoising performance is consistently worse than the other two cases, particularly where the original signal is below ~ 10 dB SNR. The results for the 0.2 Hz event are similar to those in Fig. 11, where the Bogoslof volcanic explosion was also predominantly low in frequency.

For the wind removal results in Fig. A3, there is a more complex picture than for the microbaroms. In terms of correlation, the discontinuous wind samples do not drop off, even at very low SNR. At 0.2 Hz and 2 Hz, the DLL continuous wind results decay at relatively high SNR compared to L27K. These results

are also reflected in the maximum amplitude plots, where the original amplitude is maintained in the discontinuous wind noise case. For the overlapping wind however, the maximum amplitude diverges with lower original SNR. In terms of the SNR of the denoised data, the DLL results are better than those for L27K, with the greatest difference at 0.2 Hz, though with all datasets having a net SNR improvement. This suggests that the wind-noise spectral differences between the two stations is impactful, with the lower frequency noise at DLL easier to remove. This type of result was also presented in Fig. 12. For the SI-SDR metrics, again we find better performance for the discontinuous wind samples, with little drop off even at low original SNR. The different synthetic events here produce somewhat similar results, but the 0.2 Hz signal has noticeably lower SI-SDR than in the other two cases, indicating higher distortion. This may be due to the dominant frequency being closest to the frequencies contained by the wind.

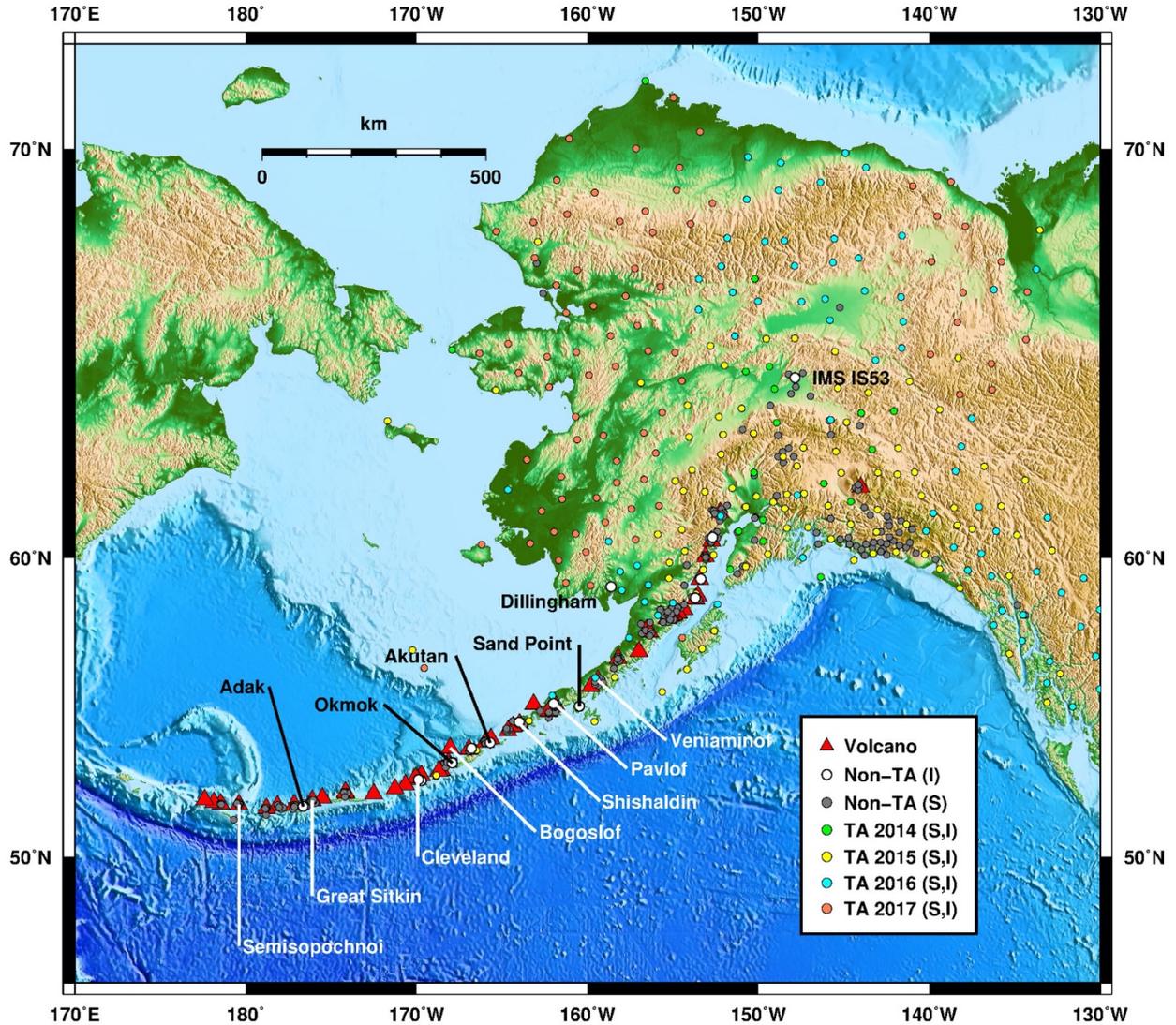


Figure 1. Locations of volcanoes with eruptions since 2016, and available seismic (S) and infrasound (I) stations that were operational during the course of the Bogoslof volcano eruption (2016–2017). Infrasound arrays that are referenced in this study are labelled in black, as well as volcanoes that have erupted since 2016 in white (Mount Cleveland volcano also has an infrasound array). Since 2020, several additional infrasound arrays have been installed by the Alaska Volcano Observatory in the Aleutian Islands. The quantity of TA stations varied through the eruption, with installations migrating westward. Some TA stations were removed entirely starting in 2021, with others adopted by existing regional networks. Topographic base-map data provided by the NOAA National Geophysical Data Center (Amante and Eakins, 2009).

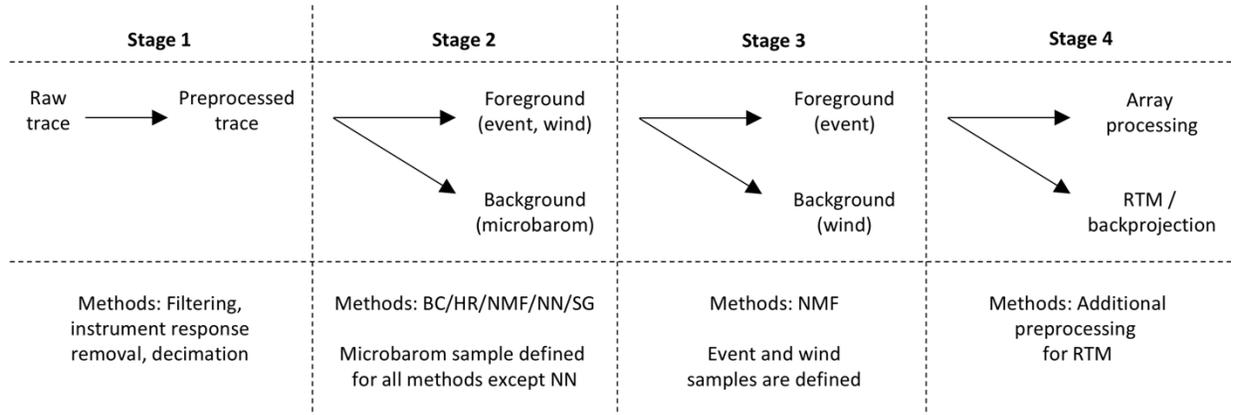


Figure 2. Summary of processing steps in denoising data prior to array processing and backprojection. Stage 1 is basic data preparation, stage 2 removes the microbarom, and stage 3 removes the wind noise. BC: BCseis; HR: harmonic regeneration; NMF: non-negative matrix factorisation; NN: nearest-neighbour filtering; SG: spectral gating.

Table 1. Explanation of acronyms that appear repeatedly in the paper. Additional acronyms defined in text.

Acronym	Definition	Category
ADKI	Adak infrasound array	Stations
AGC1	Automatic Gain Control method 1	Backprojection
AGC2	Automatic Gain Control method 2	Backprojection
AKS	Akutan infrasound array	Stations
AVO	Alaska Volcano Observatory	Stations
BC / BCseis	Block Choice seismic analysis	Denoising
CLCO	Cleveland volcano infrasound array	Stations
DF	Detector Function	Backprojection
DLL	Dillingham infrasound array	Stations
HR / HRNR	Harmonic Regeneration Noise Reduction	Denoising
IMS	International Monitoring System	Stations
IS53	IMS station IS53	Stations
NMF	Non-negative Matrix Factorisation	Denoising
NN	Nearest-Neighbour analysis	Denoising
OKIF	Okmok infrasound array	Stations
PMCC	Progressive Multi-Channel Correlation	Array processing
PSD	Power Spectral Density	Characteristics
RTM	Reverse Time Migration	Backprojection
SDPI	Sand Point infrasound array	Stations
SG	Spectral Gating	Denoising algorithms
SI-SDR	Scale-Invariant Signal to Distortion Ratio	Characteristics
SNR	Signal to Noise Ratio	Denoising algorithms
TA	EarthScope Transportable Array	Stations
UTC	Coordinated Universal Time	Reference time

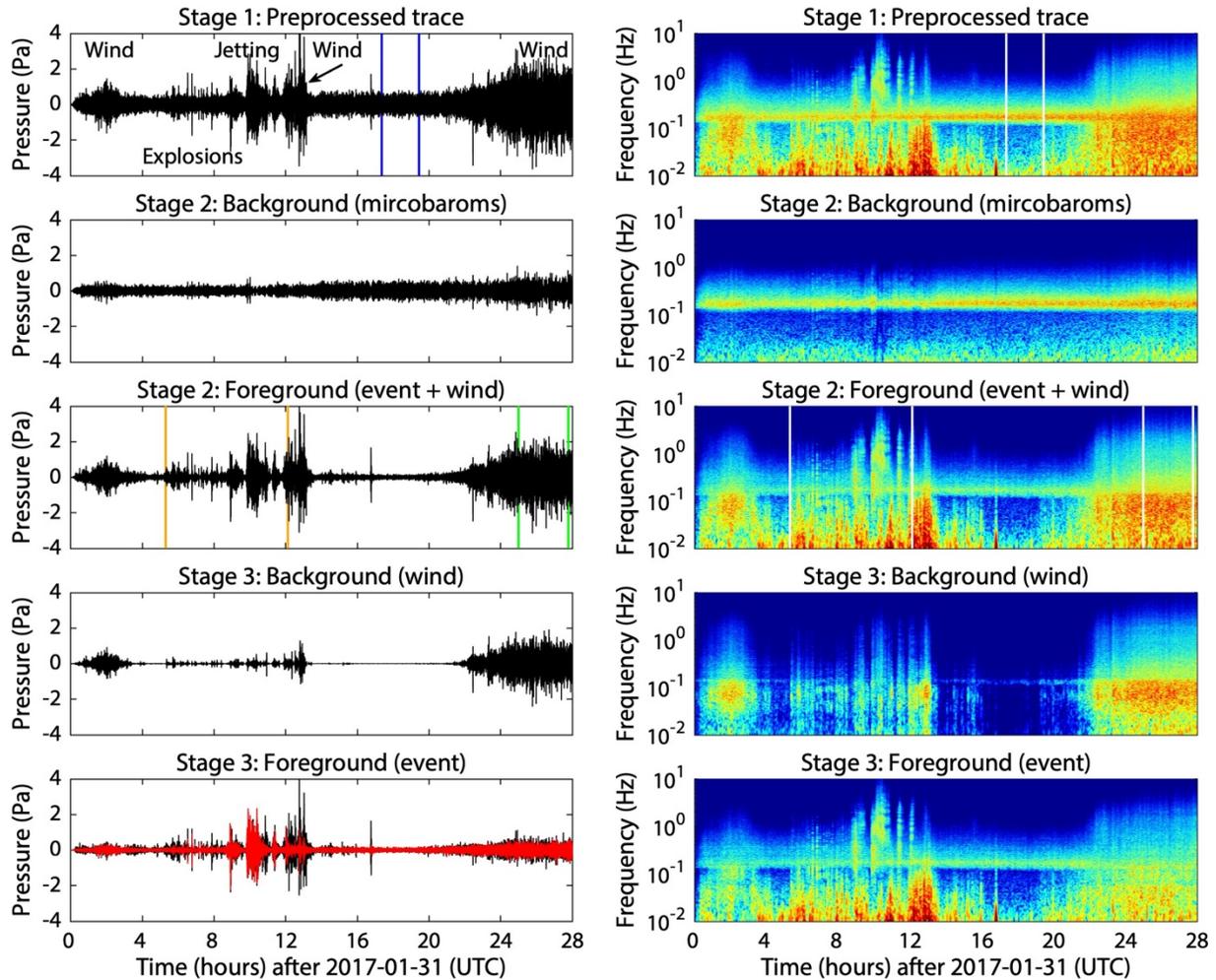


Figure 3. Denoising stages applied to infrasound records of the Bogoslof volcano eruption on 31 January 2017. Data are from OKIF channel 1. In this example we use NMF for both stage 2 and stage 3 separation steps. Data segments designated during processing are bracketed on the left-hand side on the appropriate traces as follows: microbarom (blue), event (orange), wind (green). The event timing is provided by the infrasound statistics in the AVO catalogue (travel time corrected). In the bottom panel, the red trace is the underlying black trace filtered > 0.1 Hz, removing some very low frequency wind elements. The righthand side shows the corresponding spectral content. Here, the brackets are in white only for clarity. Warmer colours in the spectrogram indicate relatively high power. The microbarom is effectively removed in stage 2, with much of the wind at the beginning and ends of the day removed in stage 3. Wind during the middle of the day largely remains as it differs in nature to the wind sample used as a reference.

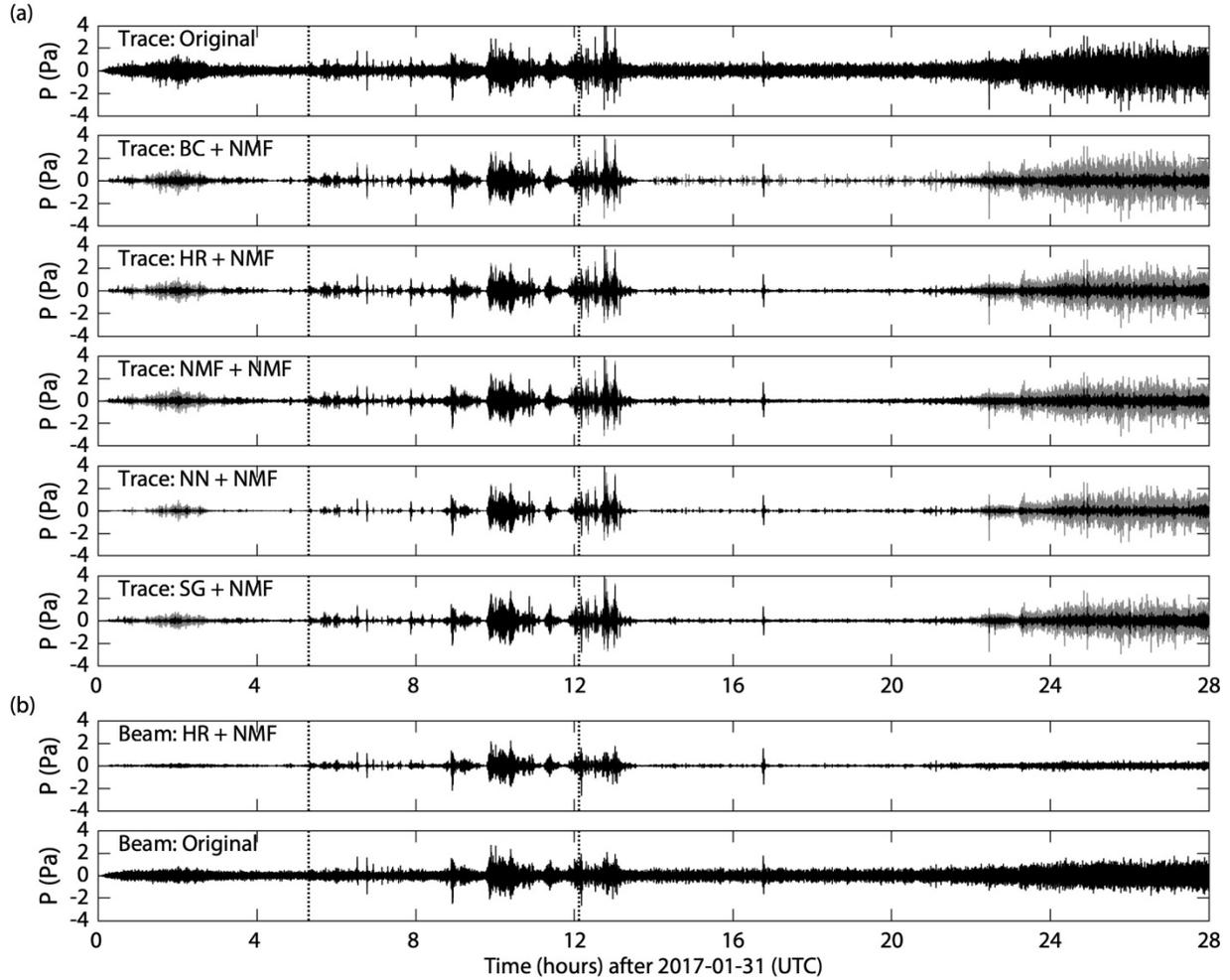


Figure 4. (a) Comparing first order differences between the original non-separated waveform (top), and each signal separation method (stage 2 = grey, stage 3 = black, P = pressure). Data are from the 31 January 2017 Bogoslof eruption, recorded at OKIF (channel 1). Dotted lines mark the duration of the infrasound from the event in the AVO catalogue (travel time corrected). The mean stage-2 SNR is 18.6 dB after denoising and 6 dB beforehand, representing a 310% increase in SNR. Individual stage-2 SNRs are within 1 dB of each other. The formula $\text{SNR}(\text{dB}) = 20 \cdot \log_{10}(\text{rms}_{\text{signal}}/\text{rms}_{\text{noise}})$ takes the root mean square (rms) of the marked microbarom period in Fig. 3 as a noise estimate, and the (noise contaminated) event section from 09:45–10:30 for the signal component. (b) Array beams for noisy traces, and for denoised (HR + NMF) traces. To a first order, resulting waveforms after stage 2, or stage 3, are similar between methods. Denoised beams have higher SNR than original beams as microbaroms do not destructively interfere.

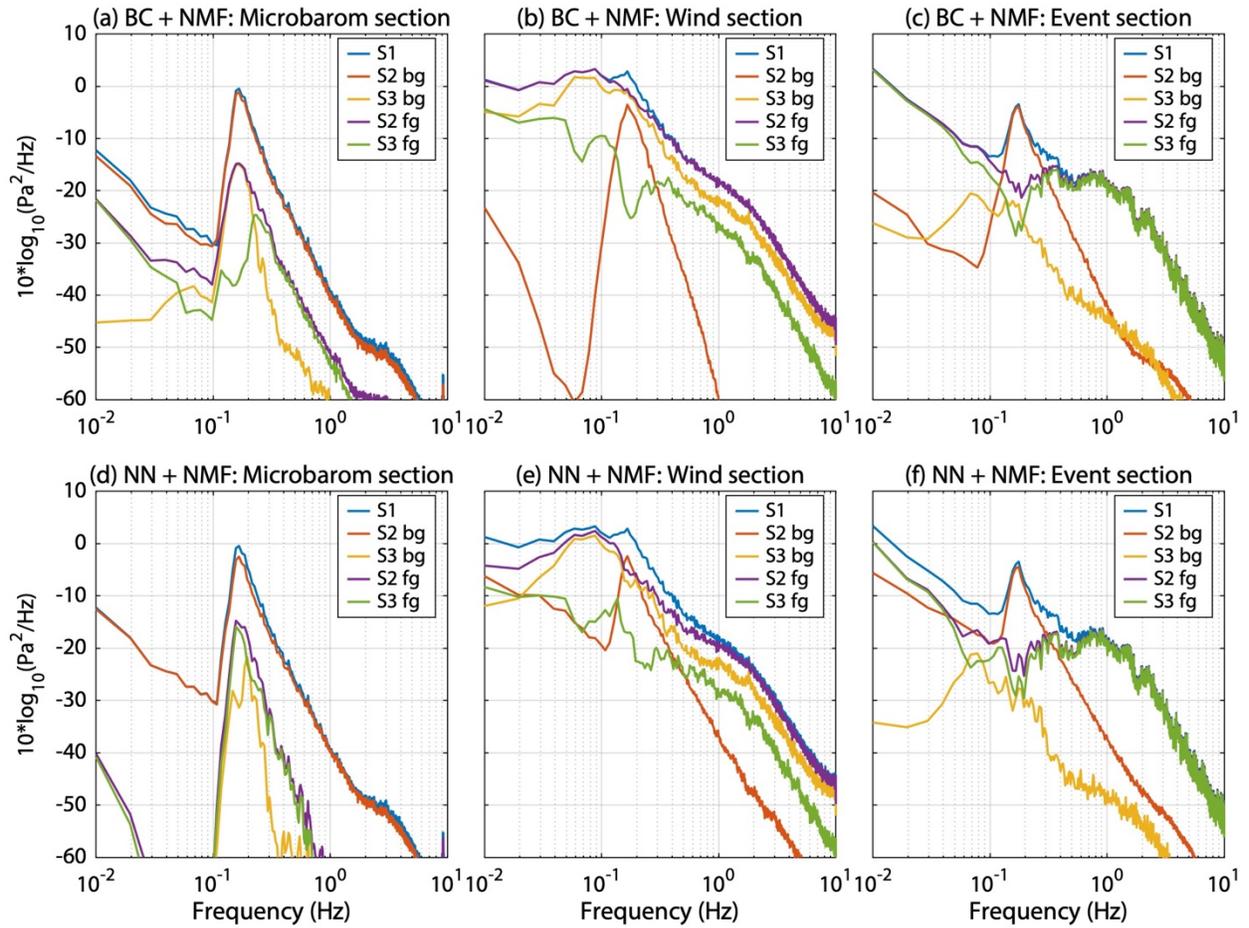


Figure 5. Examples of contrasting Power Spectral Density (PSD) estimates for labelled data segments at each processing stage (S). Fg and bg refer to the foreground and background respectively. (a–c) PSDs for BC in stage 2, and NMF in stage 3. (d–f) PSDs for NN in stage 2, and NMF in stage 3. Even though denoised traces using different methods may appear similar to a first order (Fig. 4), there may be subtler, but important differences in the frequency domain. Here we see for instance that microbarom removal in stage 2 is not evenly removed in different parts of the traces (inconsistently shaped red spectra), and also that residual microbaroms may be re-split during stage 3 (differing yellow and green spectra in panel a).

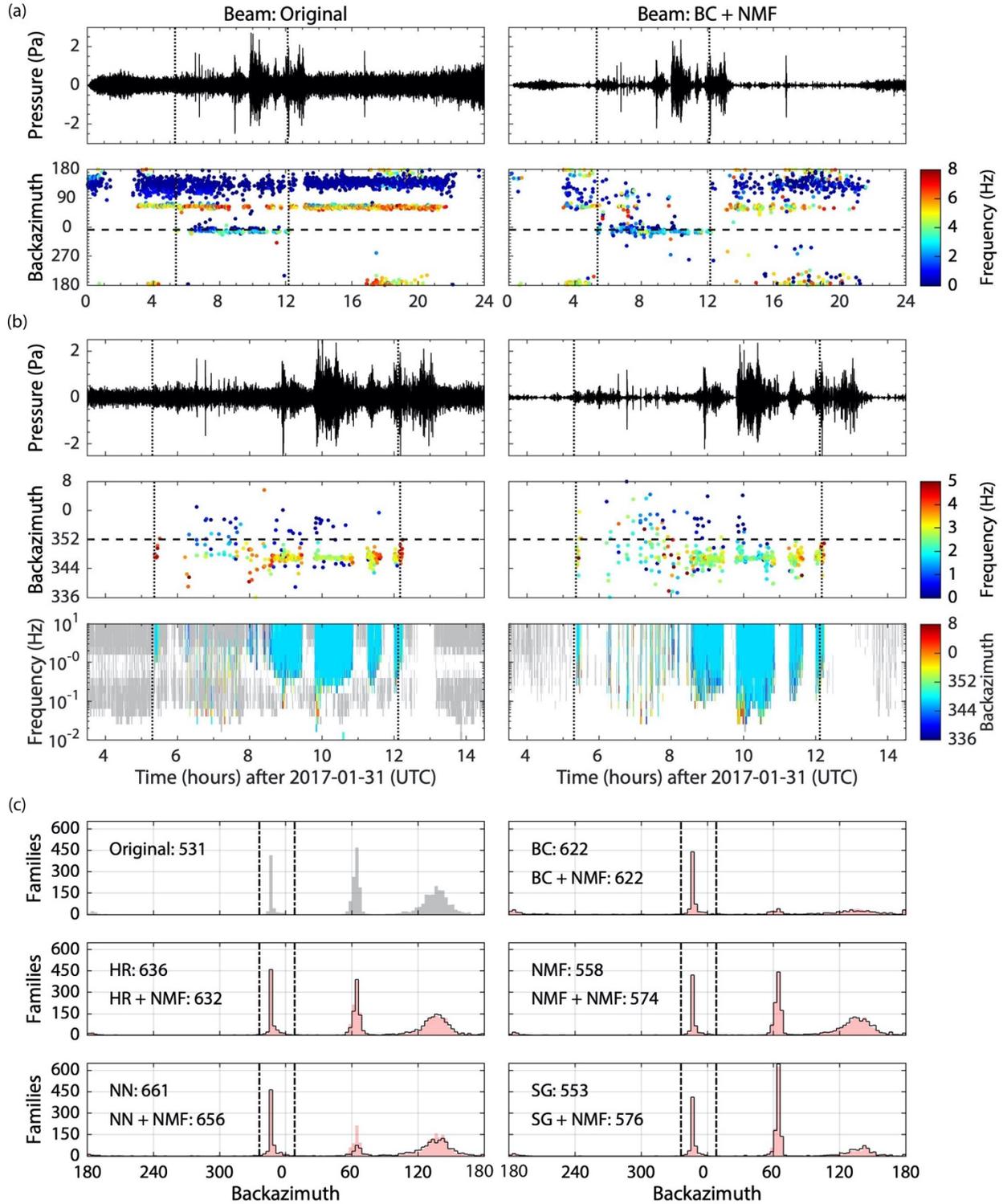


Figure 6. PMCC processing results for the 31 January 2017 Bogoslof volcano eruption, for noisy (left) and BC + NMF denoised data (right), recorded at OKIF. Denoising methods generally increase the number of event detections, and reduce both microbarom detections and microbarom amplitude. (a) 24-hour view of

results. Beams are aligned according to the means of the backazimuths and velocities from PMCC detections within $\pm 16^\circ$ of true (352°). Each point on the backazimuth versus time plot represents a PMCC family colour-coded by mean frequency. Vertical dotted lines demark the catalogued infrasound period (travel time corrected). The horizontal dashed line indicates the true backazimuth. (b) Close-up view of the eruption period. In the frequency versus time plots, pixels are colour-coded by frequency if the detections are $\pm 16^\circ$ of true. Pixels from other azimuths are grey, and white means no pixels. (c) Histograms showing OKIF PMCC detection statistics for the 31 January 2017 Bogoslof volcanic eruption for each denoising method combination (stage 2 shaded, stage 3 outlined). The first peak is for the eruption, the second peak is likely due to a processing artefact, and the third peak is for microbaroms. Vertical dashed lines demark $\pm 16^\circ$ limits around the true backazimuth to Bogoslof volcano from OKIF, which are used to calculate frequency totals provided in the subplot titles, and as per Fig. 6(b). Values shown are the number of families detected within $\pm 16^\circ$ of true.

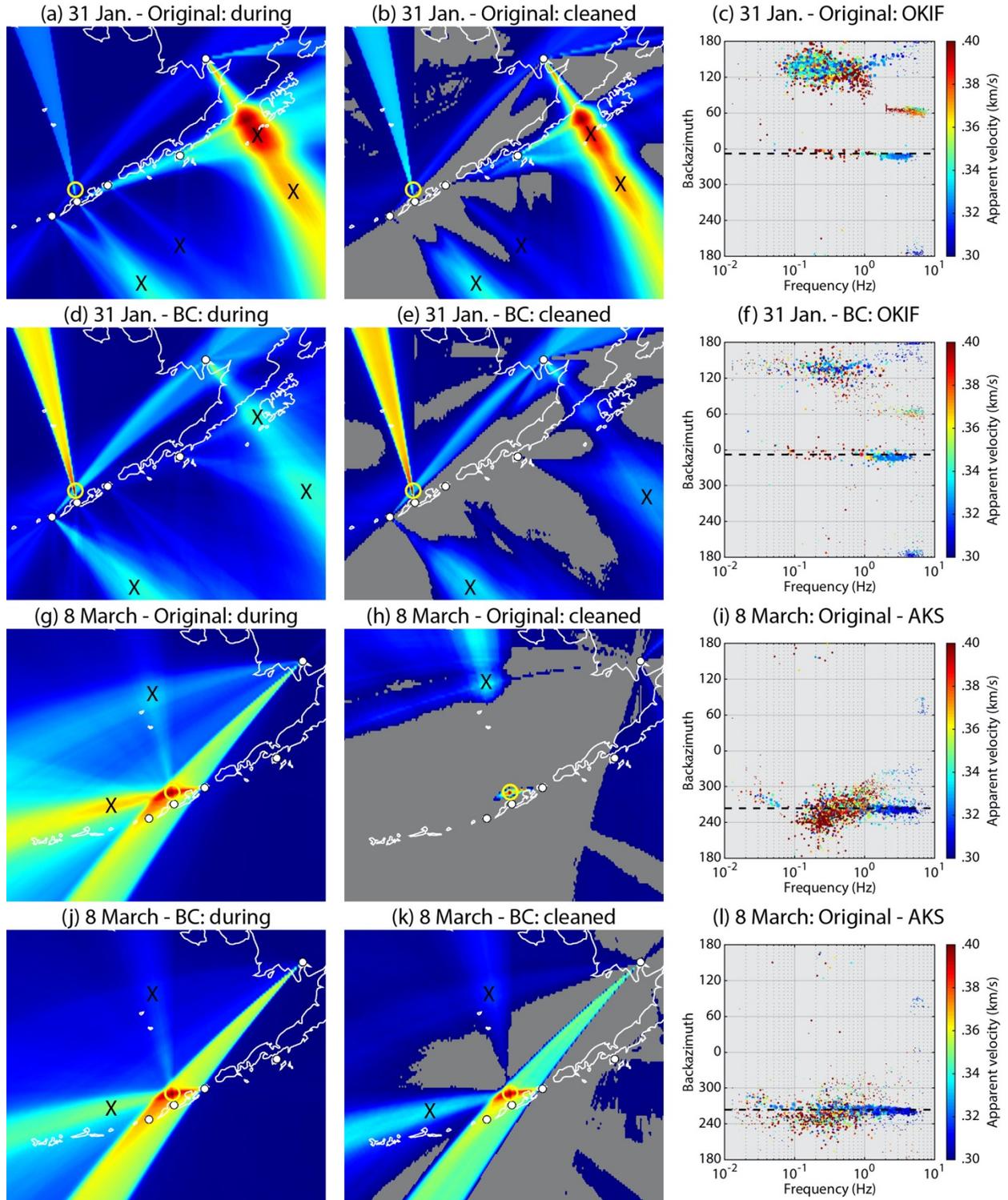


Figure 7. Mapping array cross-bearings using IMS-vASC. Reducing microbarom clutter can help isolate events of interest. This reduction can be done via data denoising and/or removing longer-term background average cross-bearing values. (a) During-grid layer, for the 31 January 2017 Bogoslof volcanic event, using

unprocessed (noisy) data. Warmer colours indicate higher numbers of overlapping backazimuths, normalized for each plot. Infrasound arrays are marked with white circles, historically active volcanoes as red triangles, with Bogoslof volcano enclosed by a yellow ring. Xs mark source areas of microbaroms. The grid resolution is 0.1° . (b) Cleaned-grid layer for data in (a). (c) Amplitude-scaled scatter plot of the frequency and backazimuth for each detection at OKIF (the true backazimuth is marked with a dashed line). (d-f) As for (a-c), but using BC-processed data. (g-l) As for (a-f), but for the 8 March 2017 Bogoslof event. Data in (l) are from AKS.

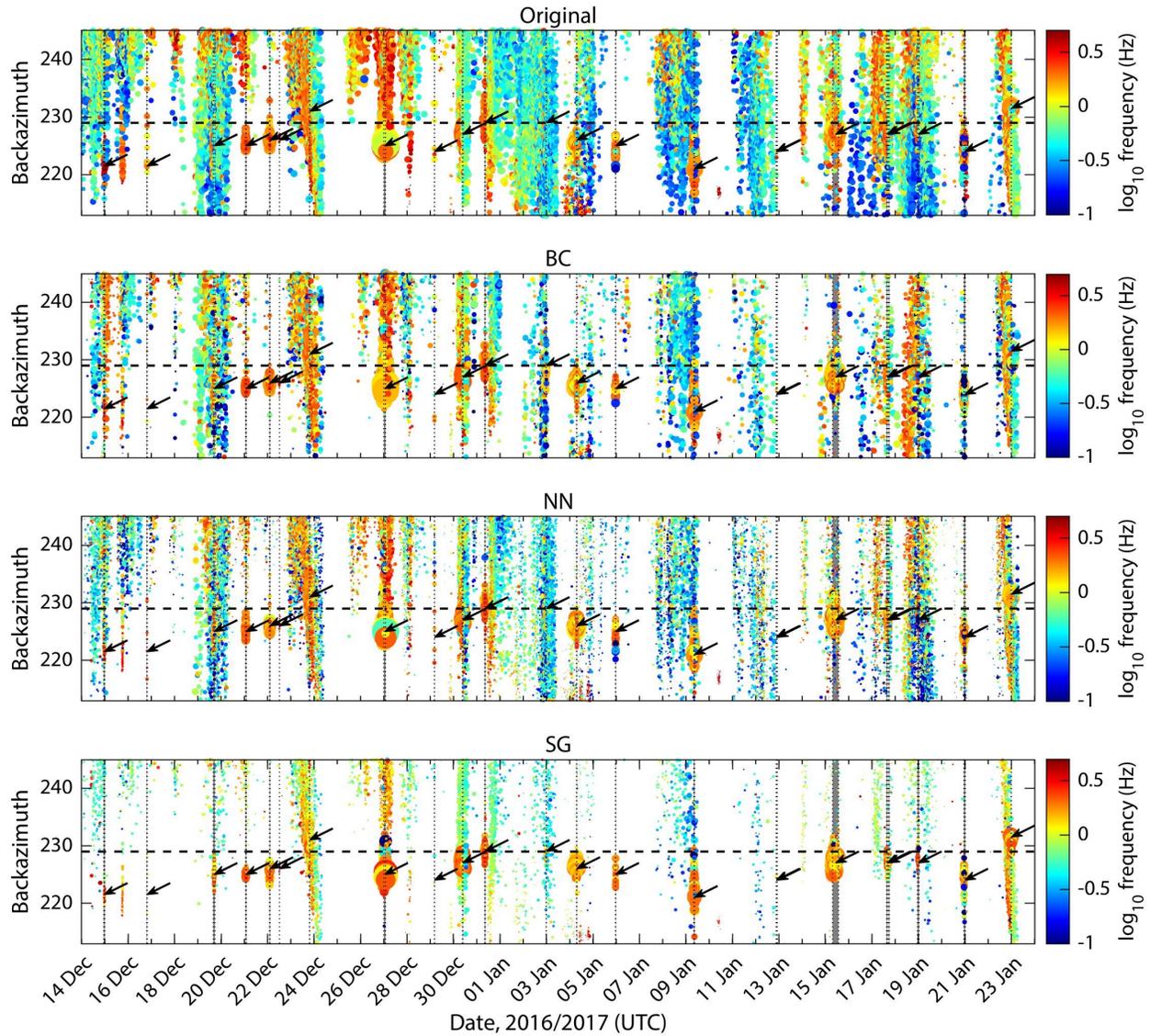


Figure 8. PMCC processing results for the period of 14 December 2016 to 24 January 2017, for noisy, BC, NN, and SG denoised data, recorded at DLL. Detections are scaled by amplitude and coloured by frequency on a log axis to help distinguish events from the microbarom. A minimum threshold of 50 pixels per family ensure only high-quality detections are displayed. Some small detections may not be visible at the figure resolution. AVO catalogued eruption periods are shaded by vertical grey bars and bound by dashed lines, using a mean stratospheric celerity of 300 m/s to calculate travel times. In this specific case, eruption end-times are taken as the latest recorded seismicity, infrasound, or lightning occurrence. Clusters of detections likely associated with these catalogued events are marked with arrows. The arrow positions are the same

for all panels despite small variations in backazimuths. Other plot details as per Fig. 6. The microbarom is generally much reduced in the denoised data in terms of detections and amplitude, allowing catalogued and potentially uncatalogued eruptions to be seen more easily.

Table 2. Denoising method ranking totals from PMCC analysis of 15 events detected at OKIF (events 1, 2, 7, 17, 28, 29, 36, 37, 39, 40, 45, 48, 49, 60, and 63 from Coombs *et al.*, 2019). Details of the classification system are in the main text. Low values indicate better performance than high values. HR provides the best overall score, although other methods are stronger in specific categories.

Category	Noisy	BC	HR	NMF	NN	SG
1. Event detections	65	50	33	52	48	58
2. Duration proportion	42	51	35	27	70	45
3. Correlation	50	61	37	47	37	82
4. F-statistic	47	47	38	56	41	86
5. Event amplitude	34	34	56	79	72	40
6. Backazimuth deviation	47	52	49	60	70	37
7. Minimum frequency	46	25	28	40	15	43
8. < 0.3 Hz detections	59	45	42	41	29	45
9. Clutter amplitude	71	62	54	52	17	48
10. Clutter detections	70	29	49	55	52	31
Total	531	456	421	509	451	515

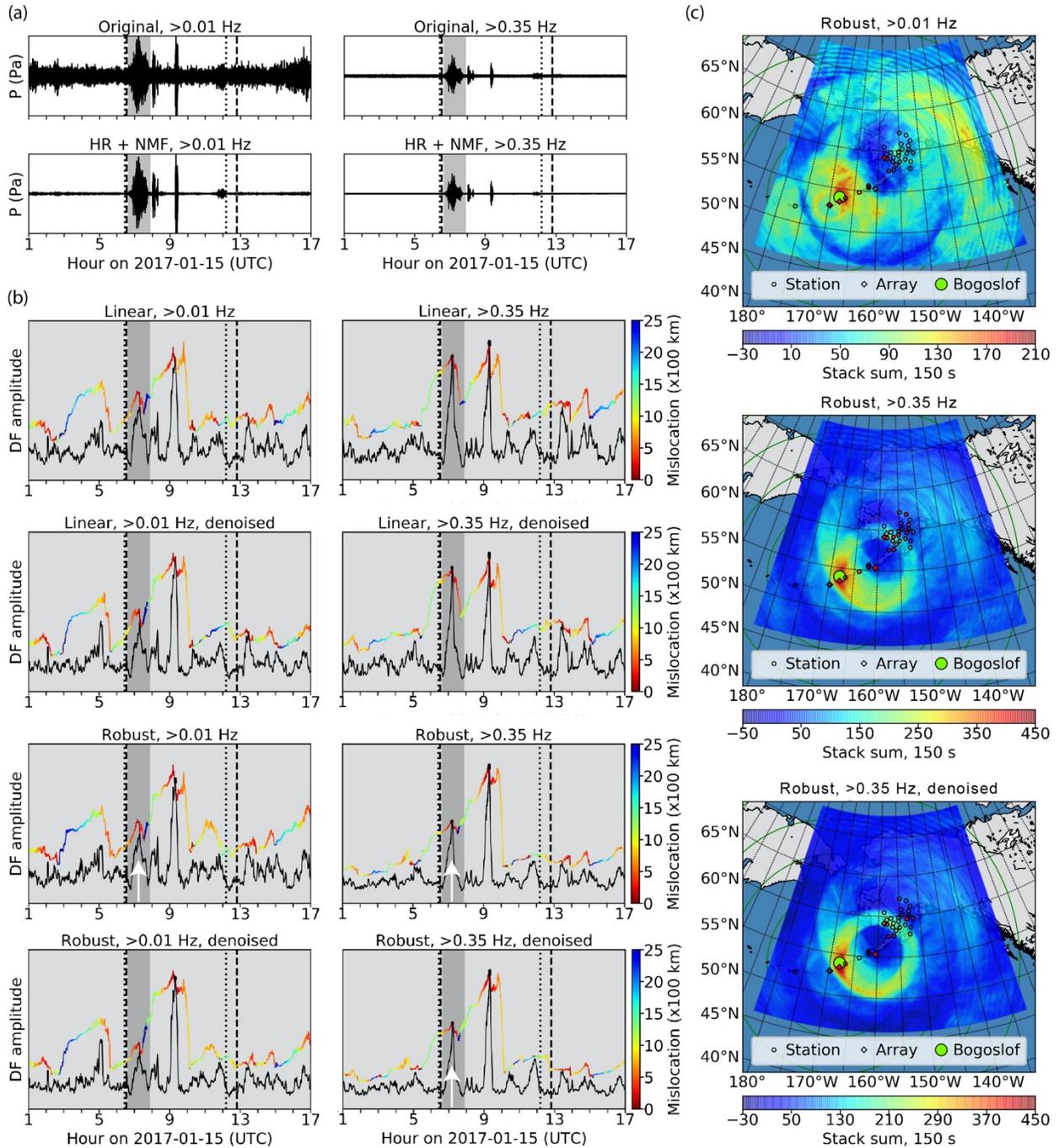


Figure 9. Examples of the effects of stacking, HR-denoising, and filtering, using Bogoslof volcanic event on 15 January 2017. Although denoising slightly improves SNR, high-pass filtering has more of an effect. (a) Beams for DLL, with differing high-pass filters. Each panel is scaled ± 1.5 Pa. The vertical dotted lines demark volcanic activity based on PMCC analysis. Vertical dashed lines demark seismic activity in the AVO catalogue, and the grey shaded area spans the catalogued infrasound activity. Infrasound travel times

have been corrected for using a celerity of 320 m/s. (b) DFs (coloured lines) all using AGC1. Black lines are stacks of traces with Bogoslof volcano as the source. Titles indicate stack type. Mislocation is the distance of the corresponding grid node from Bogoslof volcano. Intersections of the Bogoslof stack and DF imply those parts of the DF represent the true source location. Data are from 37 stations (53 channels) within 1,250 km of Bogoslof Island. Grid spacing is 0.2° latitude, 0.4° longitude. Arrows indicate the times of the maps in (c). (c) Time slices through stacks at the times of arrows shown in (b). The colour scale represents the amplitude of time-aligned data stacks at each grid location. Sensors are shaded red proportional to stack contribution. Circles are at 500 km spacing.

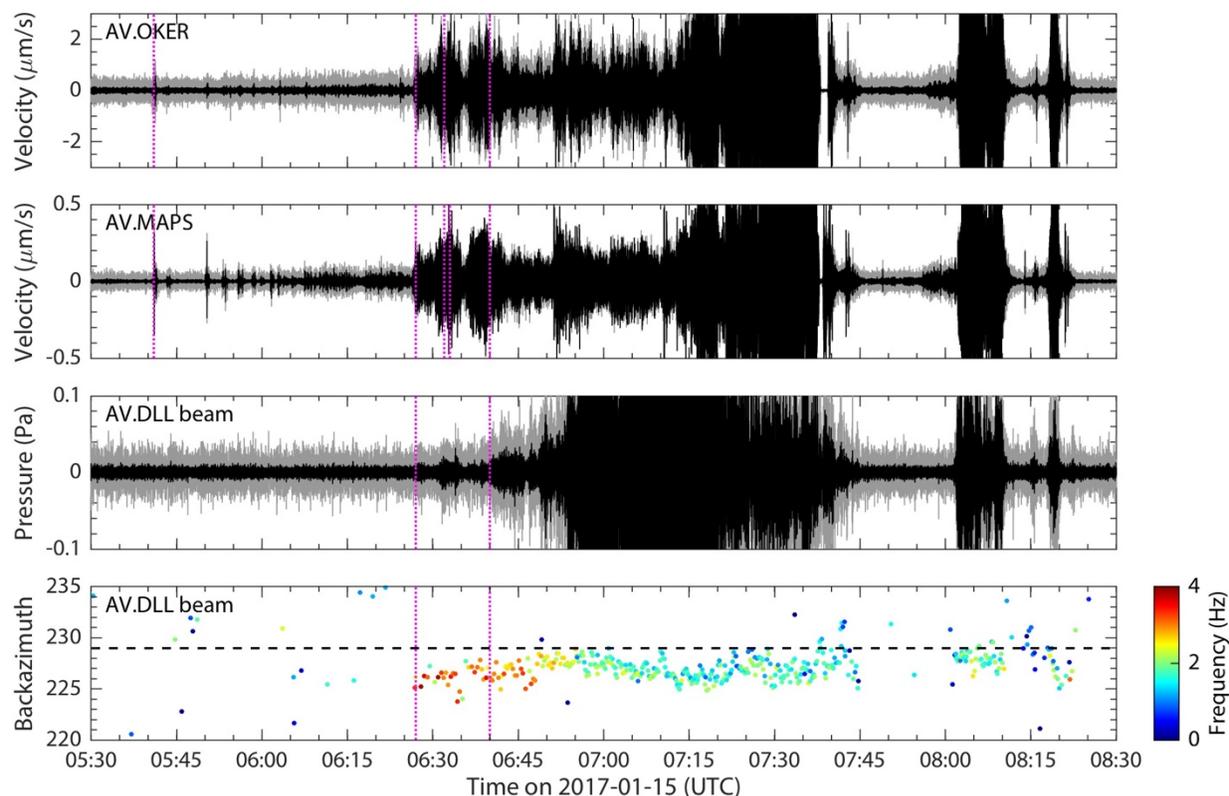


Figure 10. Comparing indicators of the eruption onset for the 15 January 2017 Bogoslof volcanic event, using seismic data from AV.OKER and AV.MAPS (1–20 Hz waveforms) and infrasound data from AV.DLL (> 0.35 Hz waveforms, and PMCC results). Data have been travel-time corrected, and clipped to emphasise smaller features. For waveforms, grey traces are instrument corrected and filtered, with black traces the SG-denoised versions. Denoising helps show waveform features previously buried in background noise, as well as reducing microbarom presence in the PMCC results. Vertical dotted magenta lines indicate the following, which include some conflicts: (1) 05:41: precursory seismic swarm onset (Tepp and Haney, 2019); (2) 06:27: explosion seismic onset at AV.OKER (Searcy and Power, 2020), with our analysis of the corresponding onset at AV.MAPS, and our estimate of the explosions and infrasound onset; (3) 06:32: end of precursory seismic swarm (Tepp and Haney, 2019), and catalogued explosion seismic onset (Coombs *et al.*, 2019); (4) 06:33: explosion seismic onset at AV.MAPS (Searcy and Power, 2020); (5) 06:40: catalogued onsets for explosion and infrasound (Coombs *et al.*, 2019).

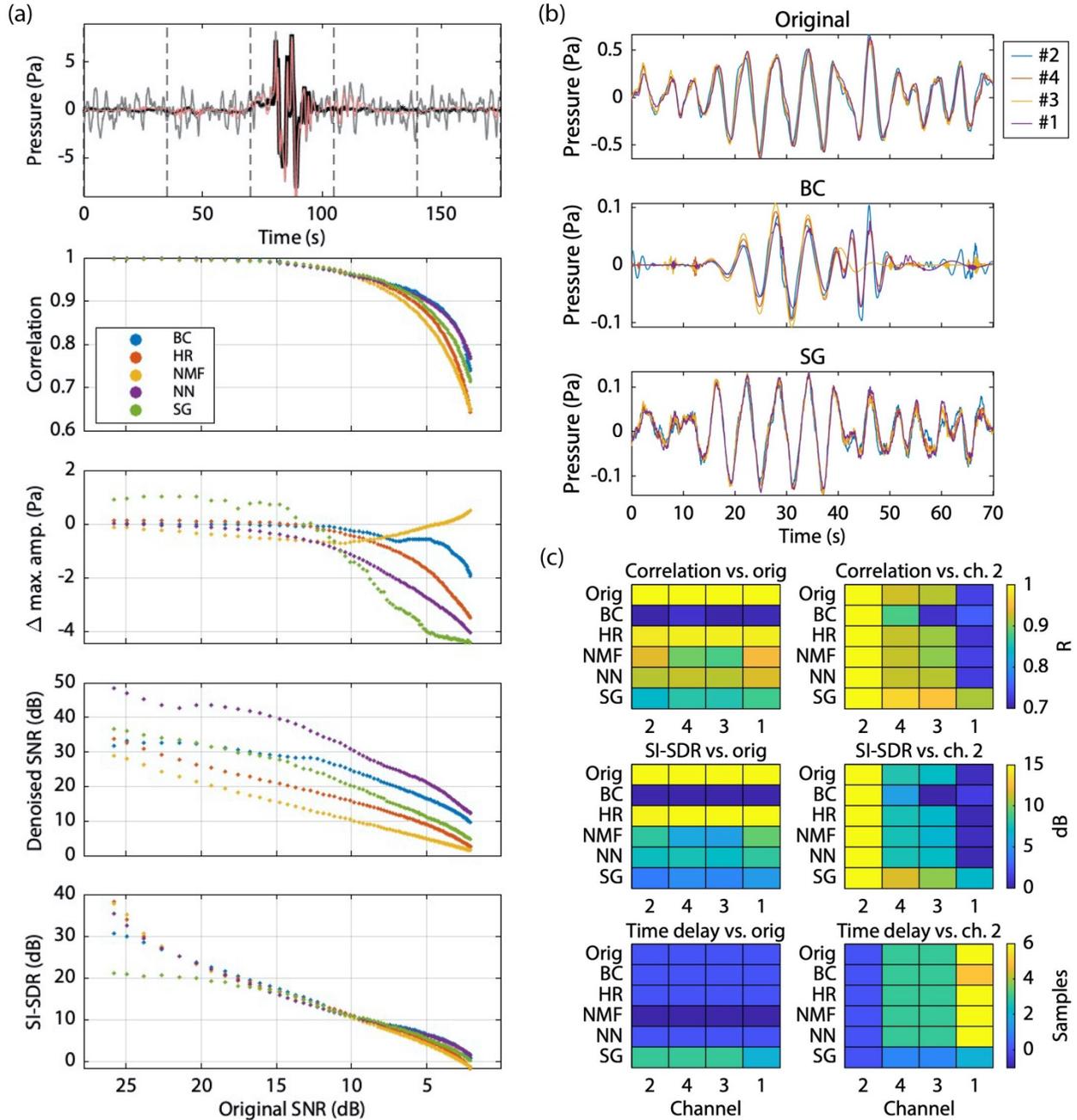


Figure 11. Evaluating the impact of denoising methods, including under different noise conditions.

(a) A high SNR Bogoslof volcanic event trace from ~02:38 20 February 2017 is contaminated by progressively higher noise levels. Top panel: Example waveform processing for OKIF channel 1: Black trace – high SNR reference waveform; Grey trace – black trace with microbarom added (scaled by a factor of 30), giving an original SNR level of ~8 dB; Red trace – denoised grey trace (using BC). Dashed lines mark periods 0–35, 70–105, 140–170 s, respectively indicating data sections used for microbarom labelling

in denoising methods that require this, the event, and noise used in SNR calculations. Panels 2–5: Correlation coefficient, change in maximum amplitude, SNR of denoised trace, and SI-SDR. Metrics 1, 2, and 4 compare denoised traces to the high-SNR reference waveform. SNR calculations are as per the Fig. 4 caption. Generally, denoising performance degrades as added noise increases, but there are wide differences in terms of denoised performance. (b) Example microbarom data prior to, and after denoising using BC and SG methods, showing variable noise reduction and phase changes. Data also from OKIF. (c) Comparing denoised to original data from (b) for each OKIF channel using correlation, SI-SDR, and time delay metrics. The left panels compare each denoised channel to its original counterpart, whereas the right panels compare each channel to channel 2 within the same processing scheme. Channels are arranged left to right by increasing distance from the microbarom source. These data show that BC and SG methods produce residual microbaroms that are more difficult to detect due to their more distorted and time-shifted signals.

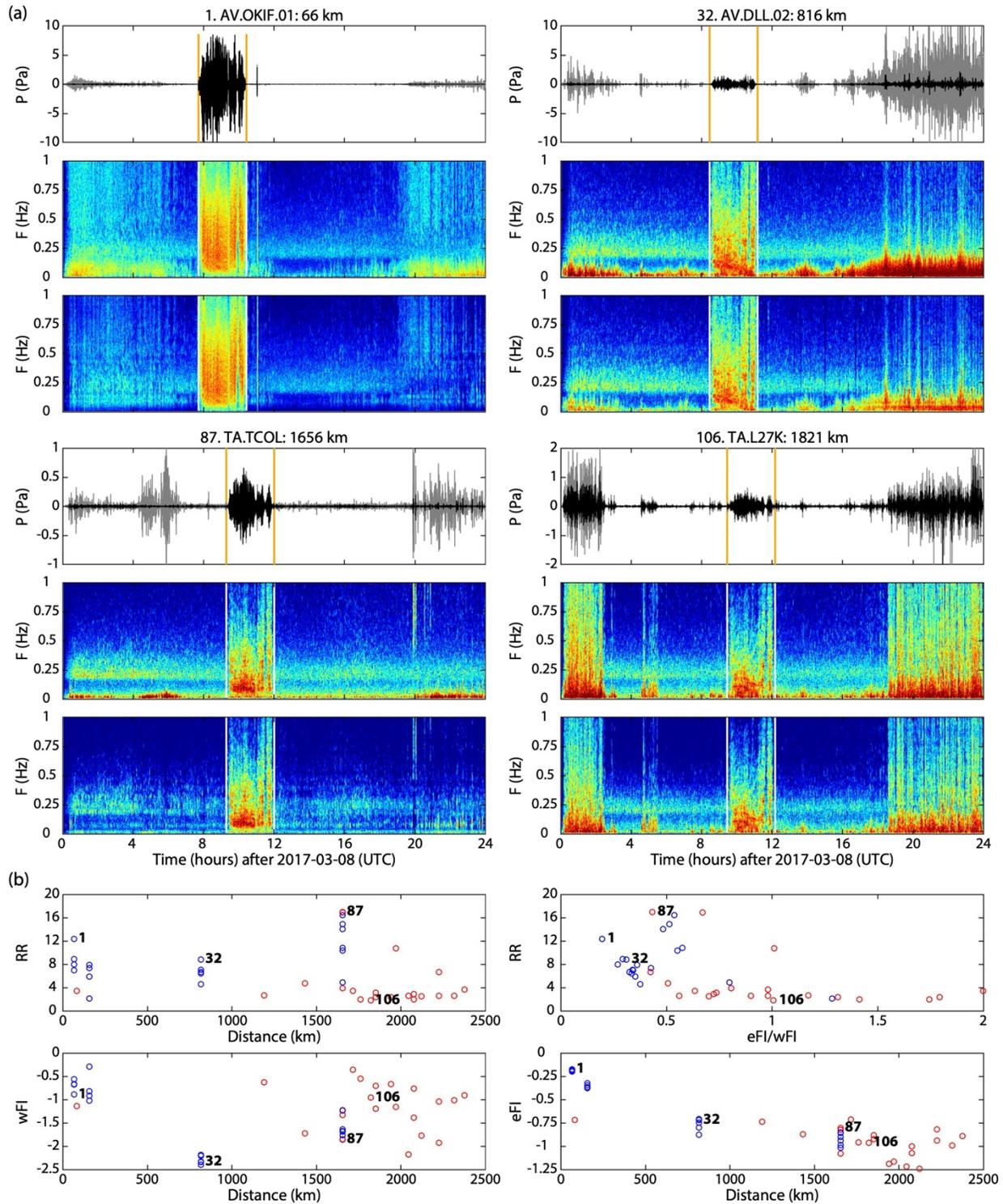


Figure 12. Assessing the effect of the relative frequency content of wind and eruptions on NMF performance. Higher SNR improvements are seen where the events have a higher frequency index relative to the wind, rather than only requiring a difference between the two. (a) Examples of HR stage 2 (grey) and

stage 3 (black) waveforms from the 8 March 2017 Bogoslof volcanic eruption, with respective spectrograms shown below. Subplot titles include the station number in terms of distance from Bogoslof volcano (not all stations shown), the corresponding distance, and the station code. Vertical bars mark the continuous infrasound phase of the eruption, according to the AVO catalogue, adjusting for a 280 m/s celerity. For each station, wind occurs primarily at the beginning and ends of the day. The upper end of each spectrogram colour scale has been adjusted such that the event portions appear to have similar maximum values. This scheme enables more direct comparisons given the different amplitudes of the signals, particularly regarding attenuation of the higher frequencies. TA.TCOL contains some non-wind noise which is not removed. P = Pressure, F = frequency. (b) Plots showing how the RR , wFI , and eFI values vary with distance, using blue for array elements, and red for non-array stations. RR is also plotted against eFI/wFI . The stations from part (a) are marked.

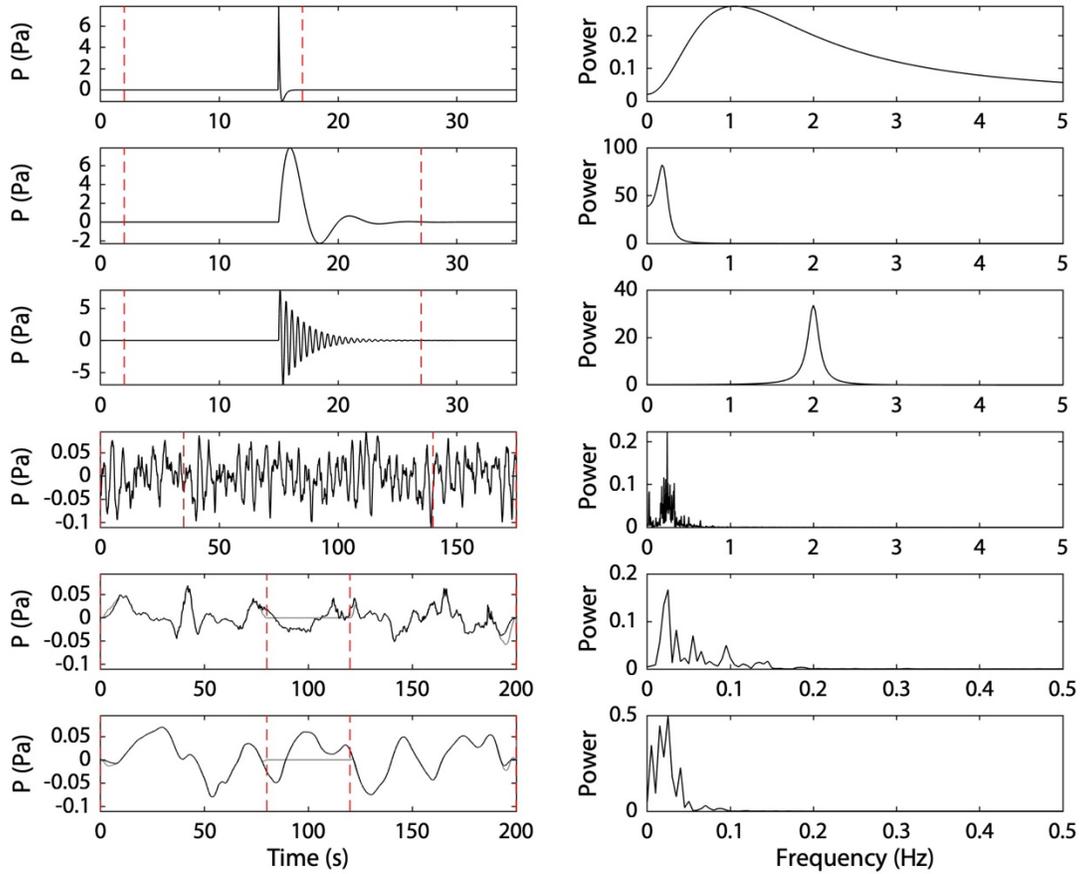


Figure A1. Waveforms and power spectra for synthetic event signals (rows 1–3), microbarom sample (row 4), and wind samples (rows 5–6). Wind in row 5 is from AV.L27K, and wind in row 6 is from AV.DLL. Each wind waveform shows the continuous sample in black, and a version where the centre is zeroed out in order to create a discontinuous version (grey). Differences at the edges are due to applied tapering. Red vertical dashed lines show the time limits used for calculations. The microbarom sample is the same as that from Fig. 11. Wind traces have already undergone stage 2 processing with the SG (spectral gating) method.

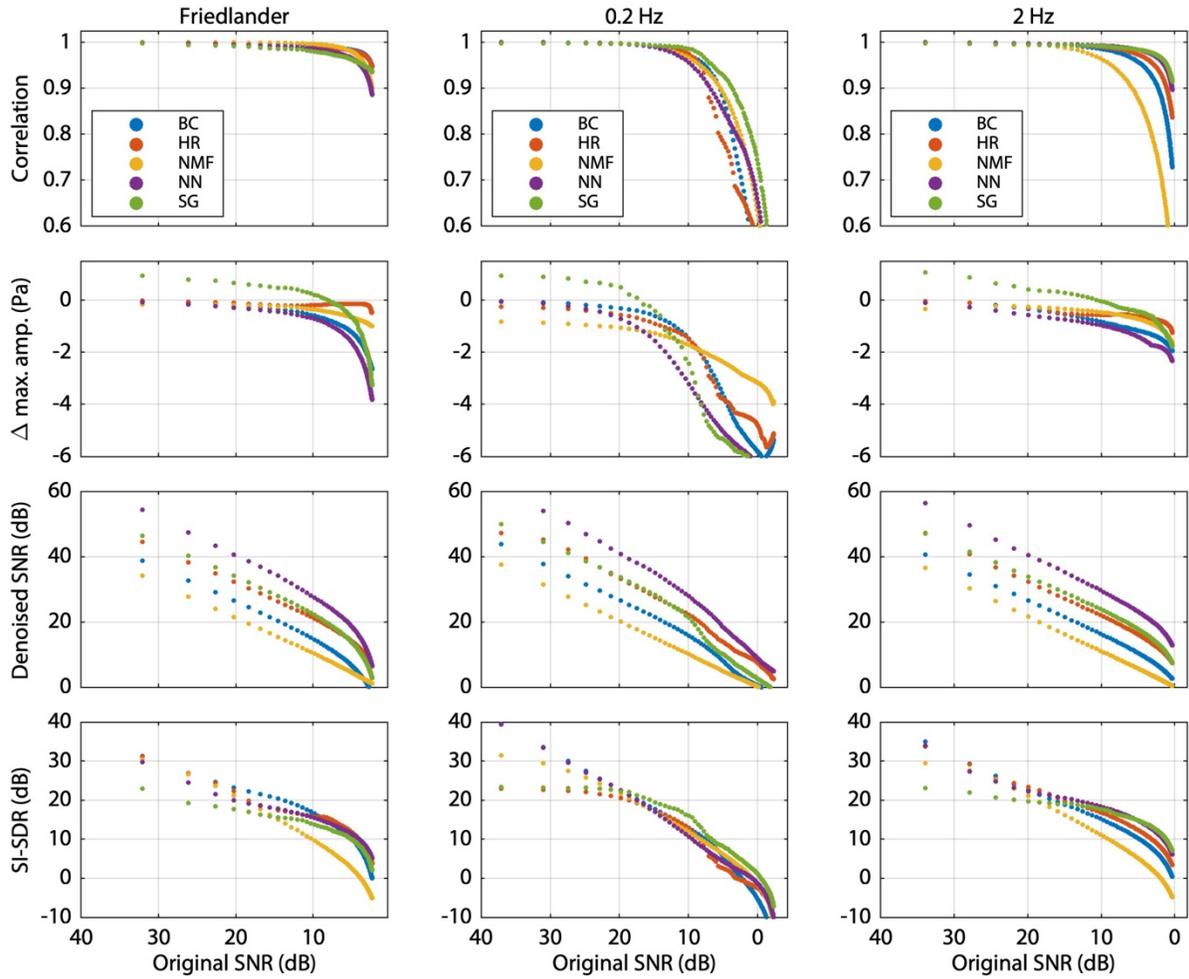


Figure A2. Evaluating the impact of denoising methods on removing the microbarom component from data, under different initial signal to noise conditions. Panels show the correlation coefficient, change in maximum amplitude, SNR of the denoised trace, and SI-SDR. Other details as per Fig. 11.

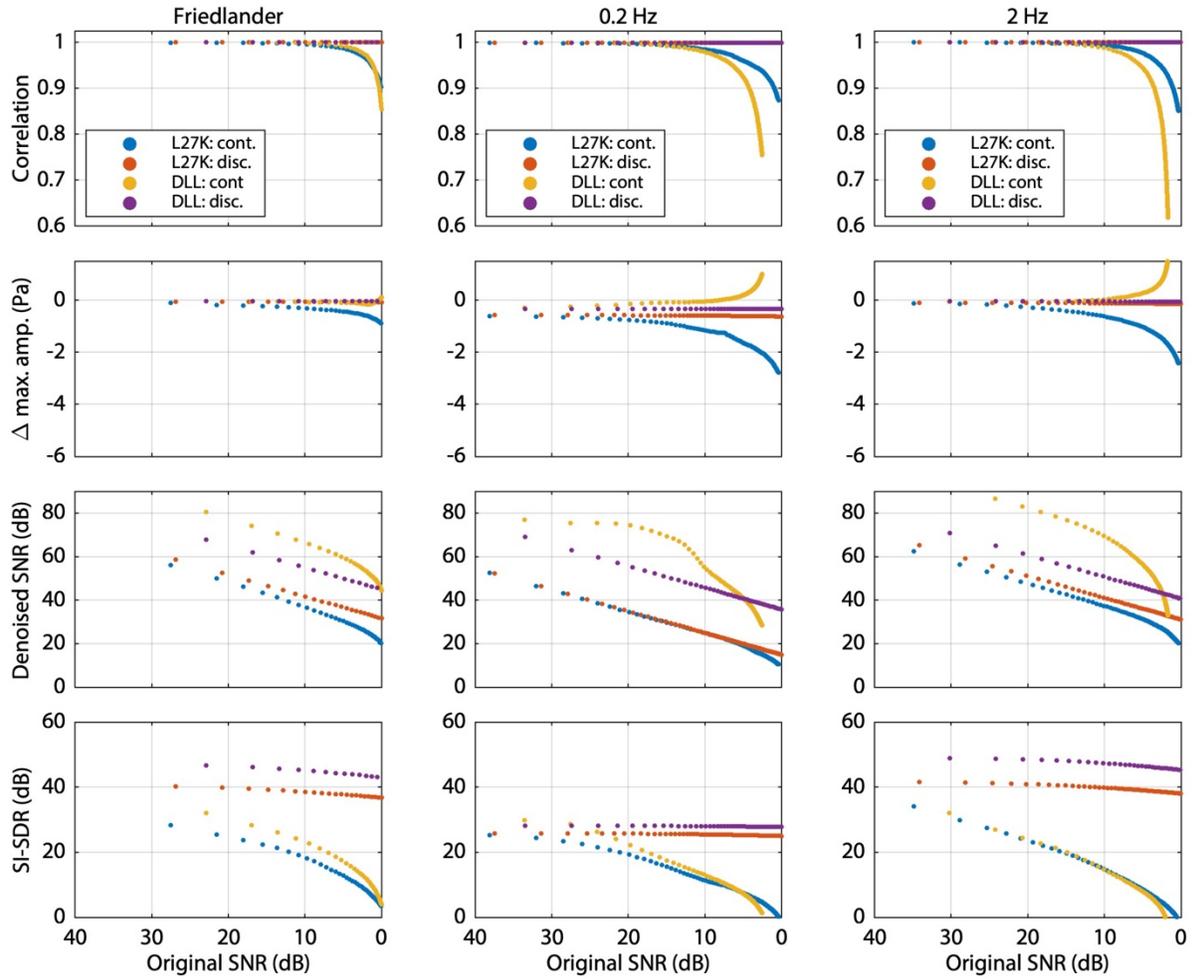


Figure A3. As for Fig. A2, but for wind. In this case, NMF is the only method applied, however. The four data sets plotted are for combinations of stations TA.L27K, and AV.DLL, and whether the wind noise is continuous through the synthetic events, or only at the ends of the traces (discontinuous).