SAND2021-14222

SANDIA REPORT

Printed

Sandia National Laboratories

Utilizing the Dynamic Networks Data Processing and Analysis Experiment (DNE18) to Establish Methodologies for the Comparison of Automatic Infrasonic Signal Detectors

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods

ABSTRACT

The Dynamic Networks Experiment 2018 (DNE18) was a collaborative effort between Los Alamos National Laboratory (LANL), Sandia National Laboratories (SNL), Lawrence Livermore National Laboratory (LLNL) and Pacific Northwest National Laboratory (PNNL) designed to evaluate methodologies for multi-modal data ingestion and processing. One component of this virtual experiment was a quantitative assessment of current capabilities for infrasound data processing, beginning with the establishment of a baseline for infrasound signal detection. To produce such baselines, SNL and LANL exploited a common dataset of infrasound data recorded across a regional network in Utah from December 2010 through February 2011. We utilize two automated signal detectors, the Adaptive F-Detector (AFD) and the Multivariate Adaptive Learning Detector (MALD) to produce automated signal detection catalogs and an analyst-produced catalog. Comparisons indicate that automatic detectors may be able to identify small amplitude, low SNR events that cannot be identified by analyst review. We document detector performance in terms of precision and recall, demonstrating that the AFD is more precise, but the MALD has higher recall. We use a synthetic dataset of signals embedded in pink noise in order to highlight shortcomings in assessing detection algorithms for low signal to noise ratio signals which are commonly of interest to the nuclear monitoring community. For comparisons utilizing the synthetic dataset, the AFD has higher recall while precision is equal for both detectors. These results indicate that both detectors perform well across a variety of background noise environments; however, both detectors fail to identify repetitive, short duration signals arriving from similar backazimuths. These failures represent specific scenarios that could be targeted for further detector development.

ACKNOWLEDGMENT

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This research was funded by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development (NNSA DNN R&D). The authors acknowledge important interdisciplinary collaboration with scientists and engineers from LANL, LLNL, MSTS, PNNL, and SNL. The authors acknowledge programmatic support for this work by Neill Symons, and gratefully acknowledge Brian Stump and Ellen Syracuse for comments on this manuscript.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Table 0-1.

Abbreviation	Definition
DOE	Department of Energy
DNE18	Dynamic Networks Experiment 2018
AFD	Adaptive F-Detector
MALD	Multivariate Adaptive Learning Detector
PMCC	Progressive Multi-Channel Correlation
IP	InfraPy
BH	Bloodhound
DOA	Direction of Arrivals
ROC	Receiver Operating Characteristic
SNR	Signal to Noise Ratio
UU	University of Utah
SMU	Southern Methodist University
PSD	Power Spectral Density

1. INTRODUCTION

The Dynamic Networks Experiment 2018 (DNE18) was a data-processing experiment designed to quantitatively assess current capabilities for multi-modal data ingestion and processing for nuclear explosion monitoring at the local/regional scale [8]. The experiment was a collaboration between Los Alamos National Laboratory (LANL), Lawrence Livermore National Laboratory (LLNL), Pacific Northwest National Laboratory (PNNL), and Sandia National Laboratories (SNL), and consisted of the joint processing of three months [12/2010-02/2011] of seismic, acoustic and radionuclide data from networks within the western US. See [48] for a complete overview of the experiment. Automated infrasonic data processing was conducted at both the station (signal detection) and network (event association and location) levels with the goal of establishing "baseline" performance metrics [14]. This manuscript focuses on the evaluation of station-level, infrasound signal detection performance.

The automation of infrasonic detectors is motivated by a need to reduce analyst workload as well as streamline large-scale infrasonic data processing [38]. A variety of automated infrasound detectors have been presented in the peer-reviewed literature including the progressive multi-channel correlation (PMCC) algorithm [13], the F-detector implemented in InfraTool [24], the Adaptive F-Detector (AFD) [5] and the Multivariate Adaptive Learning Detector (MALD) [1]. The AFD is implemented within Infrapy (IP), a Python-based array processing toolkit developed by LANL that was recently open-sourced in 2020 [9, 17, 33, 46]. The MALD is implemented in Bloodhound (BH), a Python-based pipeline data processing software developed by SNL [4].

Automated processing techniques have been used to produce bulletins of infrasound events in multiple regions such as the western United States [37, 45], the state of Alaska [42], the Korean peninsula [40, 36, 16] and across the European continent [41, 31]. Automatic detections across the 48 operational stations of the IMS infrasound network are documented by the International Data Center (IDC) and more recently by [1]. In addition, a number of bulletins have been produced specifically documenting infrasonically-observed volcanic activity (e.g., [34, 42]).

Despite the existence of large-scale, multi-year detection bulletins, most are produced by a single automated detector with few direct comparisons between detection algorithms [38] motivating this comparative study of detectors. Two measures of detector performance are the minimum signal-to-noise ratio (SNR) for signal detection or the accuracy of the direction of arrivals (DOA) parameters. These measures can be evaluated using statistical tools such as receiver operating characteristic (ROC) curves, which quantify the relationship between detection and false-alarm probability as a function of detection threshold [28]. The development of ROC curves for a particular detector necessitates a dataset where signals are known, and where signals and noise span the space of all possible characteristics [38]. One evaluation approach utilizes synthetic signals and noise, which provides full control of all effects, but is not always representative of true signal and noise scenarios. The alternative evaluation approach utilizes a catalog of real

signals detected by an analyst, which is closer to true signal and noise scenarios but assumes a subjective definition of a 'signal' which can be difficult in evaluating low SNR events [38]. A hybrid approach utilizes synthetic signals constructed using a combination of real signal and noise sequences. Results from such comparisons can be synthesized by comparing probabilities of detection (P_D) and probabilities of false alarms (P_{FA}) [32]. However, the lack of practical comparative assessments of infrasound data processing algorithms is likely due to differences in operational tuning parameters for all algorithms, which makes direct probability comparisons challenging. Additionally, algorithm assessment is complicated by the lack of a concrete and consistent baseline of what comprises a true detection (i.e. signal) or a false alarm (i.e. noise), as infrasound arrays frequently identify consistent signals originating from numerous coherent noise sources [20, 27] including surf [30], microbaroms [29], thunder [21], volcanoes [30] and anthropogenic activities such as mining, industrial activity, aircraft or urban noise [39, 35, 31].

[38] addresses these challenges by comparing Estimated Receiver Operating Curves (EROC) in order to compare the performance of PMCC and AFD taking a series of analyst detections as true detections. The authors note that while several tuning parameters are common to the two detectors, parameter differences make direct EROC comparisons difficult. An alternative method, presented in [1], compares the overall numbers of detections and their distributions in terms of each algorithms false alarm rate. In this manuscript, we present a hybrid methodology that utilizes the metrics of precision and recall to address this complex issue of assessing infrasound data processing algorithms, beginning with detection catalog comparisons. This study documents a series of automatic processing results from DNE18 supplemented by an analyst catalog and a series of synthetic waveforms. These diverse sets of test data provide an opportunity for a hybrid evaluation of the infrasonic detection algorithms. The dataset which is shared also provides a known baseline for others in the infrasound community to evaluate and compare the performance of subsequent infrasonic detectors. We provide these results as a reference point for continued detector comparisons within the infrasound research community.

2. DETECTOR OVERVIEW

The AFD as implemented within Infrapy (IP) [46] was developed to reduce false alarms from correlated or persistent noise across an array of interest [5]. The standard F detector provides a measure of the single channel SNR and utilizes Fisher statistics to identify signals of interest. It is based on calculating the signal power and the total power across an array [43]. The null hypothesis of this statistical test assumes that no coherent signal of interest is present. In the case that a signal of interest is present in the data, the F-ratio, derived from the power across the array, increases. An assumption within this method is that all the noise across the array is incoherent, while the signals are coherent. This assumption sometimes fails for infrasound data, which often contains coherent sources of background noise at a variety of spatial scales. Coherent noise elevates background F-values, resulting in false alarms unassociated with a signal of interest. AFD accounts for these elevated F-values by increasing the value of the F-statistic required to declare a detection. This adaptation is accomplished through the application of a C-value, which effectively reduces the detection threshold (p-value) and decreases the number of noise-related detections by constantly re-mapping the conventional F-statistic based on moving estimates of the background noise. [6] and [5] provide a detailed theoretical description of the Adaptive F-Detector. An overview of parameter tuning necessary to optimize detection of sources within the western US and Korean peninsula can be found in [40] and [38]. [39] and [18] discuss the application of the AFD to infrasonic signal detection using a variety of observational data.

Automatic detection processing in Bloodhound (BH) utilizes a multivariate adaptive learning detector (MALD) which is detailed in depth in both [4] and [1]. Similar to the AFD, the goal of the detector is to identify long-range infrasonic signals from explosive sources, while simultaneously reducing false alarms (FA) such as smaller local events or known coherent noise sources as discussed above. MALD accomplishes these design goals through a multi-step processing methodology where three distinct time windows are used to adaptively alter detection thresholds in order to account for the changing background noise environment. MALD utilizes semblance (S) to estimate the ratio of the power of the stacked beam to the average power of the individual traces producing values from 0 to 1. Across each moving time window, distributions of the maximum semblance value max (S), corresponding backazimuth estimate (ϕ), and variance of the backazimuth $\theta^2(\phi)$, are fit using Kernel Density Estimation (KDE). These measures provide the ability to convert test statistics for semblance and backazimuth to p-values while avoiding assumptions about the statistical properties of the background noise. The inclusion of backazimuth variance in the detector was motivated by signals that were missed by the AFD due to decorrelation at long range (an issue highlighted by [23] but clearly exhibit a stable estimate of backazimuth over time (refer to Figure 10 in [2]). This detector is similar in principle to the basis of the Hough Transform detectors of [12] and [7] combining coherence and backazimuth constraints. The p-value is the probability of obtaining a result that is equal to, or more extreme

than, what is observed in the data given the KDE. Sets of p-values for signal coherence and backazimuth variation are then combined using Fisher's method to produce ensemble p-values:

$$\chi^2 = -2 \sum_{k}^{i=1} lnp_i \tag{2.1}$$

where χ^2 is estimated for each time window by resampling the estimates of $\theta^2(\phi)$. k corresponds to the number of transforms for a given data type and p corresponds to the p-value related to that transform. The p-value is then used to determine the presence or absence of a detection based on both signal semblance and backazimuth estimates.

Both detectors were developed for use with arrays, whether seismic or infrasonic and utilize FK processing techniques to extract information (DOA, power, trace velocity). Each employ a distinct methodology to adaptively account for coherent background noise. BH builds a KDE-based background noise distribution while IP utilizes an F-distributed background noise estimate. BH includes the additional constraint that the backazimuth should be stable over some time interval. Differences in background noise thresholds may lead to discrepancies in the detection capabilities of each method.

3. INSTRUMENT DEPLOYMENT AND DETAILS

The infrasonic dataset used in this study consists of array data from six co-located seismo-acoustic stations (Figure 3-1) installed in the state of Utah by the University of Utah (UU) and Southern Methodist University (SMU) between 2006-2010 [49]. Each array consists of 4 acoustic elements; one center element with three additional elements equally spaced around the center at 100 m range. Infrasound sensors are each fit with multiple porous hoses to reduce wind noise [44]. Data were sampled at 100 samples per second. Acoustic sensors at NOQ are Chaparral Physics Model 2.0 microphones [37] with a flat frequency range of 0.1 to $>$100 Hz [6]. NOQ is equipped with a RefTek digitizer. BRP, FSU, HWU, LCM, and WMU are equipped with Inter-Mountain Lab (IML) sensors and Q330 digitizers. The frequency response for the IML sensors is flat from 2 to 30 Hz [22, 25]. Data is available from IRIS using the network code 'YJ' [26].

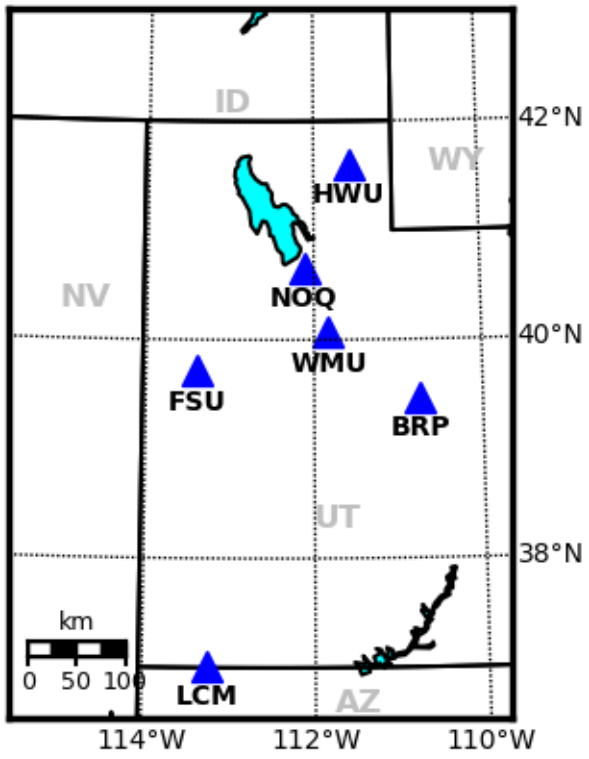

Figure 3-1. UU/SMU seismo-acoustic network used for DNE18 processing. Blue triangles indicate infrasound arrays utilized within this study.

4. VIRTUAL EXPERIMENT DETECTION RESULTS

Data from all six arrays were processed for the three-month time period as discussed in introduction to the virtual experiment. Parameters for the two detectors were defined to present as equal of a detector comparison as possible, but we highlight that the differences between the implementation of the detectors make a direct parameter comparison difficult. The parameters used for detection processing are included in Table 4-1. Parameters were set based on experience developed from prior infrasonic analyses in the Utah region [3, 18, 38, 39]. However, we note that all these publications focus on parameter optimization for the AFD. [1] notes that parameter optimization for MALD has not yet been rigorously studied, and published work focuses on parameters that worked well for test signals based on manual tuning. Therefore, we utilized a set of parameters intended to serve as a baseline for the detector comparisons. Based on our results, future work is needed to further examine and optimize these parameters. A p-value of .001 was used for detection at FSU, HWU, LCM and WMU while a p-value of .01 was used at BRP and NOQ. Lower p-value thresholds were implemented at stations where analyst inspection of automated results identified many false alarms due to multiple detections of persistent signals likely originating from anthropogenic sources. Lowering the p-value thresholds led to fewer false alarms.

Table 4-2 summarizes the total number of detections at each array from the three months based on each detector, while Figure 4-1 visually present detection results. As seen in Figure 4-1a, BH detects significantly more events at all stations except NOQ, where IP detects 224% more events. An examination of detection statistics by backazimuth is presented in Figure 4-1b and offers insight into the differences in signals identified by each detector. At BRP, both BH and IP identify signals originating from 30-350°and 220-260°; a significant portion (40%) of BH detections occur between 220-260°while only 10% of IP detections originate from that direction. Similarly, at WMU detections originate from 0-45°, 85-135°and 220-260°, but BH detections disproportionately originate from 0-45°.

An example of these repeating signals identified by BH is shown in Figure 4-2a, where the recurring source originates from 0°, is low frequency and has a duration from 20 to 30 sec. Repeating signals identified by both detectors originate from 135°, both higher in frequency and longer in duration. At HWU, BH identifies detections from 290-345°and 45-135°that are not detected by IP as seen in Figure 4-2b where signals are short in duration and low in frequency ($<$1Hz). These results suggest that BH may be detecting a recurring source that is presumed to be noise by IP. The opposite relationship between detectors is found for NOQ, where both detectors identify significant signals from 270-315°and 110-140°; IP detections are disproportionately from 110-140°, suggesting that IP is identifying a recurrent source that is presumed to be noise by BH. These discrepancies are likely related to the background noise distributions utilized for each respective detector and indicate further comparisons may be necessary. At present, these

Table 4-1. Parameters used for automatic processing.

Parameter	Value
Filter band (Hz)	1-5
Time window (s)	30
Overlap (%)	50
p-value	0.001,0.01
Semblance Threshold (BH)	0.5
Adaptive window for AFD (s)	1800

Table 4-2. Number of automatic detections at each array for IP and BH.

Array	InfraPy	Bloodhound
BRP	11541	25143
FSU	3121	3276
HWU	6747	22589
LCM	9500	16458
NOQ	3584	1106
WMU	6913	13840

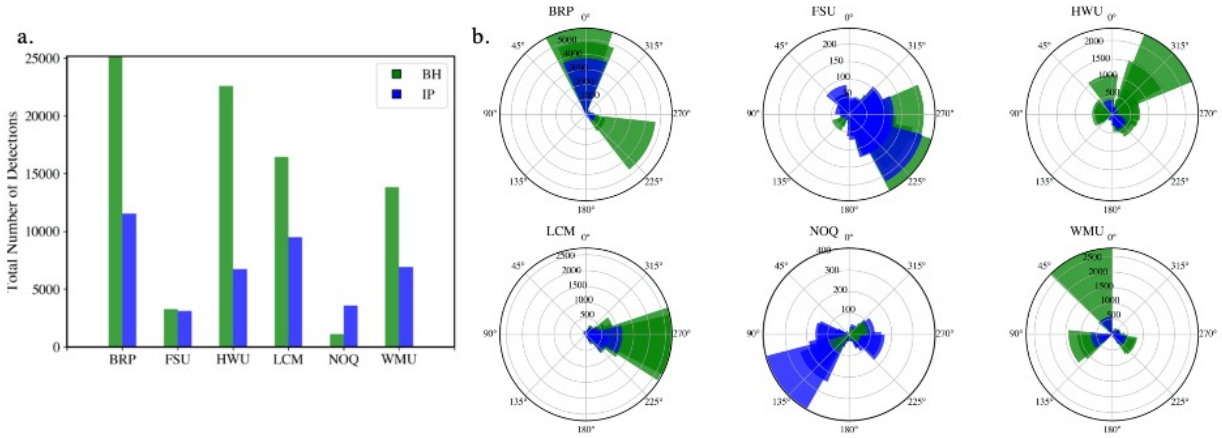

Figure 4-1. (a) Full experiment automatic detection results from Bloodhound (green) and Infrapy (blue). (b) Full experiment detection results as a function of backazimuth at each station within the network.

recurring sources, or clutter, are of unknown origin. Further investigation and identification may assist with continued detector improvements.

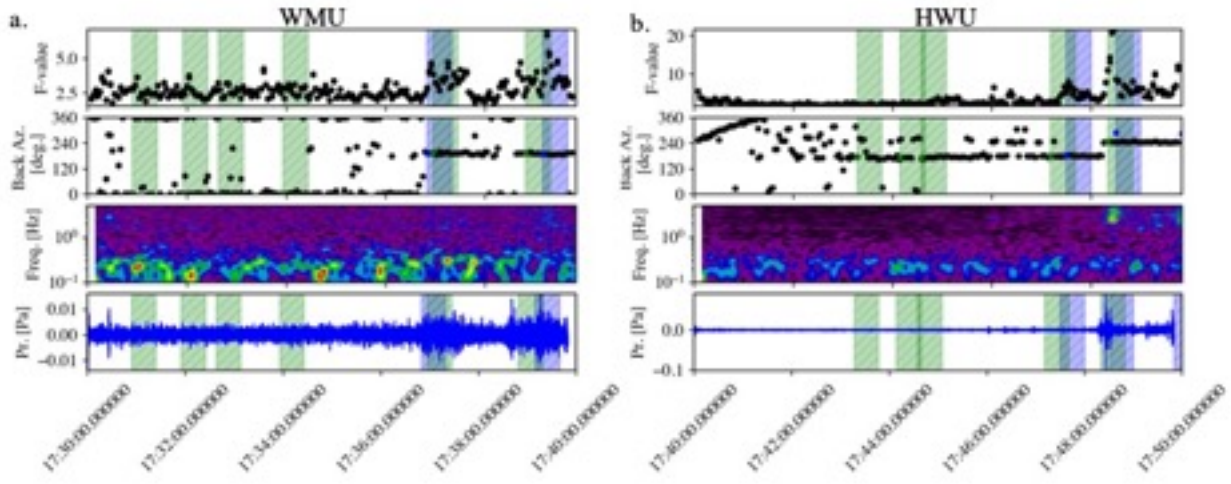

Figure 4-2. Example detections from two arrays (a) WMU and (b) HWU where panels represent (top to bottom) F-value or power across the beam, signal backazimuth, spectral density of signals and beamed amplitude (in Pa). Green windows represent BH detection onset time with a 30 s error window. Blue windows represent IP detection onset time with a 30 s error window.

5. ANALYST DETECTION CATALOG

In order to support the interpretation of the detection results, an analyst detection catalog was produced for two hours of data on 2011-01-05 at 07:00 and 17:00 UTC, corresponding to 12:00 am local time and 10:00 am local time. These hours were chosen to provide insight into the significant discrepancy in detection numbers for the two detectors, as well as producing a comparison of detection capabilities at midday and midnight (local time) consistent with different noise conditions. Analyst (A) detections were identified using FK processing on the array beam and a singular broad bandpass filter band of 0.5-5 Hz that mimicked automatic processing. The use of a single broad frequency band may cause the analyst to miss identification of narrowband signals. Following guidelines for complementary seismic analyst detection [48], a detection was declared when the analyst identified either: (a) a significant increase in beam amplitude, corresponding to an increase in F-value (equivalent to beam power); or (b) a consistent (10 sec) backazimuth trend. Table 5-1 includes the total number of detections at each array for the day and night hours, while Figure 5-1 visually compares overall hourly detection results. Mirroring automatic detector results, there are few analyst detections at the arrays during the first hour (07:00 UTC), but numerous analyst detections during the second hour (17:00 UTC). At FSU, IP detections exceed both BH and analyst detections. At NOQ, analyst detections exceed automatic detections, followed by IP detections. At the remaining four arrays, BH detections greatly exceed both IP and analyst detections.

Detections were compared in the time domain to determine if the different methods identified the same source based on a combination of temporal onset, detector window error and backazimuth estimates. Following [15], a positive detection, or a *true positive*, was declared when both the automatic detector temporal error window (30 sec) and the automatic detector backazimuth estimate ($\pm$ 5 °of error) overlapped with the analyst estimate. The 30 sec error window is determined based on the window length used for automatic processing. Figure 5-2 shows a schematic of this process. An example of these comparisons is presented in Figure 5-3, where the beamed waveform, back-azimuth estimates and F-value estimates for two, ten minute section of data in the Hour 2 time period are presented. We highlight six distinct scenarios within the time series, corresponding to differing relationships between each detection onset time and corresponding 30 sec error windows. Recurring BH-only detections, examples of false positive detections, are visible in Figure 5-3b which correspond to signals originating from a backazimuth of 220°and 260°. Signals are consistently low SNR.

Hourly results from automatic detector comparisons with the analyst detection set are presented in Figure 5-4. For the first hour (07:00 UTC), results vary across the network. At BRP, BH detects 30 events while IP and the analyst each detected 1 event. There is 1 event that was detected by all three methods. At LCM, BH detects 12 events while IP detects 6 and the analyst detects 3. Two events are detected by all three methods, while 3 events are detected by IP and BH.

Table 5-1. Number of detections for hour 1 (HR1) and hour 2 (HR2) at each array for each of the detection methods (IP = Infrapy, BH = Bloodhound, A = Analyst).

	IP HR1	IP HR2	BH HR1	BH HR2	A HR1	A HR2
BRP	1	19	30	24	1	18
FSU	0	10	0	8	0	7
HWU	0	7	1	20	0	10
LCM	6	9	12	13	3	14
NOQ	0	3	0	2	2	2
WMU	0	5	0	25	0	11

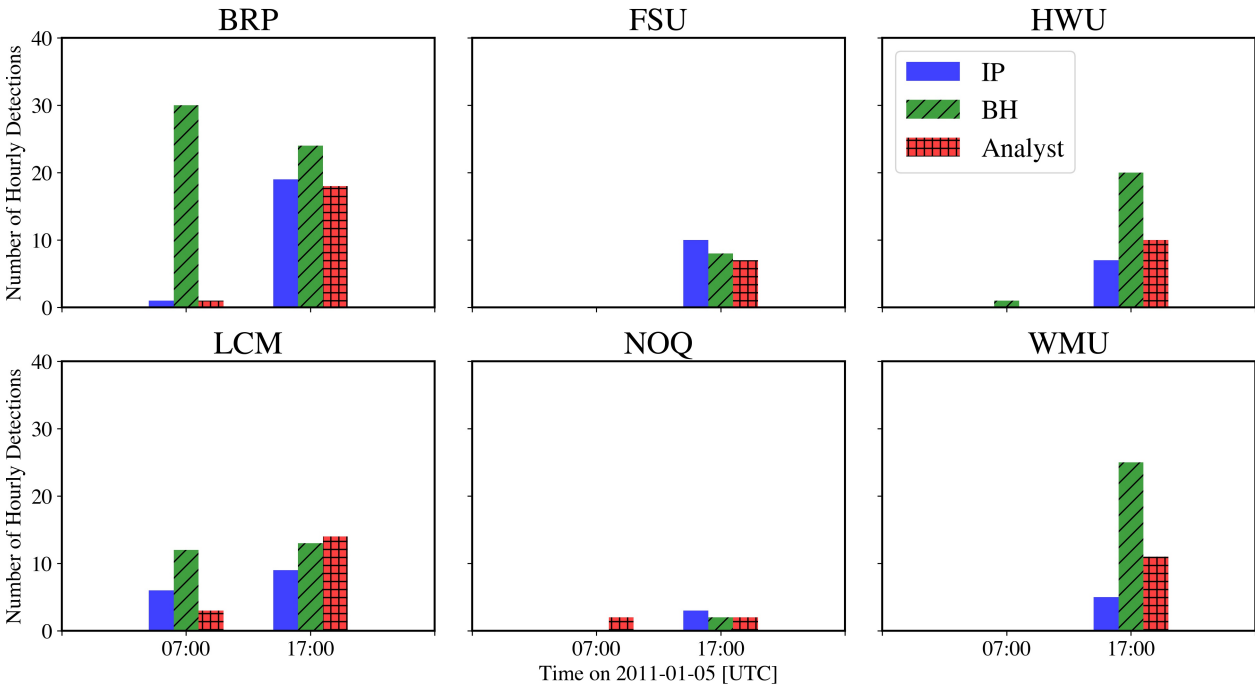

Figure 5-1. Hourly detections for hours 7:00 and 17:00 at each array within the network for IP (blue), BH (green) and the analyst (red).

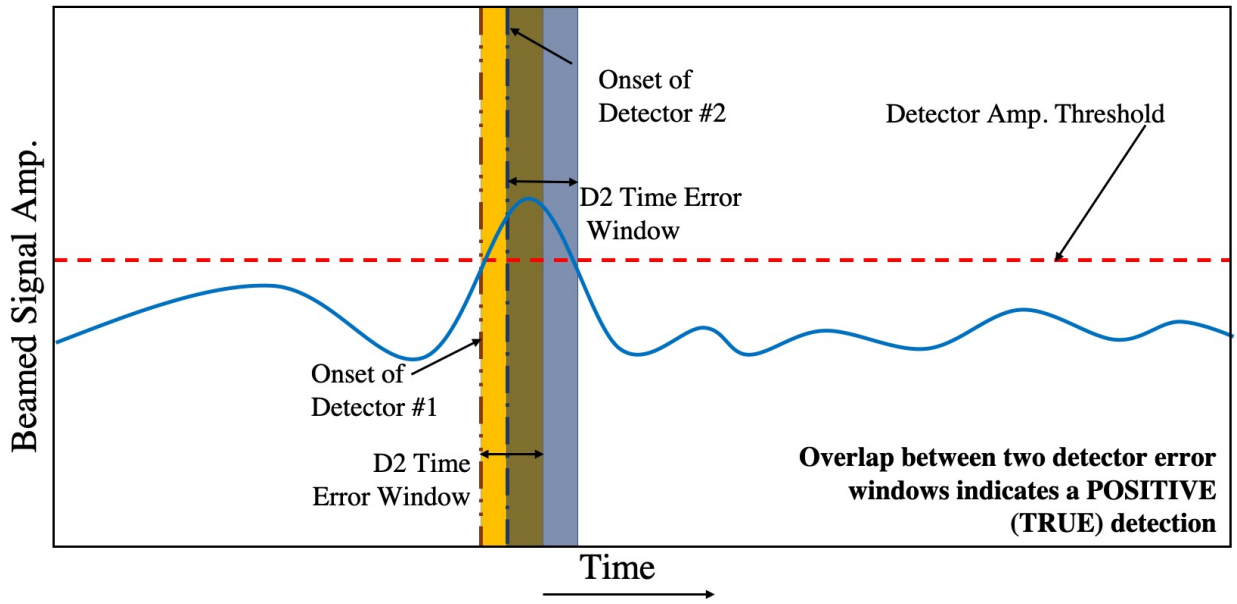

Figure 5-2. An illustration of the procedure to identify positive detections based on detector onset and temporal error windows. Dashed lines indicate the onset of each respective automated detector based on the provided amplitude (or signal correlation) threshold, while shaded regions illustrate the extent of the error window, given the detector onset time and the automated processing parameters. The overlap between the two shaded regions indicates a positive detection. Derived through personal communication with J. Carmichael (LANL).

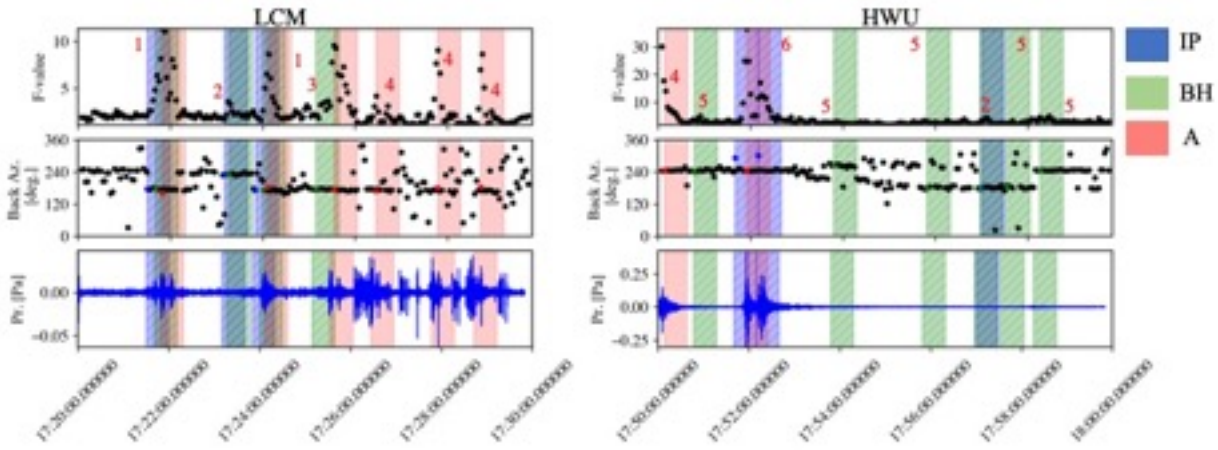

Figure 5-3. Example detection results compared using data from (a) LCM and (b) HWU. Top panel: F-value estimates from FK-processing. Middle Panel: back-azimuth estimates from FK processing. Bottom Panel: beamed signal waveforms. Red windows indicate analyst detection onset time and a 30 s window. Green windows indicate BH detection onset and a 30 s window. Green dots indicate BH detection backazimuth estimate. Blue windows indicate IP detection onset time + 30s window. Blue dots indicate IP detection backazimuth estimate. Six distinct detection comparisons are illustrated by the red numbers where: (1) indicates overlapping IP, BH and analyst detections; (2) indicates overlapping IP and BH detections; (3) indicates overlapping BH and analyst detections; (4) indicates an analyst detection only; (5) indicates a BH detection only; and (6) indicates an overlapping IP and analyst detection.

For the second hour, general trends can be identified (see Figure 5-4b). While a number of events are detected by all three methods at each station, a significant number of detections from each individual method do not overlap. For example, at HWU, BH detects 20 distinct signals, while IP detects 8 distinct signals and the analyst identified 11 distinct signals. There are only 3 common detections between IP, BH and the analyst; examples of these common detections are labeled with the number 1 in Figure 5-3. Signals are high amplitude with a consistent backazimuth trend. As seen in the example figures, signals identified by all three methods have F-values around 10 while signals confirmed by the analyst but missed by the detectors (labeled with number 4) have F-values ranging from 2-8. Additionally, signals confirmed by the analyst but missed by the detectors are generally shorter in duration. Signals identified by both detectors but not the analyst (labeled with number 2) are very low in amplitude and have little to no increase in amplitude above the background noise.

Detection results from each method can be evaluated in terms of detector recall and precision [48], where detector recall is defined as the proportion of true objects found from all true samples,

$$recall = \frac{\#of truepositives}{alltruesamples} \tag{5.1}$$

and detector precision is defined as the proportion of true objects from all found objects,

$$precision = \frac{\#of truepositives}{\#of truepositives + \#of falsepositives} \tag{5.2}$$

In these equations, true samples are signals identified by the analyst, true positives are automatic detections that were also identified by the analyst and false positives are automatic detections that were not identified by the analyst. Recall can be interpreted as the percentage of total relevant events correctly identified by each detector, while precision can be interpreted as the percentage of detections that are relevant. For comparisons, precision can be seen as a measure of detection quality while recall is a measure of detection quantity.

Detector recall and precision values for each hour are presented in Table 5-2. Metrics are incomplete for the first hour due to the low number of detections. For the second hour, metrics vary across the network. Recall is higher for BH at all stations except NOQ, while precision is higher for IP at all stations except FSU and NOQ where values are equal. Following the definitions above, these results indicate that BH identifies more analyst detections with the trade-off of a high number of false noise-related detections, while IP identifies fewer analyst detections with fewer false noise-related detections.

Evaluation of detections utilizing these terms inherently assumes that an overlap with an analyst detection is a true positive while any other detection is a false positive; however, the significant overlap between signals that were detected by both IP and BH but not the analyst, the light blue components of Figure 5-4, suggests that definition of “true positive” should be revisited. When true positive detections are considered to be either detections that overlap with analyst detections or events that were detected by both IP and BH, but not the analyst, precision and recall values for both IP and BH increase, as illustrated in Table 5-3. While overall values, expressed as percentages, increase, relationships between the two detectors remain the same. Recall is higher

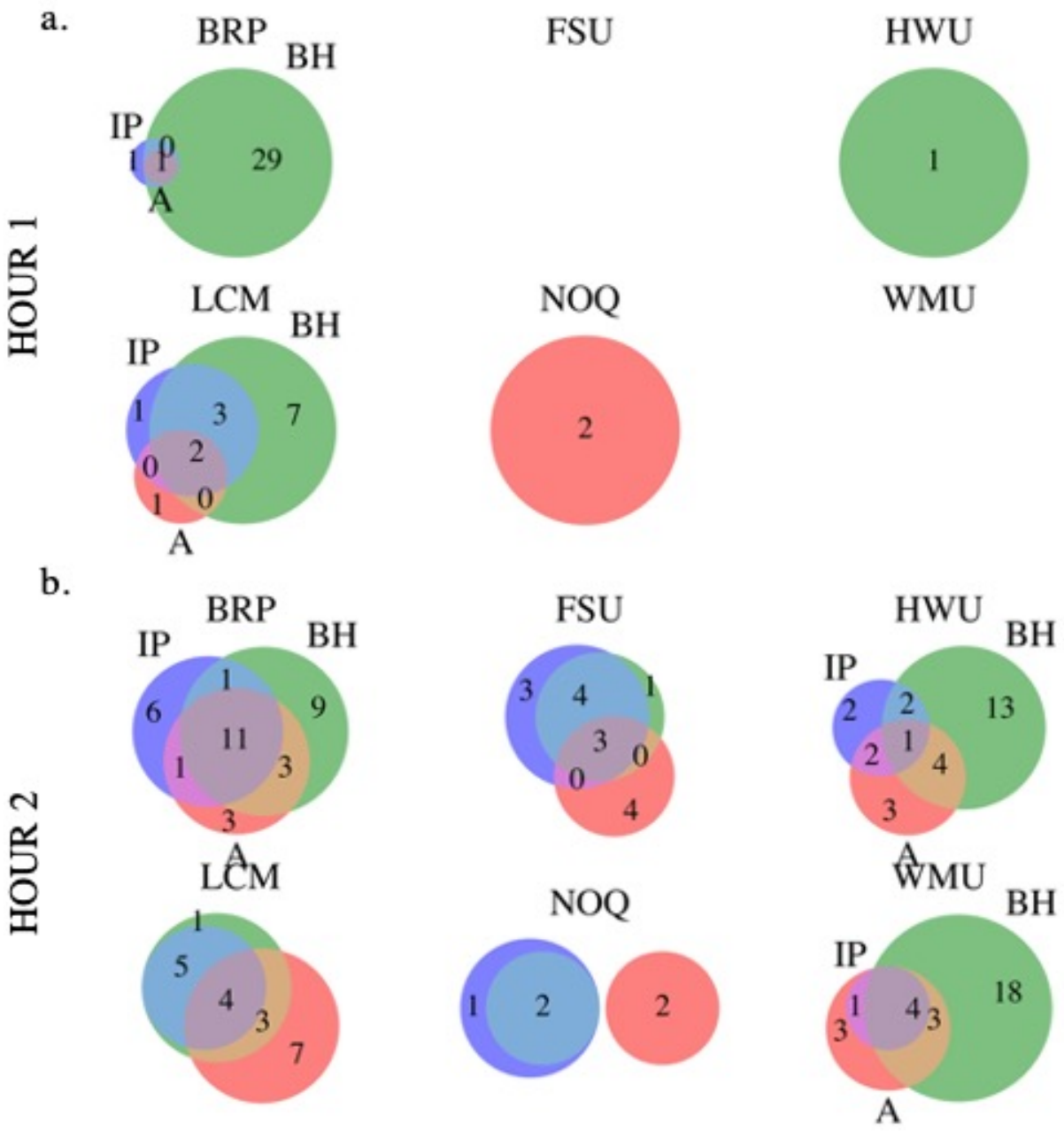

Figure 5-4. Comparisons between automatic and analyst detection catalogs for Hour 1 (07:00 UTC) (a) and Hour 2 (17:00 UTC) (b). Values inside circles indicate the number of detections in each subset where blue are IP detections, green are BH detections and red are analyst detections.

for BH at all stations except NOQ and FSU where recall values are equal. Precision is higher for IP at all stations except FSU and NOQ.

Hourly noise estimates presented in Figure 5-5 as averaged Power Spectral Density (PSD) curves [47] for the two data hours illustrate that differences in detection numbers across these two example hours are likely related to noise across the arrays. For all stations except NOQ, array noise within the band of interest, 1-5 Hz, is higher during the second hour. The first hour of data represents 12:00 am local time while the second hour of data represents 10:00 am local time. The discrepancies in both background noise and detection numbers suggest that detections within this example dataset may be driven by anthropogenic sources related to activity during the workday, similar to observations noted within [37] and [39]. Additional comparisons of hourly detections across the full three month dataset would further explore these conclusions.

Table 5-2. Recall and precision rates for each method.

		IP recall	IP Precision	BH Recall	BH Precision
HR 1	BRP	100%	100%	100%	3%
	FSU	-	-	-	-
	HWU	-	-	-	-
	LCM	67%	33%	67%	17%
	NOQ	-	-	-	-
	WMU	-	-	-	-
HR2	BRP	67%	63%	78%	56%
	FSU	43%	30%	43%	38%
	HWU	36%	50%	45%	25%
	LCM	36%	50%	50%	47%
	NOQ	50%	33%	50%	50%
	WMU	45%	100%	64%	28%

Table 5-3. Updated recall and precision rates that include signals detected by both IP and BH as true positives.

		IP Recall	IP Precision	BH Recall	BH Precision
HR 1	BRP	100%	100%	100%	3%
	FSU	-	-	-	-
	HWU	-	-	-	-
	LCM	83%	83%	83%	42%
	NOQ	-	-	-	-
	WMU	-	-	-	-
HR2	BRP	68%	68%	79%	63%
	FSU	64%	70%	64%	88%
	HWU	46%	75%	54%	35%
	LCM	55%	100%	65%	93%
	NOQ	100%	75%	67%	100%
	WMU	45%	100%	64%	28%

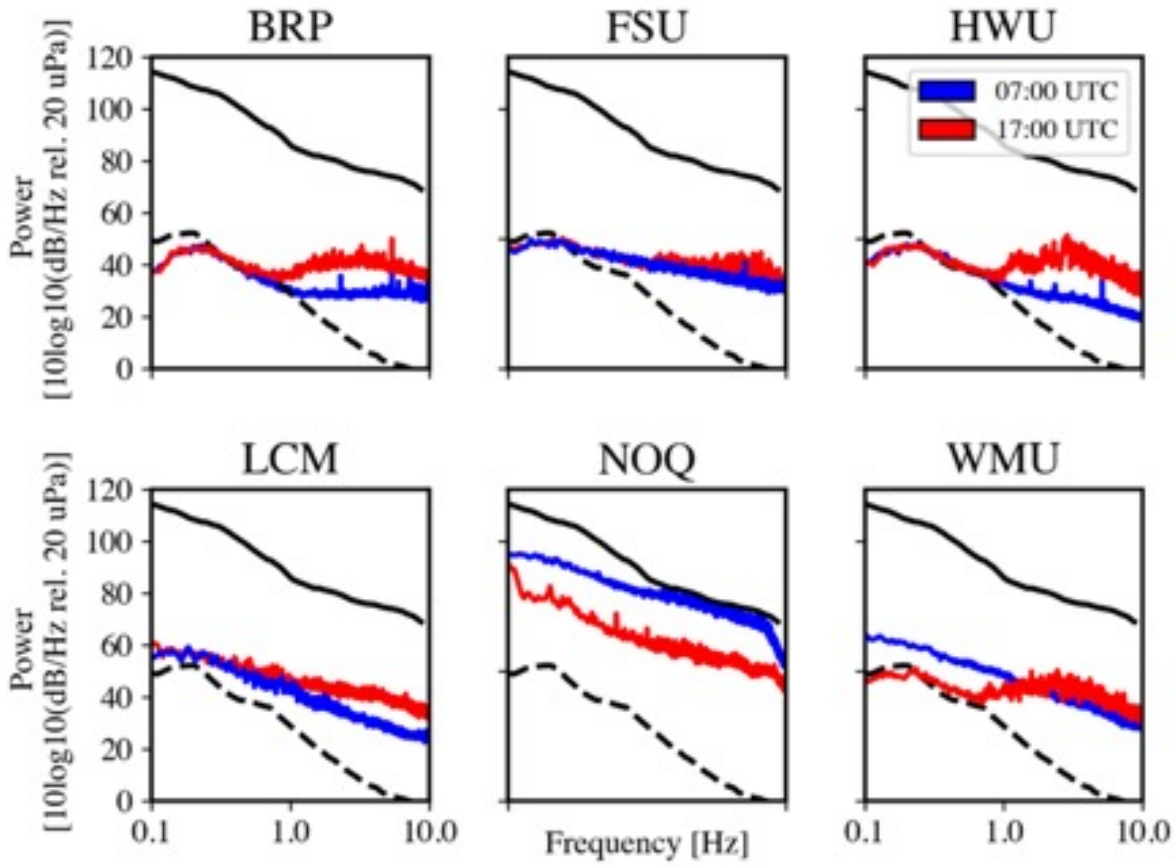

Figure 5-5. Average noise levels at each array for the two hours used for analysis on 2011-01-05. Solid black line indicates the IMS High Noise Model from [10] while dashed black line indicates the IMS Low Noise Model.

6. SYNTHETIC COMPARISONS

Given the conclusion that the majority of signals within the Western US region are due to anthropogenic sources, we construct a day-long (24 hr.) dataset consisting of 40 synthetic signals spaced approximately 30 minutes apart. This dataset provides the ability to control the signal type, duration and background noise level, ideally producing a more complete comparison of detector performance across a variety of noise conditions. Synthetic signals are constructed with a sampling rate of 20 Hz for a four-element microbarometer array with the station coordinates of BRP, where array elements are separated by 200 m. The synthetic data is comprised of both short (10-30 sec) and long (60-90 sec) duration signals, with both broad (0.5-5 Hz, for example) and narrow (0.5-0.7 Hz, for example) frequency bands. Each signal has a specified amplitude and direction of arrival (DOA). Signals are time-shifted according to the input parameters and the station location. Thus, when beamforming, the DOA and apparent velocity will correspond to the input parameters. Uncorrelated random pink noise, corresponding to a noise source with equal power across each octave, is added to each trace separately. This method was used to produce synthetic signals for benchmarking infrasound detectors based on the Hough-transform [7] and CLEAN beamforming [19].

We produce four distinct synthetic datasets with background noise levels of 0.01, 0.02, 0.05 and 0.1 Pa, referred to as relative noise levels 1, 2, 5, and 10, respectively, in order to explore detector performance in a variety of noise environments. These noise levels correspond to the amount of pink noise added to the background of each trace and alter the SNR of synthetic signals within the dataset based on the input amplitude (Figure 6-1a).

Figure 6-1b illustrates that an increase in background noise level corresponds to a decrease in SNR as characterized in the frequency domain. Detection utilizing data with relative background noise levels 1+2 represents detection under low noise environments where SNRs range from 10-60 dB/Hz relative to 20 μ Pa, while detection utilizing data with relative background noise levels of 5+10 represents detection in high noise environments where SNR drops below 10 dB/Hz relative to 20 μ Pa.

Processing parameters mirrored those utilized on the real data. A summary of results, in the form of number of detections identified through automated processing utilizing p-values of 0.01 is presented in Figure 6-2. These results summarize the total number of signals identified at each noise level, as well as determine if the two automated detectors consistently identify the same synthetic signals and evaluate the ability of each detector to identify short and long period signals. There are several significant conclusions that can be drawn from these comparisons. First, as seen in Figure 6-2a, IP identifies more signals than BH across all four relative noise levels. For example, when comparing detections identified from the dataset with a relative noise level of 5, IP identifies 34 detections while BH identifies 26 detections. This corresponds to IP identifying well above 80% of synthetic signals (dashed black line in Figure 6-2) and BH identifying well below.

Second, one would expect that automatic detectors would identify more signals during periods of low noise; however, results indicate that there is no direct relationship between relative noise level and overall detection numbers for either IP or BH. IP identifies the most signals at the highest noise level, while BH identifies the least signals at the highest noise level and the most signals at a relative noise level of 2 and 5. We presume this lack of correlation is related to the high SNR values of signals, even across the noise level of 10, SNR values remain between 1-10.

Due to lower number of BH detections, we increased the p-value for BH detection to 0.03. This parameter change significantly increased the number of BH detections to numbers comparable to IP detections, and brings successful detection above 80% for noise levels 1, 2 and 5. Overall BH detection numbers are still low for noise level 10, which indicates that optimization of algorithms or processing parameters is needed to successfully identify low SNR signals or signals in known high-noise environments. Additional processing could also utilize high noise levels to determine at which level IP detection ability begins to fail. The higher number of signals identified by IP over BH contrasts with results utilizing real data, indicates that higher detection numbers from BH may be related to the identification of persistent noise sources over true signals.

In Figure 6-2b we compare detector success rates for short duration (31 total) synthetic signals, and in Figure 6-2c we compare success rates for long duration (9 total) synthetic signals. Short duration success rates mirror overall detections results, where IP consistently identifies above 80% of signals. The use of a p-value of 0.03 significantly improves BH detection, raising successful short duration signal detection to above 80% of signals at noise levels of 2 and 5. The use of the 0.03 p-value additionally increases the number of long duration signals identified by BH, producing detection numbers equal to IP across the four noise levels. These results indicate that both detectors successfully identify long duration signals, and variability in detection numbers is driven by the successful detection of short duration, impulsive signals. Additionally, we note that the difference in overall number of successfully identified signals from IP and BH is likely due to the different way in which each algorithm uses p-value thresholds for identifying detections. The p-value in BH is an ensemble value and is used to determine the presence or absence of a detection based on both signal semblance and backazimuth estimates. The p-value in IP is used to threshold detections based on a modified F-distribution, given the background noise estimations. While both detectors utilize an adaptive mechanism to account for coherent noise, comparisons indicate that processing parameters are not equal across the two detectors and it may be difficult to produce a direct comparison between methodologies.

Overall detection numbers are supplemented by an evaluation of how often each detector identifies the same signal of interest; presented as Venn diagrams across p-values and noise levels in Figure 6-3. In this figure, blue indicates synthetic signals only identified by IP, green indicates synthetic signals only identified by BH and cyan indicates synthetic signals identified by both, as a function of noise level and p-value used for detection. The top row of Figure 6-3 compares detection datasets produced with a p-value of 0.01 across the four relative noise levels while the bottom row compares BH detection with a p-value of 0.03 and IP detection with a p-value of 0.01. As expected from direct detection numbers, IP detection with a p-value of 0.01 and BH detection with a p-value of 0.03 produce the most comparable results. In this case, each detector successfully identifies between 26-33 of the same signals across all four noise levels. Notably, the signals missed by both BH and IP are short duration consecutive signals that arrive within 60

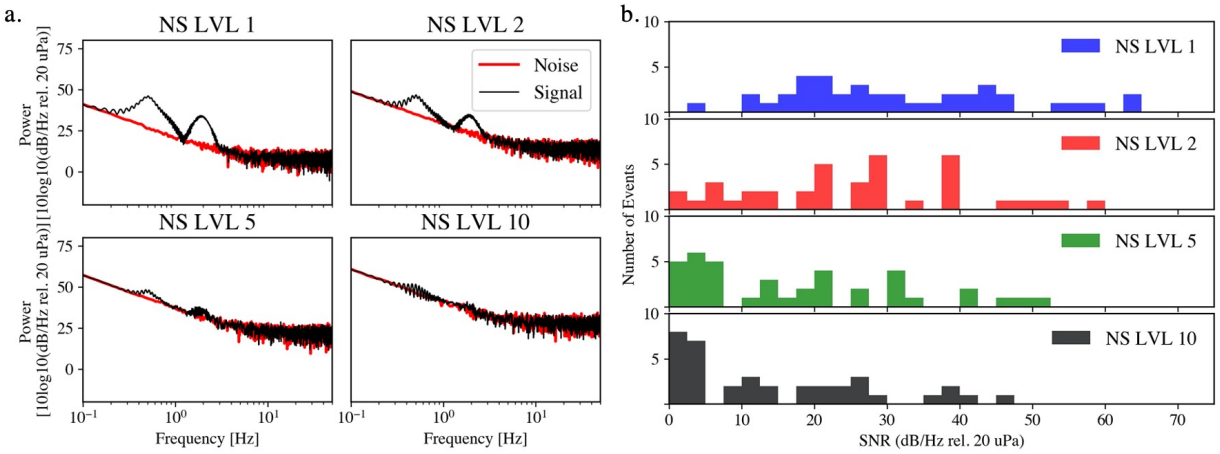

Figure 6-1. (a) Example windowed signal (black) and noise (red) comparisons across the four relative noise levels, (b) Event signal to noise ratios (SNR) across all four levels. SNR decreases as relative noise level increases.

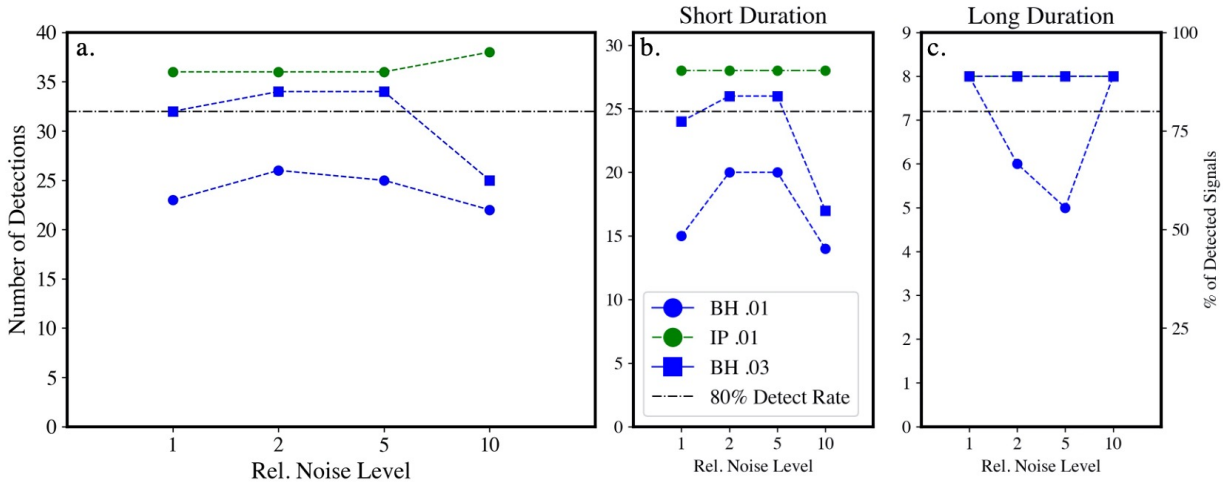

Figure 6-2. (a) Total number of detections (b) number of short duration detections, (c) number of long duration detections identified by each detector for each noise level, as a function of automatic detector and p-value (0.01, 0.03).

seconds and from the same direction as another synthetic signal. Examination of duration of detections suggests that both detectors envelope the repeating signals in a single long duration detection, indicating that both IP and BH may not perform well for identifying repeating short duration signals without additional parameter tuning or algorithm development.

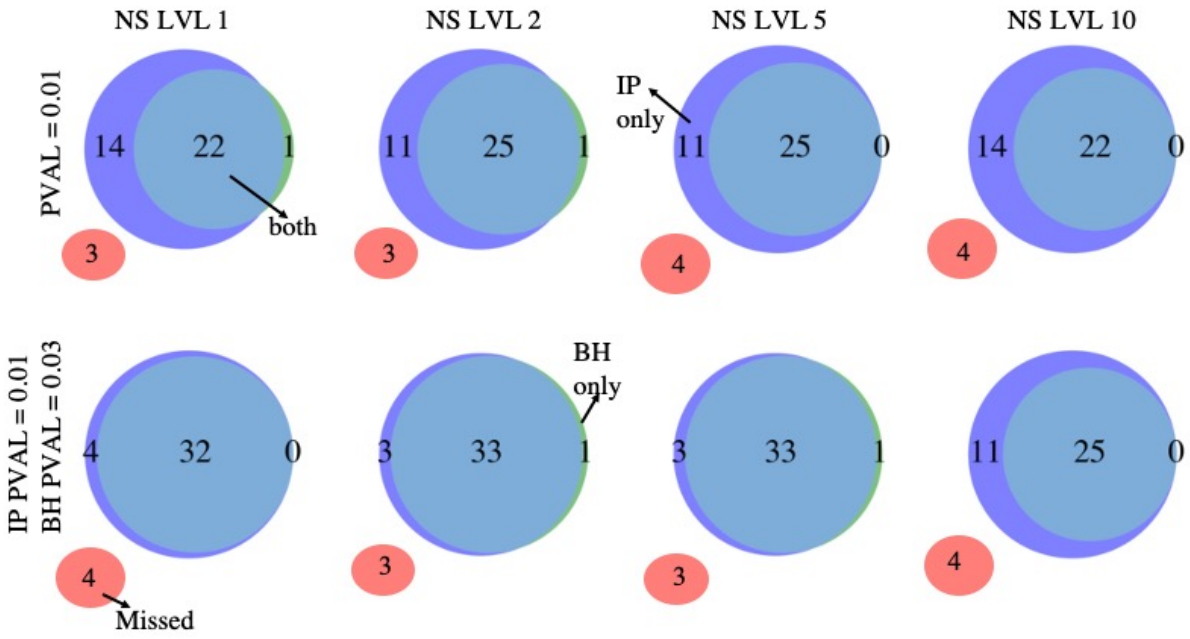

Figure 6-3. Comparisons of day-long synthetic data detection catalogs utilizing a p-value of 0.01 (top), and a combination of 0.01 and 0.03 (bottom) across the four relative noise levels (left to right: 1, 2, 5 and 10). Values inside circles indicate the number of detections.

7. RECALL AND PRECISION FOR THE SYNTHETIC DATASET

We further utilize these comparisons to evaluate detector performance in terms of recall and precision. Results are presented in Table 7-1 and offer an interesting comparison between the detectors. Both precision and recall are higher for IP, regardless of noise level. IP recall remains at 90% across all four noise levels, while altering the p-value to 0.03 improves BH recall from 55-65% to 62-85%. BH recall rates are best for detection across noise levels 2 and 5, utilizing a p-value of 0.03. BH recall declines significantly for detection across noise level 10, indicating that processing with the parameters utilized throughout this manuscript will not successfully identify short duration signals of interest in environments where either low signal amplitudes or high background noise reduce SNRs to less than 5.

BH precision remains at 100% across both p-values and all four noise levels, indicating that the detector does not falsely identify any noise as signals. IP precision rates are consistently 100%; however, at the highest noise levels IP falsely identifies noise as signals, which decreases the detector precision from 100% to 97%. Based on SNR calculations shown in Figure 6-1, this result indicates IP will perform well in high (SNR$<$5) noise environments, with a trade-off in increased false alarms. In contrast, BH will detect fewer signals, given the lower recall rates, but simultaneously identify fewer false signals. This result is significant for future work evaluating detection catalogs produced for both low SNR signals and high noise environments.

Table 7-1. Precision and recall for synthetic data processing.

		IP Recall	IP Precision	BH Recall	BH Precision
0.01	NS LVL 1	90%	100%	57.5%	100%
	NS LVL 2	90%	100%	65%	100%
	NS LVL 5	90%	100%	62.5%	100%
	NS LVL 10	90%	97%	55%	100%
0.03	NS LVL 1	-	-	80%	100%
	NS LVL 2	-	-	85%	100%
	NS LVL 5	-	-	85%	100%
	NS LVL 10	-	-	62.5%	100%

8. DISCUSSION AND CONCLUDING REMARKS

We utilize two detection datasets, one from observational data recorded by a network of infrasound arrays in the western US (Figure 3-1), and one from a set of synthetic signals embedded in realistic background noise at a single station to help assess methodologies for comparing outputs from different automatic infrasound detectors. Previous work [1, 38] notes the difficulty in directly comparing results from signal detectors. Here, we utilize a combination of overall detection numbers and precision and recall rates in order to estimate detector performance across a variety of noise environments. Results represent the first step towards developing procedures for assessing infrasound data processing algorithms across the full pipeline of signal detection, association and eventual event localization [11, 33, 37].

Comparisons between automatic and analyst detections produced during DNE18 identify several important trends. First, defining events identified by both automatic detectors but not analyst review as true positives increase both precision and recall rates across the network. This result suggests that automatic detectors may be able to identify small amplitude, low SNR events that cannot be identified by analyst review, and that ensembles of detectors can potentially be used to provide a form of ground-truth. When examining general detector characteristics for the DNE18 results, IP is more precise (avg. precision of 86.5% compared to BH avg. precision of 70%) while BH has only slightly better recall (avg. recall of 67% vs IP avg. recall of 64.5%). In practical terms, IP produces a higher quality detection dataset with fewer false alarms while BH is able to overall detect more signals at the cost of introducing a larger number of false alarms into the detection dataset. These results should naturally lead towards event catalog comparisons and a better understanding of consistent nuisance sources, such as anthropogenic noise, with the goal of eliminating "clutter" across event bulletins [40].

Interpretation of detector results are complimented by processing of a day-long synthetic dataset with 40 signals embedded in realistic background noise. Automatic detection with a variety of p-values and noise levels complements the analyst-centric approach taken. Use of a pure synthetic dataset provides a controlled experiment to assess the number of signals identified by automated detectors but lacks insight into the performance of detectors under realistic noise environments. Our hybrid approach combines embedding synthetic signals with realistic pink noise across four relative noise levels, corresponding to detection of signals with SNRs ranging from 70-1.

The IP detector with a p-value of 0.01 and BH detector with a p-value of 0.03 produce the most comparable results, successfully identifying between 26-33 of the same signals across all four noise levels. Estimates of detector recall and precision document that both precision and recall rates are higher than rates from the real dataset, which is to be expected from a purely synthetic dataset. IP has overall higher recall rates while precision rates are equal for both detectors. BH recall is best for detection across noise levels 2 and 5, suggesting that detector performance is reduced at higher noise environments where lower signal amplitudes reduce SNRs to below 10. In

general, both detectors fail to identify repetitive, short duration signals that simulate multiple arrivals from a similar source; instead, both methodologies associate the repeating signals as a singular, long duration detection. Similarly, BH fails to identify both short duration, narrowband and low amplitude signals across the four noise levels, which is likely driven by backazimuth variance across the detection window, leading to missed detections. These failures represent specific signal types and scenarios that could be targeted for further detector development.

We conclude that differences in detection numbers are likely related to how each detector uses the p-value threshold for processing, as BH performance improved as the p-value was increased and could ideally improve more significantly with further parameter optimization. Based on results from the synthetic dataset, additional processing of the DNE18 dataset utilizing a p-value of 0.03 may provide a more realistic comparison between the two detectors. However, a change in the p-value would increase the large number of repeating or clutter detections identified by BH. Discrepancies in detection backazimuths indicate that there are likely difference in the ways that each detector builds and evaluates the background noise distribution, however such evaluation is beyond the scope of this study. Despite the trade-off, high rates of precision and recall for both processing pipelines indicate that the utilization of both IP and BH as automated infrasound event processing algorithms is promising for large datasets.

The failure of each detector to identify short-duration, repetitive signals originating from a similar source motivates targeted algorithm improvements. The general consistency of results across noise levels indicates that both IP and BH perform well under a variety of noise environments. However, it is important to note that this interpretation is based on analyst picks from only two hours, and the day-long synthetic dataset which could be expanded upon. We note that the use of correlated signals and uncorrelated pink noise within the synthetic dataset represents an idealized scenario designed for the success of a power-based detector, such as the original F-detector or the Adaptive F-Detector. In addition, based on the design of the Multivariance Adaptive Learning Detector, we expect it to miss signals where the window over which the variance in the backazimuth is small. This expectation is substantiated by results indicating IP identifies more synthetic signals than BH across all four noise levels. Results could be refined with more extensive analyst review or longer synthetic datasets that combine purely synthetic signals with a variety of coherent and incoherent background noise sources such as microbaroms, wind turbines, or high incoherent wind noise. At present, there are few methods to introduce realistic noise into synthetics. These limitations motivated the use of both realistic and synthetic signals as a means of developing comparisons across ideal and not-ideal noise environments. The datasets within this manuscript are intended to serve as a benchmark for further improvements to automatic infrasonic signal detectors. We have documented limitations of current infrasound detection algorithms and intend to use lessons from this study as a path forward towards improved automated detection capability.

Finally, we note that an implicit assumption in this assessment is that the analyst has superiority over the algorithm, which may not always be the case. The impact of this assumption is partly addressed by defining additional true events within our catalogs as those identified by both IP and BH. These results demonstrate that multiple detector comparisons illuminate additional ‘true’ signals, resulting in increases in detector precision and recall. We suggest that reliance on analyst review as ‘Ground Truth’ may be inappropriate for evaluating detector performance with regards

to low SNR signals and a hybrid approach utilizing both an analyst and independent detection catalogs may enhance results.

Our results utilizing catalogs produced by both real and synthetic signals indicate that establishing a direct comparison between different signal detectors is difficult, particularly when each detector utilizes a unique set of processing parameters. Our comparisons utilized parameters substantiated by earlier work within the study region, but it is possible that a better combination exists. The synthetic dataset represents a first step towards developing a standard for tuning automated detector parameters; however, such extensive tuning was beyond the scope of this paper. We suggest that to investigate this problem, many parameters should be tested, and the resulting accuracies compared. An automated grid search of tuning parameters may illuminate further optimal combinations for the processing of regional networks such as the UU infrasound network. We additionally suggest that future detector comparisons focus first on optimizing parameters for each individual detector across the dataset of interest, and then establishing comparisons following the methodologies established within this study.

9. DATA AND RESOURCES

Infrasonic waveform data utilized during DNE18 is available from IRIS using the network code ‘YJ’ [26]. Data from NOQ is part of the University of Utah seismic network and is available using the station code ‘UU’.

10. AUTHOR CONTRIBUTIONS

FD led the analysis and writing of the manuscript. SA processed the BH datasets. GA produced the synthetic dataset utilizing input from FD. SJA provided oversight and information regarding detector development. All authors provided critical feedback and helped shape the research, analysis, and structure of the manuscript.

REFERENCES

[1] S. Arrowsmith. False alarms and the IMS infrasound network: Understanding the factors influencing the creation of false events. *Geophysical Journal International*, 215(2):1322–1337, 11 2018.

[2] S. Arrowsmith, G. Euler, O. Marcillo, P. Blom, R. Whitaker, and G. Randall. Development of a robust and automated infrasound event catalogue using the International Monitoring System. *Geophysical Journal International*, 200(3):1411–1422, 1 2015.

[3] S. Arrowsmith, D. Norris, R. Whitaker, and D. Anderson. Sources of error model and progress metrics for acoustic/infrasonic analysis: Location estimation. *Pure and Applied Geophysics*, 171(3-5):587–597, 2014.

[4] S. Arrowsmith, S. Tarin, and S. Albert. Bloodhound 0 . 8 : A python package for infrasound data analysis: Sandia National Laboratories. (August), 2018.

[5] S. Arrowsmith, R. Whitaker, C. Katz, and C. Hayward. The F-detector revisited: An improved strategy for signal detection at seismic and infrasound arrays. *Bulletin of the Seismological Society of America*, 99(1):449–453, 2 2009.

[6] S Arrowsmith, R. Whitaker, S. Taylor, R. Burlacu, B. Stump, M. Hedlin, G. Randall, C. Hayward, and D. ReVelle. Regional monitoring of infrasound events using multiple arrays: application to Utah and Washington State. *Geophysical Journal International*, 175(1):291–300, 10 2008.

[7] G. Averbuch, J. Assink, P. Smets, and L. Evers. Extracting low signal-to-noise ratio events with the Hough transform from sparse array data. *Geophysics*, 83(3):WC43–WC51, 5 2018.

[8] M. Begnaud, C. Gammans, E. Syracuse, and J. MacCarthy. Testing the effects of velocity models for seismic location in the DNE18 virtual experiment. *AGUFM*, 2019:S11E–0419, 2019.

[9] P. Blom, O. Marcillo, and G. Euler. InfraPy: Python-based signal analysis tools for infrasound. Technical report, Los Alamos, NM (United States), 5 2016.

[10] J. Bowman, G. Shields, and M. O'Brien. Infrasound station ambient noise estimates and models 2003–2006. *Infrasound Technology Workshop, Brasilia, Brazil, November 2-6, 2009*, 2009.

[11] D. Brown, C. Katz, R. Le Bras, M. Flanagan, J. Wang, and A. Gault. Infrasonic signal detection and source location at the prototype International Data Centre. *Pure and Applied Geophysics*, 159(5):1081–1125, 2002.

[12] D. Brown, R. Whitaker, B. Kennett, and C. Tarlowski. Automatic infrasonic signal detection using the Hough transform. *Journal of Geophysical Research*, 113(D17):D17105, 9 2008.

[13] Y. Cansi. An automatic seismic event processing for detection and location: The P.M.C.C. method. *Geophysical Research Letters*, 22(9):1021–1024, 5 1995.

[14] J. Carmichael, M. Begnaud, C. Rowe, F. Dannemann, and J. MacCarthy. Establishing a baseline signal detection capability for the 2018 Dynamic Networks data processing and analysis experiment (DNE18). *AGUFM*, 2018:S53E–0468, 2018.

[15] J. Carmichael, R. Nemzek, S. Arrowsmith, and K. Sentz. Fusing geophysical signatures of locally recorded surface explosions to improve blast detection. *Geophysical Journal International*, 204(3):1838–1842, 3 2016.

[16] I.-Y. Che, J. Park, T.S. Kim, C. Hayward, and B. Stump. On the use of a dense network of seismo-acoustic arrays for near-regional environmental monitoring. In *Infrasound Monitoring for Atmospheric Studies*, pages 409–448. Springer International Publishing, 10 2019.

[17] F. Dannemann and O. Marcillo. Schema for the LANL infrasound analysis tool, InfraPy. Technical report, Los Alamos, NM (United States), 4 2017.

[18] F. Dannemann Dugick, B. Stump, P. Blom, O. Marcillo, C. Hayward, and S. Arrowsmith. Benchmarking infrasonic signal detection performance utilizing the Adaptive F-Detector. *Journal of the Acoustical Society of America*, 148(5), 2020.

[19] O. den Ouden, J. Assink, P. Smets, S. Shani-Kadmiel, G. Averbuch, and L. Evers. CLEAN beamforming for the enhanced detection of multiple infrasonic sources. *Geophysical Journal International*, 221(1):305–317, 4 2020.

[20] L. Evers and H. Haak. Listening to sounds from an exploding meteor and oceanic waves. *Geophysical Research Letters*, 28(1):41–44, 1 2001.

[21] T. Farges and E. Blanc. Characteristics of infrasound from lightning and sprites near thunderstorm areas. *Journal of Geophysical Research: Space Physics*, 115(A6):n/a–n/a, 6 2010.

[22] A. Fisher. *West Texas array experiment: Noise and source characterization of short-range infrasound and acoustic signals, along with lab and field evaluation of Intermountain Laboratories infrasound Microphones*. PhD thesis, 2013.

[23] D. Green and A. Nippress. Infrasound signal duration: the effects of propagation distance and waveguide structure. *Geophysical Journal International*, 216(3):1974–1988, 3 2019.

[24] D. Hart. Automated infrasound signal detection algorithms implemented in MatSeis-InfraTool. Technical report, Albuquerque, NM (United States), 2004.

[25] D. Hart. Evaluation of Inter-Mountain Labs infrasound sensors. Technical report, Albuquerque, NM (United States), 2007.

[26] C. Hayward. Infrasound from earthquakes: Signal characteristics and depth discrimination. 2010.

[27] C. Hetzer and R. Waxler. Development and application of first- and second-pass clutter-reduction techniques for nuclear blast detection at infrasound stations. *AGUFM*, 2009:S31B–1720, 2009.

[28] S. Kay. Fundamentals of statistical signal processing volume ii detection theory. Technical report, 1998.

[29] M. Landès, L. Ceranna, A. Le Pichon, and R. Matoza. Localization of microbarom sources using the IMS infrasound network. *Journal of Geophysical Research: Atmospheres*, 117(D6), 3 2012.

[30] A. Le Pichon. Infrasound from ocean waves observed in Tahiti. *Geophysical Research Letters*, 31(19):L19103, 2004.

[31] A. Le Pichon, J. Vergoz, P. Herry, and L. Ceranna. Analyzing the detection capability of infrasound arrays in Central Europe. *Journal of Geophysical Research*, 113(D12):D12115, 6 2008.

[32] O. Marcillo, S. Arrowsmith, M. Charbit, and J. Carmichael. Infrasound signal detection: Re-examining the component parts that makeup detection algorithms. In *Infrasound Monitoring for Atmospheric Studies: Challenges in Middle Atmosphere Dynamics and Societal Benefits: Second Edition*, pages 249–271. Springer International Publishing, 10 2018.

[33] O. Marcillo, P. Blom, G. Euler, J. MacCarthy, J. Park, B. Stump, C. Hayward, and F. Dannemann. On the development and testing of a database-centric pipeline for the analysis of a regional infrasound network. *AGUFM*, 2015:S51C–2691, 2015.

[34] R. Matoza, M. Landès, A. Le Pichon, L. Ceranna, and D. Brown. Coherent ambient infrasound recorded by the International Monitoring System. *Geophysical Research Letters*, 40(2):429–433, 1 2013.

[35] S. McComas, C. Hayward, M. Pace, C. Simpson, M. McKenna, and B. Stump. Infrasound monitoring in non-traditional environments. *The Journal of the Acoustical Society of America*, 144(6):3201–3209, 2018.

[36] M. McKenna, B. Stump, and C. Hayward. Effect of time-varying tropospheric models on near-regional and regional infrasound propagation as constrained by observational data. *Journal of Geophysical Research: Atmospheres*, 113(D11), 2008.

[37] J. Park, S. Arrowsmith, C. Hayward, B. Stump, and P. Blom. Automatic infrasound detection and location of sources in the Western United States. *Journal of Geophysical Research: Atmospheres*, 119:7773–7798, 2014.

[38] J. Park, C. Hayward, C. Zeiler, S. Arrowsmith, and B. Stump. Assessment of infrasound detectors based on analyst review, environmental effects, and detection characteristics. *Bulletin of the Seismological Society of America*, 107(2):674–690, 4 2017.

[39] J. Park and B. Stump. Seasonal variations of infrasound detections and their characteristics in the western US. *Geosciences Journal*, 19(1):97–111, 3 2015.

[40] J. Park, B. Stump, C. Hayward, S. Arrowsmith, I-Y. Che, and D. Drob. Detection of regional infrasound signals using array data: Testing, tuning, and physical interpretation. *The Journal of the Acoustical Society of America*, 140(1):239–259, 7 2016.

[41] C. Pilger, L. Ceranna, J. Ross, J. Vergoz, A. Le Pichon, N. Brachet, E. Blanc, J. Kero, L. Liszka, S. Gibbons, T. Kvaerna, S. Näsholm, E. Marchetti, M. Ripepe, P. Smets, L. Evers, D. Ghica, C. Ionescu, T. Sindelarova, Ben H., and P. Mialle. The European infrasound bulletin. *Pure and Applied Geophysics*, 175(10):3619–3638, 2018.

[42] R. Sanderson, R. Matoza, D. Fee, M. Haney, and J. Lyons. Remote detection and location of explosive volcanism in Alaska with the EarthScope Transportable Array. *Journal of Geophysical Research: Solid Earth*, 125(4):1–23, 2020.

[43] R. Shumway. On detecting a signal in n stationarily correlated noise series. *Technometrics*, 13(3):499–519, 1971.

[44] B. Stump, M. Jun, C. Hayward, J. Jeon, I.Y. Che, K. Thomason, M. House, and J. McKenna. Small-aperture seismo-acoustic arrays: Design, implementation, and utilization. *Bulletin of the Seismological Society of America*, 94(1):220–236, 2004.

[45] K. Walker, R. Shelby, M. Hedlin, C. de Groot-Hedlin, and F. Vernon. Western U.S. infrasonic catalog: Illuminating infrasonic hot spots with the USArray. *Journal of Geophysical Research*, 116(B12):B12305, 12 2011.

[46] J. Webster, P. Blom, and F. Dannemann Dugick. Lanl-seismoacoustics/infrapy v0.3.4. 1 2021.

[47] P. Welch. The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[48] C. Young, S. Teich-McGoldrick, K. Aur, M. Begnaud, J. MacCarthy, R. Stead, J. Carmichael, S. Ruppert, J. Gaylord, B. Schrom, J. Mendez, and K. Koper. The 2018 Dynamic Networks Data Processing and Analysis Experiment (DNE18). *AGUFM*, 2018:S53E–0469, 2018.

[49] R. Zhou, T. Kim, R. Burlacu, B. Stump, C. Hayward, Z. Yang, Y. Chen, R. Hermann, S. Arrowsmith, and K. Pankow. Seismic observations and interpretation in NE China, infrasound observations and interpretation in Utah. Technical report, SOUTHERN METHODIST UNIV DALLAS TX, 2010.

DISTRIBUTION

Email—Internal [redacted]

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.