Enhancing Infrasonic Signal Detection and Event Localization Through Advanced Computational Methods

Fransiska Dannemann Dugick
fdannemann@smu.edu

Follow this and additional works at: https://scholar.smu.edu/hum_sci_earthsciences_etds

Part of the Geophysics and Seismology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Earth Sciences at SMU Scholar. It has been accepted for inclusion in Earth Sciences Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.
ENHANCING INFRASONIC SIGNAL DETECTION AND EVENT LOCALIZATION THROUGH ADVANCED COMPUTATIONAL METHODS

Approved by:

Dr. Brian Stump
Albritton Chair of Geological Sciences

Dr. Heather DeShon
Professor

Dr. Robert Gregory
Professor

Dr. Chris Hayward
Director, Geophysics Research Program

Dr. Philip Blom
Los Alamos National Laboratory

Dr. Omar Marcillo
Oak Ridge National Laboratory
ENHANCING INFRASONIC SIGNAL DETECTION AND EVENT LOCALIZATION THROUGH ADVANCED COMPUTATIONAL METHODS

A Dissertation Presented to the Graduate Faculty of the Dedman College Southern Methodist University

in

Partial Fulfillment of the Requirements for the degree of Doctor of Philosophy with a Major in Geophysics

by

Fransiska Dannemann Dugick

B.A., Geology, Colorado College M.S., Geophysics, Southern Methodist University

May 15, 2021
ACKNOWLEDGMENTS

This dissertation represents four years of hard work, dedication and personal growth that would not have been possible without the help and support of numerous people. First and foremost, I thank my academic advisor, Dr. Brian Stump. I am forever grateful for the opportunity to earn a MS and PhD with him, for the guidance, research support, unwavering faith in my abilities, mentorship towards my career, adherence towards my deadlines and the constant time spent on weekends or after hours working through edits and developing my dissertation. Thank you to my committee members Drs. Chris Hayward, Heather DeShon and Robert Gregory for your time, enthusiasm, encouragement, critique and challenges throughout the dissertation process. Thank you to my bonus committee members Drs. Stephen Arrowsmith and Joshua Carmichael for your support and research expertise.

It has been my privilege to complete the majority of my PhD work as a simultaneous Graduate Research Assistant on the Seismoacoustics team at Los Alamos National Laboratory. I am grateful to my research mentors and external committee members Drs. Omar Marcillo and Philip Blom for the opportunity, as well as their support, encouragement and expertise. I thank the rest of the SA team, particularly Drs. Mike Begnaud and Neill Symons for financial support that funded the chapters within this document. Thank you to Drs. Charlotte Rowe and Cathy Snelson for your mentorship as women in Global Security. Thank you to Jeremy Webster for your dog-sitting, Python expertise and always offering a laugh right when I needed it.

I joined the Geophysical Detection Programs team at Sandia National Laboratories during the home stretch of my degree completion. I thank my manager Steve Vigil for the opportunity and his support towards my professional development. Thank you to my new
teammates Dr. Daniel Bowman, Sarah Albert, Dr. Andrea Conley and Alex Sinclair for your commitment to cross-cutting research.

Throughout my time as a PhD student, I have had the opportunity to attend and present at a number of national and international scientific meetings and conferences. I am grateful for the financial support from Los Alamos to make this travel possible and the international community of colleagues I have met as a result. In particular, I thank Dr. Siddharth Krishnamoorthy for his encouragement and hot Twitter takes. I served as an American Geophysical Union Seismology Section Student Representative for three years and am grateful to the executive leadership of the organization for the opportunity. Thank you to Dr. Suzan van der Lee for your friendship. This volunteer position broadened my personal network and graduate student community. I thank a large number of these students including Monique Holt, Hank Cole, Dr. Alex Iezzi, Jordan Bishop, Dr. Devon Cole, Liam Toney, Michelle Dunn, Sydney Dybing, Dr. Elizabeth Berg and Nate Stevens for being a friendly face at a large scientific meeting and for offering peer review, coding support or reference suggestions. Thank you to my SMU grad student network: Sarah McComas, Dr. Kimber deGrandpre, Erin Summerlin-Donofrio and Anastasia Fedotova for your perspective and commiseration, and for keeping me company during the solitary writing process. Thank you to Stephanie Schwob who keeps the Earth Science department running. Thank you to Dr. Junghyun Park for your mentorship and research assistance.

To my past and present dogs, Dunes, Shiner and Otis, who bring me so much joy and love. To my friends who have become family: thank you for the weddings and bachelorette weekends, the Zoom bakeoffs, the Peloton rides, the group triathlons, the requests to read my research and for weathering the ebbs and flows of lifelong friendships. To my husband Brannon, thank you for everything. You are my rock; I could not have done this without you. We survived four years of commuting and nearly a year of pandemic-driven working from home. I cannot wait to see where we go from here. I am blessed to have married into a great
family; I am grateful to my in-laws for their love, help moving five times in six years and unwavering support. To my brother Peter, thanks for being my IT support and my personal Python reference. To my parents - thank you for challenging me, for showing me that hard work paid off and for supporting me in everything I wanted to do. Thank you for inspiring a love of the outdoors from an early age. Thank you for showing me that there is no such thing as bad weather, just bad clothes. Mom – thank you for your ‘unsolicited advice’ that always seems to come right when I need it. Thank you for your steadiness throughout the storms of life. Dad – there are no words to thank you for all that you have given me. Thank you for demonstrating the importance of relationships and how to maintain lifelong friendships. Thank you for showing me that anything is possible. Thank you for your postcards, your newspaper article cutouts, your love and your laughter.

This dissertation consists of three original works, of which I am the first author. I conducted the majority of the research, data analysis and interpretation, and writing for each. As part of my work at LANL, I contributed to the development and recent open-sourcing of an infrasound processing pipeline, Infrapy. The algorithms and packages within Infrapy were used for the data processing, visualization and analysis contained within this document. Data from chapters 1, 2 and 3 is available at the Incorporated Research Institutions for Seismology (IRIS) under the network code ‘YJ’. Synthetic waveforms from chapter 2 were constructed by Gil Averbuch at Southern Methodist University. Sarah Albert at Sandia National Laboratories contributed the Bloodhound detection dataset utilized in chapter 2. Brian Stump, Philip Blom, Chris Hayward, Omar Marcillo, Stephen Arrowsmith, Joshua Carmichael, Sarah Albert and Gil Averbuch read the manuscripts and contributed edits and constructive criticism.
The installation of worldwide infrasound and seismo-acoustic networks at both global and regional scales necessitates automated techniques and algorithms for accurate and efficient data processing and analysis. Signals recorded across the networks originate from a number of natural and anthropogenic sources. Data processing efforts focus on the separation of signals of interest from background noise, followed by the identification, or detection of signals of interest. Once signals are identified, association and location processing produces estimates of a signal’s source. This dissertation focuses on the evaluation of automated processes for identifying and locating sources of interest.

Chapter two applies two state-of-the-art automated infrasonic signal detectors to real and synthetic waveform data. Comparisons between the two detectors are produced across a variety of background noise conditions. The first detector, the Adaptive F-Detector, accounts for coherent noise across an array 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. The second detector, the multivariate adaptive learning detector, applies three distinct time windows to adaptively alter detection thresholds in order to account for the changing background noise environment. Performance is ultimately quantified in terms
of recall and precision through validation with an analyst-derived catalog, where recall is the proportion of true objects found from all true samples and precision is the proportion of true objects from all found objects. Comparisons using real waveform data document similar precision and recall rates for the two detectors of between 45-79% and 28-100% across the network, respectively, indicating that performance across both detectors is nearly equal. Comparisons against the synthetic catalog indicate that both automated detectors identify long period signals with the same success rate, while successful identification of short period signals varies based on the detector methodology and background noise level. One detector, the Adaptive F-Detector, successfully identifies most short period signals and performs well across all background noise levels, motivating further study of the detector success and failure points in chapter three. Conclusions from this chapter lead to recommendations for future automated detector development.

Following the comparison of multiple signal detectors, in chapter three a single detection algorithm, the Adaptive F-Detector, is utilized to further assess factors that influence infrasound signal detection at regional networks. Signatures from repeating above-surface explosive events, denoted ground truth events due to the known location and origin time, are detected with the Adaptive F-Detector and by a human analyst. The number of automatic detections varies as a function of array distance from the source where the arrays closest and furthest from the source detect very few events (≈ 10/45). Analyst review adds additional GT detections at all stations, and particularly increases the number of detections within the dataset at stations near the source of interest. Successful detection increases to between 24-90% of GT events depending on the stations, as compared to success automatic detection of between 14-80% of the GT events. Situations when automated methodologies fail are evaluated through a combined background noise quantification, atmospheric propagation analyses and comparison of spectral amplitudes. Results indicate that detection capability is primarily related to station proximity to the source, driven by the atmospheric propagation of tropospheric and thermospheric infrasound signals. Detection capability can
also be related to background noise levels at individual stations. This analysis provides an estimate of detector performance across the network as well as a qualitative assessment of conditions that impact infrasound monitoring capabilities.

Finally, accompanying the evaluation of successful signal detection rates across the network, in chapter four the series of explosive events are used to evaluate recent improvements to infrasonic source localization methodologies. These improvements apply modeling-based predictions for atmospheric propagation to better constrain arrival celerities and enhance temporal and spatial localization estimates at regional distances between 58-410 km from the source. The network utilized in this study has a sub-optimal geometry where six out of seven arrays are located to the SE-SSE of the source and only one station offers azimuthal resolution in the NE; this station distribution may increase bias in location results. Locations are produced using three distinct signal celerity and backazimuth deviation models; (1) a generalized celerity model, derived from ray-tracing; (2) monthly Path Geometry Models (PGMs) for celerity and backazimuth based on station range and azimuth; and (3) a data-based empirical celerity model. Application of the PGMs both underestimates and overestimates signal celerities, leading to large errors in both spatial and temporal location across the GT dataset. These errors correspond to perceived improvements in spatial accuracy, where the 90% EE areas decrease, coupled with reductions in spatial precision, although the 90% Error Ellipses (EEs) do not contain the true GT location for 14/45 or 33% of the events in this dataset. Use of the generalized celerity model and the empirical UTTR celerity models produce spatial location estimates with similar accuracy, while use of the empirical model improves localization precision. Additionally, 90% EE for all 45 events produced with both the generalized celerity model and the empirical model contain the true source location within their bounds. Results indicate that atmospheric specifications need refinement in order to accurately predict infrasonic signal propagation at regional distances; unrefined propagation results lead to errors within both spatial and temporal location estimates. These errors are driven by the predominantly tropospheric arrivals at stations as the resolution of...
current atmospheric specifications cannot account for the variability of the atmosphere near the surface of the earth. Results additionally demonstrate that bias and accuracy in localization results are driven by both network detection capability as well as network design with large location errors related to azimuthal gaps between detecting arrays. This result indicates that the applied localization methodology weighs detection backazimuth observations more heavily than travel time estimates and further suggests that additional research into backazimuth corrections may offer the opportunity to improve spatial localization results.

These three chapters provide a detailed analysis of signal characteristics from surface explosions, leading to better constraints on the performance of automated detection and location algorithms applied in a regional network setting. Results, in the form of successful event detections and locations across the network are driven by the network geometry, where stations are located between 84-410 km from the source and 6/7 stations are located towards the SE-SSE of the source. Work presented in chapter two motivates further research into optimal techniques for direct comparisons of signal detectors, particularly focused on tuning precision and recall rates rather than producing direct parameter-based detections. Results from chapters two and three indicate that further detector development is necessary in order to reduce false detections related to coherent noise sources. Additionally, a deeper understanding of repeating noise sources at individual stations is needed. Results from chapters three and four demonstrate that an understanding of infrasound signal propagation dynamics is limited by the resolution of available atmospheric models; advances in atmospheric predictions will significantly improve the ability to model infrasound propagation, particularly in the lower atmosphere. Finally, location results in chapter four strongly suggest that network design introduces bias; this bias is driven by algorithms that more heavily weight backazimuth measurements over travel time measurements. A combination of improved infrasound networks with increased azimuthal resolution and refinement of location algorithms should continue to reduce errors and bias in event location results.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................. xiv
LIST OF TABLES ............................................................... xx

CHAPTER

1 Introduction ................................................................. 1

2 Methodologies for the Comparison of Automatic Infrasonic Signal Detectors .... 9
  2.1. Introduction ............................................................. 10
  2.2. Detector Overview ..................................................... 12
  2.3. Instrument Deployment and Details .................................. 15
  2.4. Direct Detector Comparisons ........................................ 15
    2.4.1. Virtual Experiment Detection Results .......................... 15
    2.4.2. Analyst Detection Catalog ...................................... 19
    2.4.3. Synthetic Comparisons .......................................... 27
      2.4.3.1. Recall and Precision for the Synthetic Dataset ............ 33
  2.5. Discussion and Concluding Remarks ................................ 34
  2.6. Data and Resources .................................................. 39
  2.7. Author Contributions ................................................ 39
  2.8. Acknowledgments ..................................................... 39

3 Factors Influencing Automatic Infrasonic Processing Performance ............. 41
  3.1. Introduction .......................................................... 42
  3.2. Dataset .................................................................. 46
  3.3. Methodology ............................................................ 48
    3.3.1. Automatic Detection Utilizing the Adaptive F-Detector ....... 48
4.5. Discussion .................................................. 114
4.6. Conclusions .................................................. 119
4.7. Author Contributions ....................................... 121
4.8. Acknowledgments ........................................... 122

5 Conclusions ..................................................... 123

APPENDIX

A Ground Truth Events and Associated G2S Profile using for ray tracing .......... 128
B Bayesian Infrasonic Source Localization Mathematical Framework ............. 131
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>UU/SMU seismo-acoustic network used for DNE18 processing. Blue triangles indicate infrasound arrays utilized within this study.</td>
</tr>
<tr>
<td>2.2</td>
<td>(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.</td>
</tr>
<tr>
<td>2.3</td>
<td>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.</td>
</tr>
<tr>
<td>2.4</td>
<td>Hourly detections for hours 7:00 and 17:00 at each array within the network for IP (blue), BH (green) and the analyst (red).</td>
</tr>
<tr>
<td>2.5</td>
<td>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).</td>
</tr>
</tbody>
</table>
2.6 Example detection results compared using data from (a) LCM and (b) HWU. Top panel: F-value estimates from FK-processing. Middle Panel: backazimuth 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. ....................................

2.7 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. ..........................

2.8 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 (Bowman et al., 2009) while dashed black line indicates the IMS Low Noise Model. ..........................

2.9 (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. ....................

2.10 (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). ....

2.11 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. ........................................

3.1 Locations of the seven, four-element acoustic arrays (solid blue triangles) used in this study within the SMU-UU seismo-acoustic network. The red star denotes UTTR, the location of the ground truth events. ......................

3.2 Number of automatically detected GT events using the AFD at each array within the network ..........................................................
3.3 (a) A noise PDF example capturing the seasonal (winter) noise trend at HWU. (b) Seasonal trends in median noise at three arrays across the network. Colored lines indicate the median values from the noise PDFs produced during hours of the day when GT events occurred. The dashed black line represents the IMS low noise model (Bowman et al., 2009) while the solid black line represents IMS high noise model. Spring [red]: March, April, May. Summer [blue]: June, July, August. Fall [green]: September, October, November. ...................................................... 55

3.4 Full noise models, represented at 5% (top) and 95% (bottom) noise levels at each station during the spring (red), summer (green) and fall (blue). Dashed black line indicates IMS LNM, solid black line indicates IMS HNM. 55

3.5 Direct arrivals from ray-tracing for events occurring in the spring (left) summer (middle) and fall (right). Each dot represents a predicted direct arrival from ray-tracing, while the dot color represents the celerity (i.e., predicted infrasonic phase) of the arrival. Array locations are denoted by black triangles and the GT location is denoted by the red star. ......................... 59

3.6 a-c: Summary of predicted ray arrivals across the network as a function of season, where grey indicates no predicted arrivals, red indicates tropospheric arrivals, green indicates stratospheric arrivals and blue indicates thermospheric arrivals. d-f: overall predicted ray arrivals across the network for each season. ............................................................ 59

3.7 Summary of additional event detections at each station through incremental p-value parameter increases, shown against analyst review. .................. 63

3.8 Summary of relationship between total number of GT event detections compared to all detections as a function of p-value. As p-value increases, both the number of GT detections and the overall number of detections increase. Detections are defined as all detections that occur on a day with a GT event. 64

3.9 Waveform and array processing example for an event on 2012-05-14T17:56:34 at BRP with clear coherent background noise that was missed during automatic detection but identified during analyst review. a-d: Array processing results in the form of F-value, estimated trace velocity, estimates backazimuth, calculated correlation between array elements. Red line in panel c denotes the true backazimuth (316°) from receiver to source. e: Beamed waveform for time period where blue window represents the analyst-defined detection. ............................................................ 65
3.10 Example of cumulative integration processing for HWU using waveforms from an event on 09-11-2012. (a) Beamed signal (black) and residual (blue) window spectra. (b) Example of integration limits; green line denotes KDE derived from September noise model of GT hours at the frequency of interest (2.3 Hz). Black line indicates the signal upper integration limit in dB and blue indicates the residual upper integration limit in dB. (c) Cumulative distribution values for each individual frequency integration across the band of interest (0.5-5 Hz). (d) Comparison of individual residual and signal cumulative distribution values. Horizontal clustering along the x-axis indicates high signal and low noise.

3.11 Example of cumulative integration processing at HWU using waveforms from an event on 06-26-2012. a. Beamed signal (black) and residual (blue) window spectra. b. Example of integration limits; green line denotes KDE derived from June noise model of GT hours only at the frequency of interest (1.4 Hz). Black line indicates the signal integration limit in dB and blue indicates the residual integration limit in dB. c. Cumulative distribution values for each individual frequency integration across the band of interest (0.5-5Hz). d. Comparison of individual residual and signal cumulative distribution values. Horizontal clustering along the x-axis indicates high signal and low noise.

3.12 Example of cumulative integration processing at HWU using waveforms from an event on 07-31-2012. a. Beamed signal (black) and residual (blue) window spectra. b. Example of integration limits; green line denotes KDE derived from June noise model of GT hours only at the frequency of interest (0.6 Hz). Black line indicates the signal integration limit in dB and blue indicates the residual integration limit in dB. c. Cumulative distribution values for each individual frequency integration across the band of interest (0.5-5Hz). d. Comparison of individual residual and signal cumulative distribution values. Horizontal clustering along the x-axis indicates high signal and low noise.

3.13 Overall cumulative signal and residual values at each array, determined by the average value across 0.5-5 Hz. Red dots are values calculated for missed detections.
3.14 (a) Comparison between station distance from source [km] and number of successful automatic event detections, as a function of event yield in kg1/3. Marker color and size represent the percentage of GT events within each yield bin, the total number of events within each yield bin are shown in the histogram along the y-axis. (b) Comparison between station distance from source [km] and number of successful total (automatic and analyst) event detections, as a function of event yield. Marker color and size represent the percentage of GT events within each yield bin. Missed detections at each station categorized by phenomena that contributes to missed detections.

3.15 Missed detections at each station categorized by phenomena that contributes to missed detections. Missed detections due to noise are in blue while missed detections due to propagation effects are in green. Detections are separated based on how they were identified: detections identified by parameter optimization are dashed, detections identified by the analyst are hashed and true missed detections are solid.

4.1 Infrasonic network used for producing event localization estimates, where red star indicates GT source (UTTR) and blue triangles indicate arrays used for localization. Black triangle indicates BGU, the closest seismic station to the source.

4.2 (a) Number of event detections at each station across the network. (b) Total number of detections/GT event across the network.

4.3 Number of detections, as a function of detection backazimuth measured from source to receiver. Radial bins mark detection number.

4.4 Monthly localization comparisons utilizing the two distinct celerity models: Blom et al. (2020), black and monthly PGMs, dark blue. Dashed lines indicate the change in mislocation for an individual event. Filled markers indicate events where the true GT location is contained within the 90% EE and empty markers indicate events where the 90% does not contain the true GT location. The red star marks the UTTR source location. Bottom right panel compares event-to-event changes in mislocation and in 90% EE areas.

4.5 Spatial localization results for two UTTR (red star) events. (a) marks the first event (date labeled at top), and (b) marks the second, later event (date labeled at top). Results utilizing the Blom et al. (2020) generalized celerity model are shown in black while results utilizing the monthly PGMs are shown in dark blue. Ellipses denote the 90% confidence bounds and the marker denotes the maximum a posteriori solution. Blue triangles represent the arrays with an observable detection for the event and grey lines indicate the measured backazimuth of the detection.
4.6 Monthly origin time comparisons utilizing the two distinct celerity models: Blom et al. (2020), black circles and monthly PGMs, dark blue squares. Filled markers indicate events where the true GT origin time is contained within the 90% temporal confidence bounds and empty markers indicate events where the 90% temporal confidence bounds do not contain the true GT origin time.

4.7 Comparisons between detection observations (grey histogram) and model predictions for a: signal celerities (in km/sec) and b: backazimuth deviations (in degrees) for the Blom (2020) predictions (dashed black line) and the empirical produced regional UTTR model (dotted green line).

4.8 Seasonal (rows) and station-level (columns) comparisons between detection observations (grey histogram) and model predictions for a: signal celerities (in km/sec) and b: backazimuth deviations (in deg) for the Blom et al., (2015) PGMs (dark blue line).

4.9 Seasonal (rows) and station-level (columns) comparisons between detection observations (grey histogram) and model predictions for signal celerities (in km/sec) for the Blom et al. (2015) PGMs (dark blue line).

4.10 Seasonal (rows) and station-level (columns) comparisons between detection observations (grey histogram) and model predictions for backazimuth deviations (in degrees) for the Blom (2015) PGMs (dark blue line).

4.11 Monthly localization comparisons utilizing the Blom et al. (2020) generalized celerity model (black) and the empirical UTTR celerity model (dark green). Dashed lines indicate the difference in mislocation for an individual event. Top right panel indicates the OT residual and length of the 90% error bounds for each event (in km), where filled markers indicate events where the true GT origin time is contained within the 90% EE and empty markers indicate events where the 90% does not contain the true GT origin time.

4.12 Summary statistics (a, d: mislocation in km; b, e: 90% EE area in km²; c, f: origin time residual in sec) for localizations produced utilizing the Blom et al. (2020) generalized celerity model (black circles), UTTR empirical model (green triangles) and monthly PGMs.
LIST OF TABLES

Table                                                                 Page

2.1 Parameters used for automatic processing. ............................................... 18
2.2 Number of automatic detections at each array for IP and BH. .................. 18
2.3 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).... 21
2.4 Recall and precision rates for each method. ............................................. 28
2.5 Updated recall and precision rates that include signals detected by both IP and
   BH as true positives. ................................................................. 28
2.6 Precision and recall for synthetic data processing. .............................. 35

3.1 Locations of infrasound arrays used in this study, with data recovery statistics.
   Data recovery is based on availability of array data within the predicted
   infrasonic arrival times for each GT event from UTTR during 2012. ........ 47

3.2 Automatic processing parameters for detection. .................................... 52

4.1 Mean source localization estimates for 45 UTTR events. Mislocations are cal-
   culated as the distance between the true location and the MAP location
   estimate (in km). Origin time residuals are calculated as the true origin
   time – MAP origin time estimate. The 90% confidence bounds are com-
   puted from the marginal spatial and temporal distributions, respectively. ... 103

4.2 Median source localization estimates for 45 UTTR events. Mislocations are
   calculated as the distance between the true location and the MAP loca-
   tion estimate (in km). Origin time residuals are calculated as the true
   origin time – MAP origin time estimate. The 90% confidence bounds are
   computed from the marginal spatial and temporal distributions, respectively.103

A1 Means, µj, variances, σj, and weights, wj, for the two Gaussian mixture models
(GMM) defining infrasound celerity. .................................................. 133

xx
To the original Dr. D.
Infrasound signals are low-frequency (<20 Hz) acoustic waves caused by small changes in atmospheric pressure. Both seismic and below-audible acoustic signals, commonly referred to as infrasound, are often generated by similar sources when they occur near the atmosphere – solid earth boundary. The signals differ due to their propagation medium; seismic waves propagate through the solid earth while infrasound signals propagate through the atmosphere. Predicting propagation path effects of infrasound waves can be more difficult as the signals travel through a dynamic medium. As infrasound is a relatively younger field than seismology, current research efforts focus on understanding basic phenomena in order to answer fundamental questions such as:

- What sources can be identified?
- How to optimally detect and locate sources of interest?
- How to process large volumes of data originating from large networks such as the International Monitoring System?
- How to characterize the time-varying nature of the atmosphere and its influence on signal propagation?
- How to separate signals of interest from sources of noise?

These questions motivate the work reported in this dissertation.
Infrasound is one of the four technologies utilized for verifying compliance with the Comprehensive Test Ban Treaty (CTBT) (Beall et al., 1999; Brown et al., 2002). The other technologies are seismic, hydroacoustic and radionuclide. The International Monitoring System (IMS) is comprised of all four technologies with 337 facilities distributed worldwide to support global monitoring for nuclear explosions. Over 85% of these facilities are installed and collecting data. The infrasound component consists of 60 surface stations, each of which is an array. These worldwide networks produce a continuous datastream that necessitates automated techniques and algorithms for accurate and efficient data processing and analysis.

Infrasound signals are generated by both natural (e.g., ocean interactions, earthquakes, volcanoes and meteors) and man-made (e.g., nuclear explosions, mining activity, urban processes, missiles, aircraft) sources (Arrowsmith et al., 2010). This thesis focuses on the identification and characterization of signals from above ground chemical explosions with accompanying ground truth information. Signals from explosive sources can under some circumstances propagate long distances (up to 1000s of km) due to the stratification of winds and atmospheric temperatures, which control static sound speed (343 m/sec at 20°C) creating narrow ducts through which signals travel and return to the ground (Evers & Haak, 2009).

This ducting of signals from ground sources refracts the waves back to the earth’s surface as a result of temperature and wind speed gradients within the four layers of the atmosphere: the troposphere (0-12 km), stratosphere (12-50 km), mesosphere (50-80 km) and thermosphere (80-320 km). The atmospheric temperature profile, wind speed and direction control the signal propagation path with a signal propagating upwards until it reaches a height with higher effective sound speed than the velocity at the source location where it then refracts back to the ground surface where sensors are deployed. The refracting layer controls the celerity or total distance divided by total propagation time of the return. Boundary layer arrivals tend to have group velocities or celerities higher than 330 m/sec, while tropospheric
arrivals range from 310-330 m/sec, stratospheric arrivals range from 280-330 m/sec, and thermospheric arrivals range from 180-300 m/sec (Negraru et al., 2010).

At high and middle latitudes, up to 20% of infrasonic energy can be ducted in the troposphere. These tropospheric refractions are a result of tropospheric winds, thus the appearance of these arrivals is largely dependent on ambient weather conditions below the tropopause (Brown et al., 2002). These ducted signals at closer distances have larger amplitudes as the wave fronts are subjected to a reduced amount of geometric spreading and molecular attenuation, thus producing observed arrivals important for infrasonic event location.

Stratospheric and thermospheric returns are driven by atmospheric temperature gradients as well as winds with stratospheric ducts containing between 0-40% of infrasonic energy and thermospheric ducting containing 40-85% of energy (Brown et al., 2002; Drob et al., 2003). The greatest fraction of stratospheric arrivals occur at middle and high latitudes. Stratospheric wind jets are predominantly zonal (east-west) and reverse directions between the winter and summer hemispheres, moving eastward in the winter hemisphere and westward in the summer hemisphere. The least stratospheric ducting occurs at the equatorial latitudes since winds in this region are typically transient and light. Additionally, the equatorial tropopause is higher and colder relative to the ground, resulting in upward refraction of infrasonic signals. The largest fractions of thermospheric arrivals occur in the equatorial regions. Thermospherically ducted signals are often smaller in amplitude than tropospheric and stratospheric signals due to increased geometric spreading and molecular attenuation (Drob et al., 2003). Energy not ducted in the thermosphere will propagate into the upper thermosphere where it dissipates. This represents between 12-17% of global infrasonic energy.

Ducting regions vary diurnally and seasonally as a result of changes in temperatures and winds (de Groot-Hedlin et al., 2010; Drob et al., 2010). Direct measurements focused on weather conditions provide insight into the properties in the lower atmosphere (0-30 km) while atmospheric models provide insight into the upper atmosphere conditions (30-
320 km). In the lower atmospheric layers, conditions are strongly affected by dynamic weather. Weather conditions are quantified by radiosonde measurements typically made at the 06:00 and 18:00 hour intervals or through observations at ground weather stations. Because the upper atmosphere is less time variant, statistical climatology models, such as the Mass Spectrometer Incoherent Scatter/Horizontal Wind Model (MSIS/HWM) apply empirical specifications derived from historical archives (Drob et al., 2008). These models provide estimates of global winds, temperatures and pressures for the middle and upper atmosphere, but do not include modeling for the lower atmosphere. The Naval Research Laboratory Ground to Space (G2S) semi-empirical model developed by (Drob et al., 2003), produces seamless hourly global atmospheric predictions from the ground to 750 km by extrapolating NWP data to higher altitudes using specifications from the NWM-93 model. The G2S database archives atmospheric specifications on a 1°x1° spatial grid with a 1-hour temporal resolution. Specifications are defined by a geographic location, month and time of day.

Application of the G2S model provides the opportunity to model and predict infrasound propagation based on near real-time atmospheric specifications. Specifications estimate the speeds of the zonal (east to west) and meridional (north to south) wind components along with the effective sound speed at a specific location. Estimating these components allows us to model possible atmospheric ducts within the layers of the atmosphere.

Successful detection and subsequent location of events of interest depends on the ability to discriminate between signals of interest and background noise (Blandford, 1982). One of the largest challenges in analyzing infrasound data is reducing the number of signal detections stemming from correlated environmental (wind guests, microbaroms) or human-related noise. Ambient infrasonic noise is characterized over the 0.1-7+ Hz frequency band, but is highly variable with respect to season, time of day and receiver location (Arrowsmith & Hedlin, 2005; Bowman, 2005). The infrasonic noise field is affected at long periods by long-range
pressure fluctuations generated by oceans or eddies, and at higher frequencies by winds along the near surface boundary layer. At higher frequencies, noise is affected by short-range pressure fluctuations generated by local weather, topography, and nearby noise sources.

It is often difficult to identify signals of interest arriving at infrasound sensors due to these noise sources. In order to observe signals of interest, infrasound data acquisition and analysis rely on three approaches to reduce noise. First, sensors are installed at naturally quiet locations such as isolated forests (when possible) which can break up wind eddies at the turbulent boundary in order to reduce some types of background wind noise. Second, individual sensors are fitted with noise reduction technologies such as long hoses or pipes designed to spatially filter out short wavelength wind noise (Stump et al., 2004; Walker & Hedlin, 2010). Third, a series of four or more sensors are deployed at spatial separations of 100-1000m across an individual site in order to provide spatial data needed to apply array processing to extract coherent signals (Krim & Viberg, 1996; Rost & Thomas, 2002).

Array processing increases signal-to-noise ratios (SNR) through the summation of the four or more spatially separated recordings at each array (Rost & Thomas, 2002). Conventional array processing, otherwise known as beamforming (Bartlett, Capon) separates the coherent and incoherent parts of a signal through the assumption of planar waves moving across the array. The processing identifies the signal backazimuth and slowness that maximizes the coherent signal by shifting the individual time series relative to one another to account for the travel time differentials across array elements, bringing the signal into phase while deconstructively canceling the noise. In the classical, or Bartlett methodology, data records for each array element are time-shifted versions of the other with local noise. Records are shifted into the frequency domain while also including the dependence of the delays on the slowness. An estimate of $F(f)$, or the beampower, can be defined. Beampower estimates are used to estimate a Fisher value, or F-statistic, which can be used in signal detectors that depend on a specified level of coherence to statistically declare an arrival.
In addition to the global IMS network, a number of regional infrasound networks have been installed worldwide. These networks are a recent addition typically focusing on shorter distances and higher frequencies. As a result, there is need to quantify their performance under these unique circumstances building on what has been done for longer ranges signals. The three chapters in this dissertation focus on the evaluation of signals from a regional infrasound network of 7 stations deployed in the Western United States. These stations were located across the state of Utah and operated from 2010-2013 offering the opportunity to evaluate infrasound signals recorded on arrays at distances between 50-500 km from a known source. The large number of signals recorded across an infrasound network motivates the assessment of automated detection and location procedures as well as the role human analysts play within a regional setting separating noise from signals.

Chapter two outlines a methodology for comparing automated infrasound detections using different methodologies with the goal of identifying an optimal automated detector. This work follows comparisons established by Park et al. (2018b) and utilizes both real and synthetic signals in order to evaluate the performance of two distinct infrasound signal detection algorithms using measures of precision and recall in the comparison. Real signals are derived from detection bulletins produced as part of the Dynamic Networks Experiment 2018 (DNE18) (Young et al., 2018) and synthetic signals are produced following methodologies outlined in Averbuch et al. (2018) and den Ouden et al. (2020). Results provide a basis for future detector development while highlighting shortcomings in the assessment of detection algorithms for low signal to noise ratio signals.

Chapter three builds upon chapter two by utilizing the output from a single automated detector in combination with a quantification of noise and propagation path effects to provide a physical understanding the conditions under which a detector may succeed or fail. This study makes use of Ground Truth explosions to assess the complete performance of the automatic signal detector. Combining background noise quantification and atmospheric
propagation analyses with comparison of spectral amplitudes provides an approach to identify the physical mechanisms that contribute to missed detections across the network. This analysis extends to an estimate of detector performance across the entire network. The mechanisms that lead to missed detections at individual arrays contribute to network-level estimates of detection capabilities and provide a basis for deployment decisions for regional infrasound arrays in areas of interest.

Chapter four focuses on assessing improvements to recent infrasonic localization methodologies which apply modeling-based predictions for atmospheric propagation to better constrain arrival celerities and enhance temporal and spatial localization estimates. This chapter relies on the results of chapter three by utilizing the robust dataset of detections from the set of ground truth events to estimate their locations and assess their statistical characteristics. The goals of this study are two-fold: first, it aims to perform a statistical assessment of improvements to infrasonic localization capabilities following examples in seismic literature in order to produce a formal assessment of infrasonic localization capabilities and limitations. Second, it aims to identify improvements to localization including updated methods to better take into account the propagation medium. The work addresses a need within the infrasound community to refine location procedures as well as to better understand the role of changing atmospheric conditions on infrasound signal propagation and automated event location. Results indicate that atmospheric specifications need refinement in order to accurately predict infrasonic signal propagation; unrefined propagation results lead to errors within both spatial and temporal location estimates. Bias and accuracy in localization results are driven by both network detection capability as well as network design with large location errors related to azimuthal gaps between detecting arrays. Finally, the use of newly introduced empirical models significantly improves temporal localization results but has little effect on spatial locations.
These three chapters guide new analysis of infrasound signals that improve both our understanding of atmospheric propagation dynamics and the utility of automated processing pipelines for monitoring purposes. They also provide a framework for making decisions necessary for the deployment of regional infrasound networks for monitoring purposes.
CHAPTER 2

Methodologies for the Comparison of Automatic Infrasonic Signal Detectors

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

\[ \text{Dannemann Dugick, F., Albert, S., Averbuch, G., and Arrowsmith, S., "Utilizing the Dynamic Networks Data Processing and Analysis Experiment (DNE18) to Establish Methodologies for the Comparison of Automatic Infrasonic Signal Detectors\} submitted to the Bulletin of the Seismological Society of America} \]
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.

2.1. 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 (Begnaud et al., 2019). 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 Young et al. (2018) 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 (Carmichael et al., 2018). 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 (Park et al., 2017). A variety of automated infrasound detectors have been presented in the peer-reviewed literature including the progressive multi-channel correlation (PMCC) algorithm (Cansi, 1995), the F-detector implemented in InfraTool (Hart, 2004), the Adaptive F-Detector (AFD) (Arrowsmith et al., 2009) and the Multivariate Adaptive Learning Detector (MALD) (Arrowsmith, 2018). The AFD is implemented within Infrapy (IP), a Python-based array processing toolkit developed by LANL that was recently open-sourced in 2020 (Blom et al., 2016; Dannemann & Marcillo, 2017; Marcillo et al., 2015; Webster et al., 2021). The MALD is implemented in Bloodhound
BH), a Python-based pipeline data processing software developed by SNL (Arrowsmith et al., 2018).

Automated processing techniques have been used to produce bulletins of infrasound events in multiple regions such as the western United States (Park et al., 2014; Walker et al., 2011), the state of Alaska (Sanderson et al., 2020), the Korean peninsula (Che et al., 2019; McKenna et al., 2008; Park et al., 2016) and across the European continent (Le Pichon et al., 2008; Pilger et al., 2018). Automatic detections across the 48 operational stations of the IMS infrasound network are documented by the International Data Center (IDC) and more recently by Arrowsmith (2018). In addition, a number of bulletins have been produced specifically documenting infrasonically-observed volcanic activity (e.g., (Matoza et al., 2013; Sanderson et al., 2020)).

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 (Park et al., 2017) 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 (Kay, 1998). 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 (Park et al., 2017). 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 (Park et al., 2017). 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 ($P_{FA}$) (Marcillo et al., 2018). 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 (Evers & Haak, 2001; Hetzer et al., 2009) including surf (Le Pichon, 2004), microbaroms (Landès et al., 2012), thunder (Farges & Blanc, 2010), volcanoes (Le Pichon, 2004) and anthropogenic activities such as mining, industrial activity, aircraft or urban noise (Le Pichon et al., 2008; McComas et al., 2018; Park & Stump, 2015).

Park et al. (2017) 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 Arrowsmith (2018), 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.
2.2. Detector Overview

The AFD as implemented within Infrapy (IP) (Webster et al., 2021) was developed to reduce false alarms from correlated or persistent noise across an array of interest (Arrowsmith et al., 2009). 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 (Shumway, 1971). 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. Arrowsmith et al. (2008) and Arrowsmith et al. (2009) 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 Park et al. (2016) and Park et al. (2017). Park et al. (2015) and Dannemann Dugick et al. (2020) 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 Arrowsmith (2018) and Arrowsmith et al. (2018). 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 ($\phi$), 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 (Green & Nippress, 2019) but clearly exhibit a stable estimate of backazimuth over time (refer to Figure 10 in (ArrowSmith et al., 2015)). This detector is similar in principle to the basis of the Hough Transform detectors of Brown et al. (2008) and Averbuch et al. (2018) 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:

$$\chi^2 = -2 \sum_{k} \ln p_i$$

(2.1)

where $\chi^2$ 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.

2.3. Instrument Deployment and Details

The infrasonic dataset used in this study consists of array data from six co-located seismo-acoustic stations (??) installed in the state of Utah by the University of Utah (UU) and Southern Methodist University (SMU) between 2006-2010 (Zhou et al., 2010). 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 (Stump et al., 2004). Data were sampled at 100 samples per second. Acoustic sensors at NOQ are Chaparral Physics Model 2.0 microphones (Park et al., 2014) with a flat frequency range of 0.1 to >100 Hz (Arrowsmith et al., 2008a). 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 (Fisher, 2013; Hart, 2007). Data is available from IRIS using the network code ‘YJ’ (Hayward, 2010).

2.4. Direct Detector Comparisons

2.4.1. 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
Figure 2.1: UU/SMU seismo-acoustic network used for DNE18 processing. Blue triangles indicate infrasound arrays utilized within this study.
between the implementation of the detectors make a direct parameter comparison difficult. The parameters used for detection processing are included in Table 2.1. Parameters were set based on experience developed from prior infrasonic analyses in the Utah region (Arrowsmith et al., 2014; Dannemann Dugick et al., 2020; Park & Stump, 2015; Park et al., 2017). However, we note that all these publications focus on parameter optimization for AFD. Arrowsmith (2018) 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 2.2 summarizes the total number of detections at each array from the three months based on each detector, while Figure 2.2 visually present detection results. As seen in Figure 2.2a, 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 2.2b 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 2.3a, 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
Table 2.1: Parameters used for automatic processing.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter band (Hz)</td>
<td>1-5</td>
</tr>
<tr>
<td>Time window (s)</td>
<td>30</td>
</tr>
<tr>
<td>Overlap (%)</td>
<td>50</td>
</tr>
<tr>
<td>p-value</td>
<td>0.001, 0.01</td>
</tr>
<tr>
<td>Semblance Threshold (BH)</td>
<td>0.5</td>
</tr>
<tr>
<td>Adaptive window for AFD (s)</td>
<td>1800</td>
</tr>
</tbody>
</table>

Table 2.2: Number of automatic detections at each array for IP and BH.

<table>
<thead>
<tr>
<th>Array</th>
<th>InfraPy</th>
<th>Bloodhound</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRP</td>
<td>11541</td>
<td>25143</td>
</tr>
<tr>
<td>FSU</td>
<td>3121</td>
<td>3276</td>
</tr>
<tr>
<td>HWU</td>
<td>6747</td>
<td>22589</td>
</tr>
<tr>
<td>LCM</td>
<td>9500</td>
<td>16458</td>
</tr>
<tr>
<td>NOQ</td>
<td>3584</td>
<td>1106</td>
</tr>
<tr>
<td>WMU</td>
<td>6913</td>
<td>13840</td>
</tr>
</tbody>
</table>

Figure 2.2: (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.
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 2.3b 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 recurring sources, or clutter, are of unknown origin. Further investigation and identification may assist with continued detector improvements.

2.4.2. 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 (Young et al., 2018), 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 2.3 includes the total number of detections at each array for the day and night hours, while Figure 2.4 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 Carmichael et al. (2016), 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 (± 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 2.5 shows a schematic of this process. An example of these comparisons is presented in Figure 2.6, 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 2.6b 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 2.7. 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. For the second hour, general trends can be identified (see Figure 2.7b). While a number of events are detected by all three methods at each
Figure 2.3: 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.

Table 2.3: 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).

<table>
<thead>
<tr>
<th></th>
<th>IP HR1</th>
<th>IP HR2</th>
<th>BH HR1</th>
<th>BH HR2</th>
<th>A HR1</th>
<th>A HR2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRP</td>
<td>1</td>
<td>19</td>
<td>30</td>
<td>24</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>FSU</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>HWU</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>20</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>LCM</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>13</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>NOQ</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>WMU</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>11</td>
</tr>
</tbody>
</table>
Figure 2.4: 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 2.5: 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 2.6: 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.
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 2.6. 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 (Young et al., 2018), where detector recall is defined as the proportion of true objects found from all true samples,

\[ \text{recall} = \frac{\# \text{ of true positives}}{\text{all true samples}} \]  \hspace{1cm} (2.2)

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

\[ \text{precision} = \frac{\# \text{ of true positives}}{\# \text{ of true positives} + \# \text{ of false positives}} \]  \hspace{1cm} (2.3)

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.
Figure 2.7: 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.
Detector recall and precision values for each hour are presented in Table 2.4. 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 2.7, 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 2.5. While overall values, expressed as percentages, increase, relationships between the two detectors remain the same. Recall is higher 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 2.8 as averaged Power Spectral Density (PSD) curves (Welch, 1967) 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 Park et al.
(2014) and Park & Stump (2015). Additional comparisons of hourly detections across the full three month dataset would further explore these conclusions.

2.4.3. 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 (Averbuch et al., 2018) and CLEAN beamforming (den Ouden et al., 2020).

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 2.9a).
Table 2.4: Recall and precision rates for each method.

<table>
<thead>
<tr>
<th></th>
<th>IP Recall</th>
<th>IP Precision</th>
<th>BH Recall</th>
<th>BH Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR 1</td>
<td>BRP</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>FSU</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HWU</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LCM</td>
<td>67%</td>
<td>33%</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>NOQ</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>WMU</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HR2</td>
<td>BRP</td>
<td>67%</td>
<td>63%</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>FSU</td>
<td>43%</td>
<td>30%</td>
<td>43%</td>
</tr>
<tr>
<td></td>
<td>HWU</td>
<td>36%</td>
<td>50%</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>LCM</td>
<td>36%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>NOQ</td>
<td>50%</td>
<td>33%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>WMU</td>
<td>45%</td>
<td>100%</td>
<td>64%</td>
</tr>
</tbody>
</table>

Table 2.5: Updated recall and precision rates that include signals detected by both IP and BH as true positives.

<table>
<thead>
<tr>
<th></th>
<th>IP Recall</th>
<th>IP Precision</th>
<th>BH Recall</th>
<th>BH Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR 1</td>
<td>BRP</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>FSU</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HWU</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LCM</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td>NOQ</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>WMU</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HR2</td>
<td>BRP</td>
<td>68%</td>
<td>68%</td>
<td>79%</td>
</tr>
<tr>
<td></td>
<td>FSU</td>
<td>64%</td>
<td>70%</td>
<td>64%</td>
</tr>
<tr>
<td></td>
<td>HWU</td>
<td>46%</td>
<td>75%</td>
<td>54%</td>
</tr>
<tr>
<td></td>
<td>LCM</td>
<td>55%</td>
<td>100%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>NOQ</td>
<td>100%</td>
<td>75%</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>WMU</td>
<td>45%</td>
<td>100%</td>
<td>64%</td>
</tr>
</tbody>
</table>
Figure 2.8: 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 (Bowman et al., 2009) while dashed black line indicates the IMS Low Noise Model.
Figure 2.9b 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 2.10. 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 2.10a, 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 2.10) 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 2.10b we compare detector success rates for short duration (31 total) synthetic signals, and in Figure 2.10c 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.
Figure 2.9: (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 2.10: (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).
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 2.11. 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 2.11 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 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.

2.4.3.1. 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 2.6 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 ??, 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.

2.5. 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 2.1), 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 (Arrowsmith, 2018; Park et al., 2017) 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 (Brown et al., 2002; Marcillo et al., 2015; Park et al., 2014).

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 net-
Figure 2.11: 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.

Table 2.6: Precision and recall for synthetic data processing.

<table>
<thead>
<tr>
<th></th>
<th>IP Recall</th>
<th>IP Precision</th>
<th>BH Recall</th>
<th>BH Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>NS LVL 1</td>
<td>90%</td>
<td>100%</td>
<td>57.5%</td>
</tr>
<tr>
<td></td>
<td>NS LVL 2</td>
<td>90%</td>
<td>100%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>NS LVL 5</td>
<td>90%</td>
<td>100%</td>
<td>62.5%</td>
</tr>
<tr>
<td></td>
<td>NS LVL 10</td>
<td>90%</td>
<td>97%</td>
<td>55%</td>
</tr>
<tr>
<td>0.03</td>
<td>NS LVL 1</td>
<td>-</td>
<td>-</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>NS LVL 2</td>
<td>-</td>
<td>-</td>
<td>85%</td>
</tr>
<tr>
<td></td>
<td>NS LVL 5</td>
<td>-</td>
<td>-</td>
<td>85%</td>
</tr>
<tr>
<td></td>
<td>NS LVL 10</td>
<td>-</td>
<td>-</td>
<td>62.5%</td>
</tr>
</tbody>
</table>
work. 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 (Park et al., 2016).

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.

2.6. Data and Resources

Infrasonic waveform data utilized during DNE18 is available from IRIS using the network code ‘YJ’. The DOI for this dataset is https://doi.org/10.7914/SN/YJ’2010 (Hayward, 2010). Data from NOQ is part of the University of Utah seismic network and is available using the station code ‘UU’.

2.7. 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.

2.8. Acknowledgments

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 Engineer-
ing 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.
CHAPTER 3
Factors Influencing Automatic Infrasonic Processing Performance

Abstract

We explore physical and deployment factors that influence infrasound signal detection and assess automatic detection performance for a regional infrasound network of arrays in the western US, using signatures of Ground Truth explosions (yields). Despite these repeated known sources, published infrasound event bulletins contain few Ground Truth events. Arrays are primarily distributed towards the SSE and SSW, at distances between 84-458 km of the source, with one array offering azimuthal resolution towards the NE. Events occurred throughout the spring, summer and fall of 2012, with the majority occurring during the summer months. Depending upon array, automatic detection that utilizes the Adaptive F-Detector successfully identifies between 14-80% of the ground-truth events, while a subsequent analyst review increases successful detection to 24 – 90%. Combined background noise quantification, atmospheric propagation analyses and comparison of spectral amplitudes determine the mechanisms that contribute to missed detections across the network. This analysis provides an estimate of detector performance across the network as well as a qualitative assessment of conditions that impact infrasound monitoring capabilities. The mechanisms that lead to missed detections at individual arrays contribute to network-level estimates of detection capabilities and provide a basis for deployment decisions for regional infrasound arrays in regions of interest.

3.1. Introduction

Infrasound signals from large sources such as volcanic explosions (Matoza et al., 2007), earthquakes (Le Pichon et al., 2005b; Mutschlecner & Whitaker, 2005), tsunamis (Le Pichon et al., 2005b), mining explosions (Che et al., 2019; Hagerty et al., 2002) as well as nuclear and chemical explosions (Che et al., 2009, 2014; Evers & Haak, 2007; Park et al., 2014; Pasyanos & Kim, 2019; Walker et al., 2011) can be observed at distances of hundreds to thousands of kilometers from the source. The larger propagation distances reflect the stratification of wind speeds and atmospheric temperatures, which control static sound speed and create narrow ducts through which signals travel (Beasley & Georges, 1977; Evers & Haak, 2009). Waves propagating these distances refract back to the earth’s surface as a result of atmospheric temperature and velocity gradients within the four layers of the atmosphere that include: the troposphere (0-12 km), stratosphere (12-50 km), mesosphere (50-80 km) and thermosphere (80-320 km). From a signal detection perspective, boundary layer arrivals tend to have group velocities or celerities higher than .330 km/sec, tropospheric arrivals range from .310-.330 km/sec, stratospheric arrivals range from .280-.330 km/sec and thermospheric arrivals range from .180-.300 km/sec.

A global network of 60 planned infrasound monitoring arrays, part of the International Monitoring System (IMS) operated by the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO PrepCom) to detect atmospheric explosions is nearing completion, with 49 certified arrays as of June 2017 (Christie & Campus, 2010; Marty, 2019). The average spacing between IMS infrasound arrays is 1920 km (Christie & Campus, 2010), which is sufficient for the detection of 1 kT atmospheric explosions within 500 – 4500 km of the station (Green & Bowers, 2010); however, this spacing is insufficient for the detection of smaller, local infrasound sources that may only be detectable within a few
hundred kilometers. Regional infrasound arrays operated for scientific experiments (Assink et al., 2016; Ceranna et al., 2009) or as augmentation to existing seismic stations ((Che et al., 2019; Stump et al., 2007) supplement IMS coverage in regions of interest. Regional infrasound networks, which provide closer and more densely-spaced observations, render an opportunity to improve event detection, localization and identification of smaller sources. For example, regional observations of signals from the underground nuclear explosions in North Korea in 2009, 2013 and 2016 improved infrasonic localization estimates when compared to estimates solely from IMS stations (Che et al., 2014; Park et al., 2018a). Similarly, automated detections from a regional infrasound network supplemented sparse recordings from the 2005 Buncefield Oil Storage Facility explosive event that were observed on IMS arrays; ultimately producing an improved event location estimate within 35 km of the true source (Ceranna et al., 2009). Finally, regional infrasonic networks offer the opportunity for long-term monitoring of natural hazards such as volcanos (De Angelis et al., 2012) and can provide data for constraining energetic natural events such as tornados (Frazier et al., 2014).

The growing number of regional infrasound networks motivates the need to quantify the best data processing practices needed to produce optimum event catalogs. This study focuses on evaluating the performance of automated detection algorithms using real ground truth events observed across a regional infrasound network occurring over a range of atmospheric and noise conditions. Both the IMS global network and smaller regional networks utilize automatic processes to identify signals of interest using a consistent objective detection metric in order to reduce analyst workload considering the large number of infrasonic signals (Marty, 2019). A critical component of automatic processing is detector optimization, which depends on detection performance at individual arrays which depends on signal coherence across the array elements, signal frequency content and array-specific noise characteristics.

Within the context of this study, we define signals of interest to be explosion generated signals with ground truth while noise is results from any source of background emissions.
which produce infrasound energy that obscures the signals of interest. Infrasound background noise levels are known to be highly frequency and wind dependent, varying by as much as 60 dB [relative to 1Pa] at a single frequency, primarily dependent on wind conditions (Bowman, 2005). Typically, recurrent noise sources, both coherent and incoherent, can increase the background noise level and reduce array detection capability by decreasing the coherence between individual array elements (Bowman et al., 2009). Coherent noise sources which produce signals that are spatially correlated between array elements (Brown et al., 2014) while incoherent noise sources such as wind are not spatially correlated between array elements. Multiple studies have documented that repeating or continuous coherent noise sources can degrade the identification of signals of interest using arrays (Woodward et al., 2005). The most well-known example of such sources of noise are the microbaroms which commonly produce coherent signals in the 0.1 - 0.5 Hz frequency band. Additional sources of repetitive signals within the 0.5 - 5 Hz band include noise related to surf (Arrowsmith & Hedlin, 2005; Le Pichon, 2004), thunder (Farges & Blanc, 2010), volcanoes (Matoza et al., 2007) and anthropogenic activities such as mining, industrial activity, aircraft or urban noise (Bowman et al., 2009).

The Adaptive F-detector (AFD) was developed (Arrowsmith et al., 2009) to account for both correlated and uncorrelated noise through modification of the conventional F-statistic, which is based on a F-distribution that assumes under the null hypothesis there is a ratio of two random variables with chi-square distributions. The revised detector reduces false alarms from coherent noise by applying an adaptive analysis window that updates the detection threshold, which may be elevated due to coherent background noise. The AFD reduced false alarm detections attributable to correlated noise across arrays in the production of a two-year infrasound event bulletin, relative to the non-adaptive detector (Park & Stump, 2015; Park et al., 2014, 2016), using data from a regional infrasound network in the Western US. AFD shows a fair predictive capability to detect small explosions (Carmichael et al., 2020).
Seismoacoustic event bulletins (Park et al., 2014; Walker et al., 2011) illustrate that the Utah Test and Training Range (UTTR) is a major source of infrasound signals in this region. Despite these repeated known sources, the Park et al. (2014) event bulletin, produced using the AFD and including automatic association and location procedures, only includes 5 of the 47 known blasts from UTTR in 2012 which motivates this more in-depth investigation.

Array-centric detection performance (Green & Bowers, 2010; Le Pichon et al., 2008) is controlled by the efficiency of signal propagation from source to receiver, signal arrival characteristics relative to background noise at the receiving station and the ability of the detector and/or array elements and the array detector ability to identify a unique coherent signal of interest within a time series. The small number of identified events in the bulletin of Park et al. (2014) relative to the total number of known UTTR events reflects several issues: (1) A lack of an atmospheric waveguide so that energy does not efficiently propagate from the source (UTTR) to the receivers; (2) A high noise level at the receiver obscuring the signal; (3) Problems with methodologies for detecting and associating signals of interest across the network; or (4) A combination of factors.

The UTTR events are infrasound sources where knowledge of a facility location, combined with seismically determined event times, provides ground truth (GT) for infrasound signals to address these issues. The primary goal of this study is to utilize this GT data set to quantitatively assess and interpret the detection performance of a regional infrasonic network. This study will improve our insight in the causes of missed arrivals. Arrays within the network are located from 84 to 458 km from the source where arrays located near the source (84 km) to 200-250 km of the source provide the opportunity to assess automatic signal detector performance at distances within the acoustic ‘stratospheric shadow zone’ (Herrin et al., 2007; Negraru et al., 2010). A subsequent analysis will address association and location.

This manuscript is organized as follows: Section 3.2 describes the regional infrasonic network utilized in this study. Section 3.3 describes five components used to assess auto-
mated detector performance as well as evaluating mechanisms for missed detections across the network. Section 3.4 discusses these results in the context of an estimate of detector performance at each array or station, as well as an assessment of conditions that impact infrasound monitoring capabilities.

3.2. Dataset

The Utah infrasound network (Arrowsmith et al., 2008a; Stump et al., 2004) consisted of a total of 12 infrasound arrays, 9 integrated into the University of Utah seismic network and 3 operated separately. Due to limited data availability as a result of telemetry failures and individual array deployment times, only 7 of these arrays were used in this study. Their locations are shown in Figure 3.1. Table 3.1 summarizes array locations and data recovery statistics for 2012. Each array consists of 4 acoustic sensor elements; one center element with three additional elements on 100 m legs, equally spaced around the center. Infrasound sensors are each fit with 8 porous hoses to reduce wind noise (Stump et al., 2004).

Data were sampled at 100 samples/sec. Acoustic sensors at NOQ are Chaparral Physics Model 2.0 microphones with a flat frequency range of 0.1 to 100 Hz (Arrowsmith et al., 2008a; Park et al., 2014) and recorded with a RefTek digitizer. BRP, FSU, HWU, LCM, PSU and WMU are equipped with Inter-Mountain Lab sensors and Q330 digitizers. The frequency response for the IML sensors is flat from 2 to 30 Hz (Fisher, 2013; Hart, 2007). Data from the six stations with IML sensors were corrected for instrument response to simulate a widened flat response down to 0.1Hz using response corrections from IRIS in Obspy (Beyreuther et al., 2010).

Ground Truth data includes 47 missile motor or propellant explosions conducted at the Utah Test and Training Range (UTTR) during the spring, summer and fall of 2012 with yields from 1665-17651 kg. Origin times were verified with seismic arrivals recorded at the closest seismometer, BGU, approximately 26 km from UTTR. Events occurred irregularly
Figure 3.1: Locations of the seven, four-element acoustic arrays (solid blue triangles) used in this study within the SMU-UU seismo-acoustic network. The red star denotes UTTR, the location of the ground truth events.

Table 3.1: Locations of infrasound arrays used in this study, with data recovery statistics. Data recovery is based on availability of array data within the predicted infrasonic arrival times for each GT event from UTTR during 2012.

<table>
<thead>
<tr>
<th>Array</th>
<th>Data Recovery Rate</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Number of Elements</th>
<th>BH Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRP</td>
<td>98%</td>
<td>39.47</td>
<td>-110.74</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>FSU</td>
<td>100%</td>
<td>39.72</td>
<td>-113.39</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>HWU</td>
<td>91%</td>
<td>41.61</td>
<td>-111.56</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>LCM</td>
<td>98%</td>
<td>37.01</td>
<td>-113.24</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>NOQ</td>
<td>98%</td>
<td>40.65</td>
<td>-112.12</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>PSU</td>
<td>81%</td>
<td>38.53</td>
<td>-113.85</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>WMU</td>
<td>100%</td>
<td>40.08</td>
<td>-111.83</td>
<td>4</td>
<td>100%</td>
</tr>
</tbody>
</table>
over the time period, with 1-2 events per week on weekdays between the hours of 16:00-22:00 UTC, which corresponds to 10:00-16:00 local time.

3.3. Methodology

Within this section we evaluate automated detector performance and identify mechanisms for missed detections across the network. In Section 3.3.1, the automatic detector is applied to the array data to establish a baseline of automatic detections and subsequent performance measure across the network. Section 3.3.2 and Section 3.3.3 utilize network noise estimates and atmospheric modeling predictions to assess these detection results. In Section 3.3.4, the impact of assumed detection parameters is quantified and contrasted against analyst review of the data that identifies signals missed by the automated process. Finally, in Section 3.3.5, we assess events the automatic detector did not identify in order to understand the physical phenomena that contributed to the missed detection including: (1) High noise immediately prior to and during the expected arrival time which masks the signal of interest; (2) Atmospheric conditions unfavorable to propagation from the source to the receiver; and (3) Signal amplitudes below the noise floor at the sensor.

3.3.1. Automatic Detection Utilizing the Adaptive F-Detector

Infrasound detection separates signals of interest from noise. Multiple detectors have been developed for infrasound signals, each with its own characteristics and include the progressive multi-channel correlation algorithm (Cansi, 1995) (PMCC), the gravity wave detection method (de Groot-Hedlin et al., 2014), the standard F-detector (Blandford, 1974) and the adaptive F-detector (AF) (Arrowsmith et al., 2009). PMCC utilizes the consistency of time delays between subnetworks of array elements to estimate coherent plane wave characteristics for signals including signal backazimuth and apparent velocity. The gravity wave detector uses sub-arrays with long baselines such as those (70 km) from the USArray
Transportable Array (TA) elements to identify coherent long-period gravity wave signals. This methodology has been shown to be sensitive to impulsive events and can distinguish distant events due to the large number of arrays in the network despite the large spatial separation between individual sensors. AFD with its scaling of the F-statistic was developed to account for both correlated and uncorrelated noise through modification of the conventional F-statistic and detection threshold. Testing with the AFD demonstrates reduced false alarms due to correlated noise across array elements, as compared to the conventional F-detector (Park et al., 2014).

Prior infrasound studies (Park et al., 2014; Walker et al., 2011) in the Western US used a frequency band of 1-5 Hz for automatic processing based on observations of regional infrasound signals within this band. An expanded frequency band of 0.5-5 Hz is used in this study as larger chemical explosions, such as those between 1665-17651 kg from UTTR produce infrasonic signals with energy below 1 Hz (Arrowsmith et al., 2008b; Carmichael et al., 2016; McKenna et al., 2007). This slightly broader band is used to expand the range of possible signals of interest.

The AFD utilizes the standard F-statistic calculation (Blandford, 1974):

\[
F = \frac{J - 1}{J} \frac{\sum_{n=n_0}^{n_0+(N-1)} \left[ \sum_{j=1}^{J} x_j(n + l_j) \right]^2}{\sum_{n=n_0}^{n_0+(N-1)} \left( \sum_{j=1}^{J} x_j(n + l_j) - \frac{1}{J} \sum_{m=1}^{J} x_m(n + l_m) \right)^2}
\]

(3.1)

In this equation, \(J\) is the number of sensors, \(x_{jn}\) is the waveform amplitude of the \(n\)th sample of the mean-free time series from sensor \(j\), \(l_j\) is the time-alignment lag from beamforming, \(n_0\) is the starting sample index for the processing interval and \(N\) is the number of samples in the processing window. The detector is implemented using the maximum average cross-correlation for beam formation and the associated p-value, which is the probability of obtaining a F-statistic as extreme as the calculated values under the F-distribution: \(pF(t)\), from all triplets of elements in an array for each time window.
AFD accounts for temporal changes in noise by applying an adaptive window to update the detection distribution, which makes a distinction between the signal and correlated noise (Arrowsmith et al., 2009). In the presence of correlated noise, the F-statistic is distributed as:

\[ CF(2BT, 2BT(N - 1)) , \tag{3.2} \]

where \( B \) is the bandwidth of the filtered data, \( T \) is the length of the detection window over which the power is averaged, \( N \) is the number of array elements and \( C \) is given by:

\[ C = (1 + N \frac{P_S}{P_n}) \tag{3.3} \]

with \( PS/PN \) denoting the correlated noise power to uncorrelated noise power (Shumway, 1971). This equation quantifies how narrowband noise (smaller \( B \)) reduces the effective degrees of freedom in the F-statistic’s capability to detect signals in noise.

Free detector parameters include: analysis time window; overlap between consecutive windows, adaptive window length for noise assessment and subsequent \( C \)-value estimation; and the p-value used for signal identification. For consistency, initial parameters used in this study are those outlined in Park et al. (2014) and detailed in Table 3.2. We use these parameters to automatically process data from the UU/SMU infrasound network for the time period of 1 January 2012 – 31 December 2012 with the AFD. In order to assess detector performance, we estimated detection rate with a simple measure:

\[
\frac{\text{number of automatically detected GT events}}{\text{total GT events}}. \tag{3.4}
\]

This measure requires that: (1) data for the acoustic channels at each array exist; (2) the signal of interest in fact propagates from the source to receiver; and (3) the absolute size of
the source is large enough to generate a signal large enough to be above the ambient noise across the array. The adequacy of these assumptions will be reviewed in the subsequent assessment of the detection statistics. Following Che et al. (2011), we associate detections with a GT event if the infrasound arrival time falls within a range derived from the ground truth seismic origin time, the distance to the array and assumed infrasonic celerities from 0.2 to 0.4 km/sec (Negraru et al., 2010) accompanied by back-azimuth estimates within ±10° of the true backazimuth. Figure 3.2 displays the resulting total number of automatically detected events at each array with an average percentage of detected events at an individual array of 51%. The number of detections varies as a function of array distance from the source with the arrays closest (NOQ at 84 km) and furthest from the source (LCM at 458 km) detecting very few events, while arrays between 140 – 300 km identifying 60 to 80% of the GT events. Many of these arrays fall within the proposed geometric shadow zone for stratospheric arrivals (Herrin et al., 2007; Negraru et al., 2010). Given the range of source-receiver separations and azimuth distribution, as well as arrivals predicted from propagation modeling, observations at arrays deployed within 200 km of the source are likely tropospheric arrivals while observations at arrays beyond 200km are likely stratospheric or thermospheric arrivals.

3.3.2. Utilizing Noise Estimates to Assess Station-Specific Missed Detections

Array-specific noise estimates are made to quantify time-varying noise conditions at each array across the network. These estimates were used to produce probability density functions (PDFs) to quantify seasonal noise estimates applicable to the data analysis, following published methodologies (Bowman, 2005; Bowman et al., 2009; Brown et al., 2014; McNamara & Buland, 2004). These PDFs represent the distribution of noise conditions at each array as a function of frequency and can be used estimate a signal-to-noise ratio (SNR) for a source of a known size at an assumed distance. Power spectral density (PSD) estimates were made
Table 3.2: Automatic processing parameters for detection.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency band (Hz)</td>
<td>0.5-5</td>
</tr>
<tr>
<td>Time window (sec)</td>
<td>30</td>
</tr>
<tr>
<td>overlap (sec)</td>
<td>15</td>
</tr>
<tr>
<td>adaptive window (sec)</td>
<td>3600</td>
</tr>
<tr>
<td>p-value</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Figure 3.2: Number of automatically detected GT events using the AFD at each array within the network
with 200 s windows and 50% overlap across 15-min of instrument-corrected infrasound data. Multiple 15-minute windows were averaged to produce a single, hourly PSD estimate. Noise PDFs are then calculated using:

$$P(T_C) = \frac{N_{P_{TC}}}{N_{TC}}$$  (3.5)

where $N_{TC}$ is the number of spectral estimates that fall within a 1-dB power bin, where bins range from 0-120 dB/Hz relative to 20$\mu Pa$ and $T_C$ is a center period. $N_{TC}$ is the total number of spectral estimates over all powers with a center period of $T_C$.

As this study focuses on the detection of repeating GT signals with known origin times, the noise analysis was restricted to time periods consistent with the GT event times (from 16:00 to 22:00 UTC). Data from all days of the week was utilized in the estimates. Infrasound noise includes both natural and anthropogenic sources which may vary between weekdays and weekends (Park & Stump, 2015) although work presented here does not try to separate these effects.

Park and Stump (2015) conducted a similar noise study for this network, producing 5% and 95% relative noise estimates without instrument correction in order to focus on seasonal changes. The expansion of this analysis using PDFs documents notable differences in noise trends as the densities include the full spectrum of noise characteristics across the network. In contrast, the 5% and 95% noise estimates illustrate the seasonal noise extremes at each station. As most of the network is comprised of IML sensors whose response rolls off below 2 Hz, characteristics of low frequency noise across the network were not accurately represented in this previous work. Using data corrected for the IML response provides absolute, low frequency noise estimates, that include quantification of the 0.2 Hz peak associated with microbaroms. The microbaroms are consistently observed across the network, as highlighted at HWU in Figure 3.3a. Noise at 0.2 Hz varies annually and appears higher during the winter months at most arrays, as a result of microbaroms generated by winter storms over
the oceans (Bowman, 2005; Drob et al., 2003; Landes et al., 2012). We interpret the strong low frequency noise peak (+10 dB) that centers near 0.2 Hz in the winter noise estimates as resulting from microbaroms that originate in the Pacific Ocean and arrive as a result of stratospheric wave propagation. We attribute the lack of a similar peak in the spring and summer noise estimates to seasonal reversals in the stratospheric jet direction, which drives a change in the dominant microbarom source from the Pacific Ocean to both the Atlantic and southern oceans. Figure 3.3b and Figure 3.4 display temporal and spatial noise variability across the network. While median noise estimates remain within the bounds of the IMS low and high noise models (Bowman et al., 2009), our noise estimates are generally high across the network relative to these bounds. Background noise peaks in the spring and decreases in the summer and fall. Relative noise levels vary across the network; levels at LCM are consistently 20 dB higher than levels at FSU or BRP. These estimates will be used to explore the relationship between the noise variance and automatic detection performance at each station. Although the two noisiest arrays (NOQ and LCM) show the lowest detection rates, we cannot easily establish a direct relationship between noise level and automatic detection performance for these two arrays, because of disparate deployment range (84 km and 458 km, respectively) and azimuth (309° and 3°, respectively) to the source. These differences motivate a more comprehensive analysis that also includes a comparative propagation path assessment.
Figure 3.3: (a) A noise PDF example capturing the seasonal (winter) noise trend at HWU. (b) Seasonal trends in median noise at three arrays across the network. Colored lines indicate the median values from the noise PDFs produced during hours of the day when GT events occurred. The dashed black line represents the IMS low noise model (Bowman et al., 2009) while the solid black line represents IMS high noise model. Spring [red]: March, April, May. Summer [blue]: June, July, August. Fall [green]: September, October, November.

Figure 3.4: Full noise models, represented at 5% (top) and 95% (bottom) noise levels at each station during the spring (red), summer (green) and fall (blue). Dashed black line indicates IMS LNM, solid black line indicates IMS HNM.
3.3.3. Utilizing Modeling to Understand Station-Specific Missed Detections

While high noise at an array may contribute to missed detections, there is the possibility that signals of interest may be missed when they do not propagate from the source to the array. Signal transmission loss over propagation distances can also reduce amplitudes, so they do not exceed background noise amplitudes.

In order to examine trends in expected signal propagation across the network, range-dependent ray-tracing through realistic atmospheric models using the infraGA acoustic ray-tracing program (Blom & Waxler, 2012) was conducted. Propagation is modeled assuming a ground elevation of 1.6 km above sea level indicative of the average station elevation across the network. Three-dimension, range-dependent ray tracing was conducted for each event using the 1-hour Ground-to-Space (G2S) (Drob et al., 2003, 2010, 2013) profiles centered at the source for the hour in which the GT event occurred. For time periods where the exact 1-hour profile was unavailable, we substituted a profile from the closest time interval (within 24 hours; see Appendix A for full list of profiles used). Ray-tracing utilized the following parameters: launch angles from 0.5 - 45°; 75 bounces in order to capture near-source direct arrivals; and an inclination step of 0.1 in order to provide a higher density of arrivals. Utilizing the expected frequency content of signal arrivals, 0.5 - 5 Hz and an average wave velocity of .3 km/sec, we estimate signal wavelengths between 0.15 - 1 km. Extrapolating for error, direct arrivals were thus assumed to be comprised of any ray that arrived within 2 km of the center of each array.

Figure 3.5a-c includes examples of predicted direct arrivals for individual GT events for spring, summer and fall. Winter is excluded as no GT events occurred during this season in 2012. The three ray-tracing examples highlight dominant seasonal trends. Figure 3.5a documents ray-tracing from early April, a transitional time between winter to summer stratospheric winds as reflected in the westward stratospheric propagation. By mid-April stratospheric winds weaken and the summer waveguide towards the east establishes between late
May or early June (Figure 3.5b). The tropospheric jet is the familiar ‘jet stream’ and is dominantly eastward in the western US with variable north and south components. Unlike the stratospheric jet, the jet steam varies on a daily to weekly scale, which is evident in Figure 1 4a-c where ray-tracing predicts arrivals consistently to the north through the spring and summer changing to the south-east in the October example (Figure 3.5c). The model predicts thermospheric returns along all cardinal directions during all three seasons; however, we require assessments of the attenuation and predicted signal amplitudes to determine whether a thermospheric signal will be observable at arrays of interest within the 0.5 - 5 Hz frequency band used for detection (Akintunde & Petculescu, 2014; de Groot-Hedlin et al., 2010; de Groot-Hedlin, 2016). Figure 3.6a-c summarizes the percentage of predicted phase types at each station and associated range in the network based on season with pie charts (Figure 3.6d-f) quantifying the total predicted arrivals across the network for each season to the right of each panel. Percentages are calculated from the total number of direct rays for each phase, derived by predicted celerity. During the spring and summer months, geometric arrivals are not predicted at arrays across the network for 50% and 48%, respectively, of the GT events. The remainder of arrivals are dominated by a combination of tropospheric and thermospheric arrivals at the arrays of interest. In the fall months, 78% of predicted arrivals are predominately tropospheric with a small number of thermospheric arrivals predicted at BRP, PSU and LCM. While there is a clear stratospheric waveguide in all three example profiles, stratospheric arrivals are not predicted at any of the arrays within the network. These seasonal differences are driven by a directional change in the dominant tropospheric jet which increase the number of tropospheric arrivals at all arrays across the network; at arrays close to the source, this corresponds to change from ‘no propagation predicted’ to tropospheric arrivals and at arrays further from the source, this corresponds to a change from thermospheric to tropospheric arrivals.

The relationships between automatic detection rates and the percentage of events with predicted arrivals vary across the network and are primarily impacted by source to receiver
distance. At arrays closest to and farthest from the source, propagation modeling indicates that sparse raypaths spatially coincide with low automatic detections. Arrivals modeled at LCM take thermospheric paths during the spring and summer months and tropospheric paths during the fall. Predictions of thermospheric arrivals during the spring and summer suggest that although models predict raypaths, signal amplitudes may attenuate such that they are too small relative to background noise levels at the receiver, as few signals with thermospheric celerities were identified. At arrays between 200-300 km from the source, high rates of automatic detection correlate with higher numbers of arriving raypaths. For arrays between 150-200 km from the source, high rates of successful automatic detection contrast with low percentages of predicted arrivals. This suggests atmospheric models do not include wave propagation path characteristics for these mid-range distances. Further modeling, possibly with refined atmospheric characterizations that considers arrival scattering sourced by wind shear (Blixt et al., 2019), may be necessary to refine the relationships between the success of signal detectors and propagation predictions at all ranges.
Figure 3.5: Direct arrivals from ray-tracing for events occurring in the spring (left) summer (middle) and fall (right). Each dot represents a predicted direct arrival from ray-tracing, while the dot color represents the celerity (i.e., predicted infrasonic phase) of the arrival. Array locations are denoted by black triangles and the GT location is denoted by the red star.

Figure 3.6: a-c: Summary of predicted ray arrivals across the network as a function of season, where grey indicates no predicted arrivals, red indicates tropospheric arrivals, green indicates stratospheric arrivals and blue indicates thermospheric arrivals. d-f: overall predicted ray arrivals across the network for each season.
3.3.4. Parameter Optimization and Analyst Review

We varied detector parameters to determine if additional signals could be identified using AFD, thereby producing a more robust set of automatic detections and providing insight into the effectiveness of the detector based on the ground truth. Following a similar study conducted in the Korean peninsula (Park et al., 2016), we ran automatic detection with p-values of 0.03, 0.05, 0.07 and 0.09. Additional detections were identified following the procedure detailed in Section 3.3.1; Figure 3.7 summarizes the number of additional GT events identified by automatic processing with each p-value increase. A total of 5 additional detections at NOQ, 4 additional detections at BRP and HWU, 3 additional detections at PSU and WMU and 2 additional detections at LCM and FSU resulted. The overall number of additional GT detections is small; however, it represents an additional 7-11% of the total GT identified, depending on the station. Note that detection with a p-value of 0.09 only added one additional detection at LCM, suggesting that automatic detection with a p-value of 0.07 and even 0.05 at some arrays is sufficient for the detection of GT events utilizing the UU-SMU infrasound network. The initial p-value of 0.01 was chosen to minimize false alarms while providing a set of automatic detections for the events; increasing the p-value in order to enhance GT detections comes with a trade-off of increasing potential false alarms as well increasing analyst burden through production of a larger overall detection dataset (Carmichael et al., 2016). Figure 3.8 compares the total number of detections for automatic processing with each p-value, derived as the number of detections that occur on each day with a GT event, with “true” GT event detections.

We compare detection numbers both across arrays within the network as well as across the span of p-values used for detection. Without knowledge of infrasound-producing sources within the region, we cannot clearly attribute increased detections to true events versus coherent noise sources. This incomplete understanding of sources limits our ability to assess the relationship between true detections and false alarms. We therefore focus on tradeoffs
between increasing GT detection rates and both analyst and processing burden. At all arrays, an increase in p-value produces an increase in GT event detections accompanied by an exponential increase in total detections. For example, at HWU detection with a p-value of 0.01 identifies 28 GT events while detection with a p-value of 0.05 identifies 31 GT events, corresponding to a 7% increase. In contrast, detection with a p-value of 0.01 identifies 471 other detections while detection with a p-value of 0.05 identifies 1235 additional detections, corresponding to a 165% increase. In the case of this ground truth study, the known source time provides a basis for separating known signals from noise. However, these comparisons demonstrate that modifying automatic detector parameters drastically increases overall detection rates, indicating that while automatic detection with high p-values is necessary for maximizing the detection of low SNR signals, it comes with a tradeoff of significant processing and analyst burden.

Analyst review of the data was undertaken to see if additional GT event detections could be identified in order to interpret and possibly improve the automated results. Analyst review identified time periods where a signal of interest consistent with the ground truth information was clearly present in the timeseries waveform data, but unidentified by the detector. These detections add to the completeness of the detection catalog, while also serving to identify detections missed by the automatic detector. In contrast to the formulation of the automatic detector, the human analyst benefited from a priori knowledge of an event occurring at a given day and time. Analyst detection utilized interactive FK array analysis to declare a detection when beamforming produced a consistent (> 10 sec) backazimuth estimate within ±10 of the source backazimuth (Che et al., 2011), in tandem with increased correlation estimates across the array as documented by the F-value associated with the estimated backazimuth. Figure 3.9 shows an example detection for an event on 2012-05-14T17:56:34UTC at BRP. The blue window in panels a-e highlights the analyst-declared detection based on consistent backazimuth estimates at 316°, an elevated signal correlation and an elevated F-value. Figure 3.7 documents the total number of supplemental analyst detections in purple.
Analyst review added a significant number of detections at all arrays within the network. As is the case for seismic detectors, these results suggest that analyst review improves automated detections.
Figure 3.7: Summary of additional event detections at each station through incremental p-value parameter increases, shown against analyst review.
Figure 3.8: Summary of relationship between total number of GT event detections compared to all detections as a function of p-value. As p-value increases, both the number of GT detections and the overall number of detections increase. Detections are defined as all detections that occur on a day with a GT event.
Figure 3.9: Waveform and array processing example for an event on 2012-05-14T17:56:34 at BRP with clear coherent background noise that was missed during automatic detection but identified during analyst review. a-d: Array processing results in the form of F-value, estimated trace velocity, estimates backazimuth, calculated correlation between array elements. Red line in panel c denotes the true backazimuth (316°) from receiver to source. e: Beamed waveform for time period where blue window represents the analyst-defined detection.
3.3.5. Identifying Mechanisms Contributing to Missed Detections

Kernel Density Estimate (Latecki et al., 2007) (KDE) analysis provides a basis to quantify automatic detection performance in an integrated way by capturing the effects of both background noise and propagation path effectiveness based on continuous background noise assessments. We apply this analysis to identify whether a series of missed detections are attributed to: (1) high noise masking the signal of interest at the time of arrival, i.e., a true ‘missed detection’; (2) significant signal attenuation along the propagation path; or (3) a combination of both.

For a given array geometry, the overall best beam signal and residual can be extracted as a means of removing incoherent noise and enhancing a signal of interest (Evers, 2008). We define the best beam to be the sum of all time-aligned traces for the slowness at the maximum detector value. This can either be done for individual frequency bins or for the maximum detector value from a set of bins. The derivations of the best beam and residual can be found in Appendix D.1 of (Evers, 2008). The assessment below compares the spectral amplitudes of the beamed waveform and the beam residual, derived for the maximum detector value across 0.5 - 5 Hz to background noise KDEs in order to identify time periods when the beamed signal spectral amplitude is higher than typical levels, due to high noise across the array, or when the beamed signal spectral amplitude is comparable to, or below the typical noise levels. The latter indicates that a lack of positive GT event detections is likely due to a lack of successful signal propagation from source to receiver.

As discussed in Section 3.3.2, we use background noise statistics that are based on waveform data from the hours during which GT events occur (16:00-22:00 UTC) to produce station-specific monthly noise estimates with a KDE algorithm, which summarizes the probability of noise amplitudes at the time of event as a function of frequency. We calculate the cumulative distribution $C_{sig}, C_{resid}$ of the signal beam and the residual spectral amplitudes $(P_{sig}, P_{resid})$ at a frequency $(f)$ of interest as:
\[ C_{\text{sig}}(f) = \int_{-\infty}^{P_{\text{sig}}(f)} \rho(f, P) dP \] (3.6)

\[ C_{\text{resid}}(f) = \int_{-\infty}^{P_{\text{resid}}(f)} \rho(f, P) dP \] (3.7)

where \( \rho(f, P) \), is the KDE estimate of the noise spectral amplitude, in dB, at frequency \( f \) under the assumption that the signal of interest is comprised of the beamed waveform with the background noise characteristics contained within the residual. The observed beamed signal or residual spectral amplitudes, \( P_{\text{sig}}(f) \) or \( P_{\text{resid}}(f) \) are derived by normalizing the PSD of the best beam and residual for amplitude (dB/Hz rel. 20\( \mu \)Pa) using the relation

\[ P_{\text{sig}}(f), P_{\text{resid}}(f) = 10 \log_{10}(PSD_{\text{sig}}, PSD_{\text{resid}}). \] (3.8)

We compute the beamed signal and residual coefficients by marginalizing the predicted noise amplitude distribution \( \rho(f, P) \) from \(-\infty\) to the observed beamed signal or residual spectral amplitudes, \( P_{\text{sig}}(f) \) or \( P_{\text{resid}}(f) \), in each frequency bin. These integrations are computed across a time window of data, defined either by an analyst-identified window where a detection occurred, or for time periods where detections were not identified, by the expected signal arrival window spanning 0.2-0.4 km/sec. We then average values across the frequency band of interest, to obtain the singular coefficient values, \( C_{\text{sig}} \) and \( C_{\text{resid}} \).

The cumulative distribution coefficients \( C_{\text{sig}} \) and \( C_{\text{resid}} \) are bounded between 0-1 and can be used to interpret how the spectral amplitude of a particular beamed signal and its residual compare to the background noise amplitude statistics represented by the noise KDE. A number of relationships between the coefficients exist:
• If $C_{\text{resid}}$ is close to 0, the residual spectral amplitude is lower than typical values. This case suggests that background noise of the timeseries is below average.

• If $C_{\text{resid}}$ is close to 1, the residual spectral amplitude is higher than typical values; suggesting that background noise of the timeseries exceeds average values.

• If $C_{\text{resid}}$ falls around 0.5, the residual spectral amplitude is equal to typical noise values.

• If $C_{\text{sig}}$ is close to 0, the amplitudes of any signal within the time series are unlikely to exceed background noise amplitudes, which may be driven by attenuation as a signal propagates from source to receiver.

• If $C_{\text{sig}}$ is close to 1, the spectral amplitude is higher than typical noise amplitudes, corresponding to a high amplitude signal that should be easily detected.

Relationships between the two coefficients can be used to interpret various detection scenarios:

• If a signal of interest is present, we expect $C_{\text{sig}} \gg C_{\text{resid}}$.

• When $C_{\text{sig}} \leq C_{\text{resid}}$, the beamed signal power is less than the measured residual, which suggests that the lack of an arrival is due to propagation effects, in relation to the time dependent noise.

• If the time series contains high background noise, $C_{\text{resid}} \geq 0.75$, we interpret that while a signal may be present, spectral amplitudes do not exceed the residual spectral amplitudes driven by increased background noise. This indicates that successful event detection is likely impeded by both background noise at the station as well as propagation-related effects.

• High background noise may also manifest in a relationship where $C_{\text{sig}} \approx C_{\text{resid}}$, with both relatively large, indicating that while a signal is present, background noise is high.
If $C_{\text{sig}} \approx C_{\text{resid}}$ and both are relatively low, we interpret that while background noise is low, beamed spectral amplitudes are low as well, indicating that signals did not efficiently propagate to the station of interest.

We use these ratios to interpret characteristics of the time series, particularly for time periods where we lack an automated signal detection from a GT event.

Figure 3.10 shows an example application of this technique to our data with two additional examples included in supplementary materials. Figure 3.10a displays the signal (black) and extracted residual (blue) spectral estimates from an UTTR event on 09-11-2012 that the automated detector successfully identified at HWU. A 2-3X increase in spectral energy in the signal compared to the residual from 0.7-4 Hz is illustrated; output from FK processing supports back-azimuth and trace velocity estimates, which remains consistent with arrivals from UTTR. Figure 3.10b depicts an example of KDE integration for the signal and residual amplitudes at the peak signal frequency of 2.3 Hz. This processing integrates the KDE from the lower limit ($-\infty$) to either the upper signal limit, or the upper residual limit, at each frequency interval; we compute the upper limit is from the maximum power of the beamed signal or residual spectra at the frequency of interest. Figure 3.10c shows the results from this integration as a function of frequency, in the form of cumulative distributions ranging from 0 to 1, with signal values in black and residual values in blue. These integrated results are consistent with visible trends in the spectra; a clear separation between cumulative distribution values is documented for the frequency band from 0.3 to 4 Hz, consistent with a high SNR signal present in the data. Figure 3.10d compares the $C_{\text{sig}}$ to $C_{\text{resid}}$ values at each frequency. $C_{\text{sig}}$ ranges from 0.1 to 1.0, while $C_{\text{resid}}$ remains near 0.2. The low $C_{\text{resid}}$ values and corresponding high $C_{\text{sig}}$ values indicate that this signal has a high amplitude and occurred during a period of low background noise. This produced an arrival with a high SNR signal that was easily identified by the automatic detector.
Figure 3.10: Example of cumulative integration processing for HWU using waveforms from an event on 09-11-2012. (a) Beamed signal (black) and residual (blue) window spectra. (b) Example of integration limits; green line denotes KDE derived from September noise model of GT hours at the frequency of interest (2.3 Hz). Black line indicates the signal upper integration limit in dB and blue indicates the residual upper integration limit in dB. (c) Cumulative distribution values for each individual frequency integration across the band of interest (0.5-5 Hz). (d) Comparison of individual residual and signal cumulative distribution values. Horizontal clustering along the x-axis indicates high signal and low noise.
Figure 3.11: Example of cumulative integration processing at HWU using waveforms from an event on 06-26-2012. a. Beamed signal (black) and residual (blue) window spectra. b. Example of integration limits; green line denotes KDE derived from June noise model of GT hours only at the frequency of interest (1.4 Hz). Black line indicates the signal integration limit in dB and blue indicates the residual integration limit in dB. c. Cumulative distribution values for each individual frequency integration across the band of interest (0.5-5Hz). d. Comparison of individual residual and signal cumulative distribution values. Horizontal clustering along the x-axis indicates high signal and low noise.
Figure 3.12: Example of cumulative integration processing at HWU using waveforms from an event on 07-31-2012. a. Beamed signal (black) and residual (blue) window spectra. b. Example of integration limits; green line denotes KDE derived from June noise model of GT hours only at the frequency of interest (0.6 Hz). Black line indicates the signal integration limit in dB and blue indicates the residual integration limit in dB. c. Cumulative distribution values for each individual frequency integration across the band of interest (0.5-5Hz). d. Comparison of individual residual and signal cumulative distribution values. Horizontal clustering along the x-axis indicates high signal and low noise.
Integrated KDE results for all arrays where events were not detected, averaged across the 0.5-5 Hz band, are displayed in Figure 3.13. Examination of the relationships between $C_{\text{sig}}$ and $C_{\text{resid}}$ at each array provides insight into the mechanisms contributing to missed detections. At most arrays, two distinct ‘clusters’ group missed detections: one in the upper righthand quadrant that corresponds to missed detections due to high noise (dashed blue ellipse) and one in the lower left quadrant that corresponds to missed detections due to propagation effects, circled in dashed green.

The three examples illustrated in Figure 3.10, Figure 3.11 and Figure 3.12 show that these clusters are useful in assessing sources of a missed detection. For events where a signal was not automatically detected, the missed detection cannot be attributed to high noise if the average $C_{\text{resid}}$ value is low and is therefore expected to be attributed to propagation effects. Alternatively, if the average $C_{\text{resid}}$ value is high, the missed detection can be attributed to high noise (either coherent or incoherent) across the array. We consider it likely that there is overlap in these cases and some detections may be missed due to a combination of both factors. The use of this methodology provides insight into the phenomena causing missed detections and may assist in future automatic detector improvements. It also provides a tool for assessing a network of arrays and input for improving a set of regional arrays for improved monitoring.
Figure 3.13: Overall cumulative signal and residual values at each array, determined by the average value across 0.5-5 Hz. Red dots are values calculated for missed detections.
3.4. Discussion

3.4.1. Establishing a Detection Baseline

An automatic detection catalog was produced that applied a realistic set of processing parameters in order to optimize the detection of a series of GT events and subsequently produce a physical interpretation of automatic detector limitations. These baseline metrics are useful to assess network detection performance by providing a measure of how well the detector identifies signals that propagate from the source to the receiver. In total, the automatic detector that was parameterized with a grid of p-values successfully identified between 16% to 85% of GT events across the network, with the lowest rate of successful detection occurring at NOQ, the closest and noisiest station in the network. Subsequent analyst review evaluated if signals visible to the analyst were missed during automatic processing.

Figure 3.9 shows an example of an event that was missed by the detector but identified by the analyst. In general, signals that the automatic detector missed, but that were identified by the analyst show F-values that exceed background F-values by only one to two units. This marginal signal strength indicates low signal coherence across the array, as well as low SNR signals. Unlike the automatic detector, the analyst did not apply a duration threshold and identifies short duration, <10 s, signals that may not be long enough in duration to trigger the detector. Once automated processing identifies an infrasound event, additional arrivals may be isolated and identified by careful analyst review within the context of a priori analyst knowledge of both predicted arrival times as well as known backazimuths between receiver and source. The successful identification of additional signals from GT events suggests that the analyst technique of utilizing a priori origin time and location knowledge to predict signal arrival times could be included in automatic detection processing in an iterative effort to improve detection rates. We anticipate that this hybrid technique would increase true
detection rates by considering not only the highlighted signal characterizations, but also in some cases a priori information relative to a known set of sources or source locations.

Additionally, results indicate that p-values greater than 0.01 could be used for optimizing detection of UTTR events. In the context of infrasonic detection studies, it is extremely difficult to separate what constitutes a true detection from what is considered to be a false alarm. Arrowsmith (2018) defines a monitoring false alarm (MFA) to be any event hypothesis that is physically correct but where the event itself is not of interest to explosion monitoring, while an event of interest is defined to be any large transient event that could be consistent with an atmospheric or underground explosion. Without a formal definition or understanding of additional sources of infrasound within the western US region, we utilize overall detection numbers to illustrate the relationship between p-value and increased analyst burden. Automated processing with a variety of p-value produces a more robust detection data yet comes with the tradeoff of a significant increase in the number of daily detections which leads to additional processing and analyst burden. Our data show that increases in non-GT detections can exceed GT detections by more than a factor of 20. Fortunately, known source characteristics (origin time and location) in the case of UTTR GT events allow for a targeted assessment of detections, which alleviates most of the increased analyst burden. Without this source information, the benefit of additional event detections may be outweighed by the burden to analyst workload. Processing following the methodologies outlined may not be realistic for monitoring situations with unknown sources.

3.4.2. Event Yield and Detection Capability

Investigation of a relationship between event yield and detection capability was motivated by a similar study of infrasound detection at regional distances (McKenna et al., 2007). This study utilized iron mine explosions between 9100-45500 kg to evaluate performance of a singular infrasound array at 390km from the source, with no documentable relationship
between positive event detection and explosion yield. As the McKenna et al. (2007) study only used one array, we use the variable distance distribution of our network to examine this relationship at variable source to receiver distances.

Figure 3.14 document the relationship between array distance and event yield, determined as the percentage of successfully detected events for each explosion size. Following scaling convention from Denny & Johnson (1991), we present explosion yield as the cube root of given event yields. GT events were binned into five distinct yield ranges and percentages are evaluated as the number of successful detections at each station for the number of events within each bin. As previously shown, detection capability is driven by station distance from the source, which is reflected in the yield evaluation. In Figure 3.14a, we show similar distance-driven patterns in the yield comparisons, where at stations near to and far from the source, there are higher rates of automatic detection for events with larger yields. At stations between 100-300 km from the source, there is no clear relationship between station-level detection capability and event yield across the network. In Figure 3.14b, we demonstrate that while analyst detections add to the completeness of the detection dataset, there again is no clear relationship between detection capability and event yield. These figures demonstrate that there is no clear dependence between array distance from source, event yield and detection capability; however, we note that this relationship is only evaluated over a small set of event yields and at regional distances between 84-458 km from the source. Therefore, this conclusion is only valid within the range of explosion yields demonstrated.
Figure 3.14: (a) Comparison between station distance from source [km] and number of successful automatic event detections, as a function of event yield in kg\(^{1/3}\). Marker color and size represent the percentage of GT events within each yield bin, the total number of events within each yield bin are shown in the histogram along the y-axis. (b) Comparison between station distance from source [km] and number of successful total (automatic and analyst) event detections, as a function of event yield. Marker color and size represent the percentage of GT events within each yield bin. Missed detections at each station categorized by phenomena that contributes to missed detections.
3.4.3. Understanding True ‘Missed’ Detections

Following automatic detection and supplemental analyst review, we analyzed signal spectral amplitudes using KDEs that are based on background noise estimates. This analysis provided a statistically-justified understanding of missed detections across the network. Figure 3.15 illustrates that the mechanisms that led automatic detection misses, which we later identified through automatic detector parameter changes, vary across the network (blue and green hashed areas). These comparisons suggest that across the network, the initial p-value of 0.01 used for automated detection does not successfully detect signals of interest in situations where propagation effects lead to low energy arrivals at a station, resulting in decreased signal-to-noise ratios (SNR) that are insufficient to trigger the power-based AFD. At some arrays, the initial p-value used for detection does not successfully detect signals of interest in high noise environments. Optimal parameter choices for automatic processing may need to vary across a network as a function of signal propagation distance as well as station noise characteristics.

We attribute detections that are missed by automatic processing, but later identified through analyst review, to be due to both high noise and propagation effects at arrays across the network (Figure 3.15, blue and green dashed bars). These results demonstrate that analyst review is necessary for producing a robust GT detection set, as automatic processing utilizing the AFD misses a number of signals across the network. These signals are missed due to both high noise at the array, which reduces the signal coherency to levels insufficient for triggering the automated detector and due to insufficient propagation paths between the sources and receivers, leading to low signal amplitudes which similarly are insufficient to trigger the power-based detector.
Figure 3.15: Missed detections at each station categorized by phenomena that contributes to missed detections. Missed detections due to noise are in blue while missed detections due to propagation effects are in green. Detections are separated based on how they were identified: detections identified by parameter optimization are dashed, detections identified by the analyst are hashed and true missed detections are solid.
Additionally, in some cases, cumulative distribution values for events that were successfully identified during initial automatic detection processing fall within both the dashed blue ‘high noise’ circles and the dashed green ‘poor propagation’ circles seen in Figure 3.13. These distributions are assumed to be due to the methods used for producing them; the analysis considers spectral amplitudes but does not include the coherency of the residual. Events that fall within ‘missed detection’ clusters may contain coherent noise that increases the detection threshold and such cases cannot be resolved or accounted for in the current methods. A more in-depth analysis of the background noise including both the amplitude and coherence structure is needed to more fully characterize the detection statistics. We concede that the analysis in this paper focusing on amplitudes is only a useful first step for further understanding the characteristics of automated, analyst and missed detections.

We deemed the remaining detections as ‘true’ missed detections because neither automatic processing, parameter optimization, nor analyst review identified these events. Utilizing cumulative distribution values from the KDE analysis, we determine that between 20-35% of missed detections at the closest arrays (NOQ, HWU) and 21-39% of missed detections at arrays furthest from the source (PSU, LCM) are primarily due to propagation effects. High noise drives between 3-5% of missed detections at these arrays. In contrast, missed detections at arrays between 125 – 300 km from the source can be attributed primarily to high noise, between 11-25% of missed detections while 5-8% of missed detections at these arrays are due to propagation effects. There are no identifiable seasonal trends in missed detection rates, despite significant variability in both seasonal atmospheric propagation and noise levels across the network.

Although LCM and NOQ have the highest noise levels, analysis of propagation predictions and typical noise levels at the times of events implies that missed detections at these two arrays remain primarily caused by propagation effects instead of noise-related effects. This suggests that noise assessments alone cannot be used to completely assess array capability.
Conversely, the noise models shown in Figure 1 3b demonstrate that at some of the quieter arrays within the network (BRP, HWU, PSU and WMU), a majority of missed detections are due to uncorrelated noise. This result indicates that despite noise levels being low on average, these arrays are sensitive to incoherent noise that is most likely attributable to wind gusts (Berglund et al., 1996; Shields, 2005).

We suggest missed detections due to propagation effects relate to typical atmospheric propagation across the network. In general, we predict tropospheric arrivals to the NNW-NNE of UTTR, resulting in direct phase arrivals at individual arrays for 30 to 70% of the time. We expect stratospheric arrivals to the west during summer months and the east during winter months, but do not predict such signals to propagate directly from the source to any receivers within the network, given the dominantly north-south distribution of the network and the lack of detonations in the winter. Our modeling also predicts thermospheric arrivals in all directions with varying degrees of attenuation. The high number of missed detections at arrays within 200 km of the source (NOQ, WMU, HWU) can be attributed to a lack of conducive tropospheric duct for propagation, or to GT events occurring during time periods when the tropospheric duct narrows or changes direction. Similarly, we can attribute missed detections at LCM that locates 458 km from the source to one of three effects: either (1) GT events occur during time periods where the predicted thermospheric duct range terminates prior to 458km; (2) excessive attenuation along thermospheric propagation paths; or (3) signals arriving within a frequency band than mismatches with that used by automatic detection.

The combination of automatic detections and analyst picks provides statistics on signal propagation across the network with observable signals at a minimum of one receiver within the network for 91-96% of the GT events, depending on the station. Propagation modeling statistics derived from Figure 3.6d-f indicate that signals would propagate from the source to a single receiver for between 55-80% of the events and suggests that either
some of the propagation paths are non-geometric (e.g., scattered from fine scale structures in the atmosphere) or that atmospheric models do not fully capture the variability for the regional propagation distances used in this study. Others document a similar discrepancy between predicted and observed infrasound arrivals, particularly at arrays within the classical ‘shadow zone’ (Herrin et al., 2007; Negraru et al., 2010), which we presume is caused by small-scale variabilities in the atmosphere, wind shear, or internal gravity wave perturbations to the smoothed background atmospheric profile (Blixt et al., 2019). Finally, several studies of sources with infinite frequency theory thoroughly documents that geometric ray-tracing does not accurately predict stratospheric arrival tails from large explosive events (Nippress et al., 2014; Vergoz et al., 2018). These limitations may explain the discrepancies between observed and predicted arrivals at stratospheric distances, i.e., BRP and LCM. More extensive analysis of individual arrivals is needed to fully understand the phase characteristics of arrivals from UTTR events.

3.5. Conclusions

The systematic detection and characterization of a set of arrivals from GT sources in the western US throughout the spring, summer and fall of 2012 is used to evaluate factors influencing infrasonic signal detection and automatic processing performance across a network of regional infrasound arrays. Automatic detection followed procedures and parameters that have been previously tested for the detection of regional infrasonic signals utilizing the AFD in both the western US and the Korean Peninsula. Results indicate that automated processing with a p-value of 0.01 is insufficient for identifying events of interest at this particular regional network where arrays are located between 84-458 km from the source and most fall within the stratospheric shadow zone. Although increasing the p-value for automatic processing successfully identifies more events, it comes with a significant processing and analyst burden, quantified in Figure 3.8 by the cumulative number of daily detections. In this study, analyst burden is alleviated through knowledge of the event origin time and
location. Processing following the procedures outlined within this manuscript may not be applicable for monitoring purposes when sources are unknown.

A subsequent analyst review increased the percentage of successful detection to between 50-90% of GT events, depending on array. Analyst-identified detections differ from automatic detections in the following ways:

- Generally low in signal power (low SNR).
- Low signal coherence across the array.
- Arrivals are short duration of 10 sec.

Results also indicate that analyst review is necessary for producing a robust GT detection set, as automatic processing utilizing the AFD misses a number of signals across the network. The analyst was aided significantly by source information, in particular a priori knowledge of predicted signal arrivals.

The combination of analyst review and KDE analysis of spectral amplitudes demonstrates that signals may be missed when propagation conditions are poor from the source to the receiver, reducing signal amplitudes and resulting in insufficient signal strength for the array to detect. We predict that arrivals at arrays that are far from the source are highly attenuated thermospheric arrivals, which may lack the signal strength to exceed high correlated background noise. Similarly, arrays within 100 km of the source detect very few events due to both high background noise and lack of conductive propagation paths from source to receiver.

Although seasonal trends in propagation, noise and detections from unidentified sources are well-documented (Green & Bowers, 2010; Le Pichon et al., 2008, 2012; Nippress et al., 2014; Park et al., 2014), detections of GT events within this study do not follow any clear
seasonal trends. Prior studies (Park & Stump, 2015; Park et al., 2014) identified seasonal trends in detections and noise due to both seasonal changes in source distribution and atmospheric waveguides (Evers & Siegmund, 2009). Alternatively, this study focuses on repeating events from a stationary source. The lack of seasonal detection trends is due to predominately tropospheric and thermospheric arrivals at arrays within the network; the paths for these arrivals are less seasonally dependent than stratospheric paths. This study represents the first quantification of AFD performance utilizing events with GT information as well as the first assessment of performance in terms of coherent and incoherent noise types. As this study mirrored automatic detection procedures from the Park et al. (2014) study, the higher rates of successful GT event detection across the network indicate that the lack of GT events present in the event bulletin is due to failures within either automatic association or location procedures and not failures in the ability of the AFD to identify GT events originating at UTTR.

We find no direct relationship between event yield and detection capability, i.e., larger events are not consistently identified by more arrays across the network. This observation complements results from McKenna et al. (2007), where a regional infrasound array located 390 km from a mine in Minnesota was able to detect mining blasts between 20,000-100,000 lbs (9100-45500 kg, for comparison) with no identifiable relationship between detection capability and explosion yield. We note that this conclusion is only valid for the relatively small range of yields within this study. Results indicate that successful event detection at regional distances is driven primarily by variations in wind conditions and atmospheric conditions effecting signal propagation and is less dependent on event yield.

Lastly, we offer a robust catalog of GT events (provided as Appendix A) and signal detections that were produced from this study for subsequent association and localization studies. The evaluation of mechanisms leading to missed detections that include noise effects at the site, or atmospheric propagation from the source to the receiver, can be exploited in
a variety of ways. We assert that additional research that exploits this catalog can improve
detection algorithms through subsequent adaption for both correlated and uncorrelated noise,
as well as documenting the propagation of known signals from a GT source.

3.6. Acknowledgments

The authors acknowledge the support of the National Nuclear Security Administration
Office of Defense Nuclear Nonproliferation R&D for funding this work. Los Alamos National
Laboratory completed this work under the auspices of the U.S. Department of Energy. In-
strumentation support came from the Incorporated Research Institutions for Seismology
(IRIS) Program for Array Seismic Studies of the Continental Lithosphere (PASSCAL) pro-
gram as well as from Southern Methodist University. The University of Utah in cooperation
with Southern Methodist University installed and operated the arrays as well as provided
information and data. Data were distributed and archived at the IRIS Data Management
Center under DOI https://doi.org/https://doi.org/10.7914/SN/YJ_2010. Junghyun Park
completed seismic analysis of UTTR arrivals at BGU, which enhanced the completeness of
the GT event dataset. The authors would like to thank the editor and reviewers of this
manuscript, whose detailed comments significantly improved the structure and focus of the
paper.
**Abstract** Recent improvements to the Bayesian Infrasonic Source Localization (BISL) methodology account for realistic infrasound signal propagation across a region in order to improve the precision and accuracy of spatial and temporal source localization estimates. As a test of these advancements, the location capabilities of a regional infrasonic network of stations located between 84 – 458 km from the Utah Test and Training Range, Utah, USA, is assessed using a series of near-surface explosive events with complementary ground truth information. Ground truth origin times are refined using a nearby seismic station while the ground truth source location is known from blasting logs. Signal arrival times and backazimuth estimates are determined with an automatic F-statistic based signal detector and compared to estimates from an analyst waveform review in order to address the efficacy of the proposed automated tools. A singular celerity and backazimuth deviation model is constructed using ray tracing analysis based on an extensive archive of historical atmospheric specifications. Similarly, a set of multi-variate, season and location specific models are constructed and compared to an empirical model that depends on the observations across the infrasound network and the ground truth events, which accounts for atmospheric propagation variations from source to receiver. Spatial location accuracy is driven by a combination of signal propagation model and the azimuthal gap of detecting stations, where increased station azimuthal gaps or discrepancies between observed and predicted signal celerities result in location results with poor accuracy. These observations indicate that detection backazimuths may drive location biases over travel time estimates. The empirical model

---

0Dannemann Dugick, F., Blom, P., Stump, B., Hayward, C., Carmichael, J, Arrowsmith, S. and Marcillo, O. *submitted to Geophysical Journal International*
improves both spatial localization precision and accuracy; all estimates retain the true GT location within the 90% confidence bounds while average mislocation is reduced to 15.49 km and average 90% error ellipse areas reduce to 4141 km². The empirical model additionally reduces origin time residuals; origin time residuals from the other location models were in excess of 160 seconds while residuals produced with the empirical model were within 20 seconds of the true origin time.

4.1. Introduction

Accurate infrasonic event location estimates contribute to monitoring many types of events, including nuclear explosions. Research developments in location approaches include several proposed regional infrasonic location methodologies such as the Least Squares inverse method (Le Pichon et al., 2009), the Reverse Time Migration method (Walker et al., 2011), the Bayesian Infrasound Source Localization (BISL) method (Modrak et al., 2010) and the Probabilistic Global Search approach (Arrowsmith et al., 2020; Koch & Arrowsmith, 2019). In addition, interest in the location of sources of volcanic signals observed at closer distances has prompted development of local-scale methodologies (Johnson, 2018; Kim & Lees, 2014; Sanderson et al., 2020). These regional and local methods invert signal detection time and/or backazimuth measurements to produce an estimate of a spatial location and origin time. A simplifying but critical assumption implicit in many of these methodologies is that location estimates can be based upon an assumed single infrasound celerity value despite the time variation of infrasonic propagation in realistic atmospheric characterizations. This simple atmospheric assumption approximates the velocity of infrasonic waves propagating through the stratosphere with a constant celerity of 0.3 km/s which introduces significant errors in localization estimates as infrasonic energy is refracted by multiple ducts within the atmosphere. Additional arrivals from tropospheric (0.33 km/sec) or thermospheric (0.25 km/sec) ducts are common and further degrade locations based on this simplifying assumption.
Historically, errors in seismic arrival time predictions have been reduced through the application of waveform correlation techniques to estimate signal onset times, improved velocity models, station corrections and assessment of azimuthal gaps in network coverage (Bondár & McLaughlin, 2009; Yang et al., 2004). However, reductions in infrasound location errors are more complex due a dynamic and sparsely sampled propagation medium. In a realistic atmosphere, infrasonic energy refracts as a result of both temperature and wind gradients making the characterization of propagation effects more difficult due to the time-varying properties of the atmosphere. Despite advances in atmospheric data collection, model creation and modeling techniques (Drob et al., 2003, 2010), atmospheric predictions remain poorly resolved at the scales required to interpret signal propagation at regional distances. Two distinct approaches exist to account for this uncertainty. First, comparisons between observed and predicted signals based on propagation modeling can be used to update and ideally improve atmospheric estimates (Arrowsmith et al., 2013; Blom & Marcillo, 2017; Drob et al., 2009; Evers et al., 2014; Lalande et al., 2012; Le Pichon et al., 2005a; Vanderbecken et al., 2020). Second, stochastic models can be constructed by modeling propagation through a large number of historic atmospheric specifications, with the goal of defining distributions of arrival characteristics (Blom et al., 2015, 2018; Marcillo et al., 2013; Morton & Arrowsmith, 2014). Such predictions can be used to better assess localization methodologies by including improved uncertainty estimates based on assumed atmospheric propagation variability.

One such approach is presented in Blom et al. (2015), that documents the mathematical and theoretical framework for including spatially and seasonally specific propagation models (path geometry models, PGMs) that may improve localization estimates based on regional signals. Results from previous regional (400-1000 km) location studies in the western US (Blom et al., 2015; Marcillo et al., 2013) suggest that application of these seasonal PGMs reduces the area of event error ellipses. These analyses did not include observations within 300 km of a source, where stations are within the classical stratospheric shadow zone and tropospheric arrivals are expected. At these distances, signal propagation can vary on hourly to
daily scales, as compared to the longer-scale variability observed within the middle-and upper atmosphere layers and recorded at stations further from the source (Baldwin & Dunkerton, 2001; Garcés et al., 1998; Sorrells et al., 1997).

Assessment of event localization capabilities benefits from “ground truth” events which include high-confidence epicentral estimates combined with high-quality arrival times (Green et al., 2010; Yang et al., 2004). Seismoacoustic event bulletins (Park et al., 2014; Walker et al., 2011) indicate that the Utah Test and Training Range (UTTR) is a major source of infrasound signals in the Western United States. Despite this location of a known repeating source, the automated infrasonic event bulletin produced by Park et al. (2014) includes only 5 of the 47 ground truth events from UTTR for the year of 2012. The small percentage of events in the Park bulletin motivates this study to understand the cause(s) of the missed events, which may include the methodologies used to detect signals of interest, to associate detections across the network and to locate the events using the atmospheric conditions at the time of the event. Recently, Dannemann Dugick et al. (2020) documented that at a station-level, automatic detectors successfully identify between 14-80% of signals associated with these UTTR GT events, while subsequent analyst review increased successful detection to 24-90%, depending on event. An evaluation of mechanisms contributing to missed detections found that signals were missed for three primary conditions: first, both coherent and incoherent high amplitude noise levels suppressed signal observations across an array; second, propagation path effects where efficient ducts for propagation from source to receiver did not exist or signal attenuation reduced signal amplitudes to below background noise levels; and third, some combination of one and two.

As Dannemann Dugick et al. (2020) mirrored automatic detection procedures from Park et al. (2014), the high number of successful GT event detections across the network suggests that the lack of GT events in the event bulletin may be primarily due to failures in automatic association or location procedures and not arrival detection. Park et al. (2014) estimated
infrasound event locations using the original BISL methodology (Modrak et al., 2010); where significant bias in the location estimates was introduced with by the assumption of a uniform velocity probability distribution in the Bayesian location algorithm. While these errors may be reduced by utilizing a generalized celerity probability model as detailed in Blom et al. (2020), propagation path uncertainties remain across the network. Therefore, this study focuses on the association of signals and subsequent locations of the 47 UTTR events from 2012 in order to explore improvements to the automatic infrasonic processing pipeline utilized in Park et al. (2014) and quantify improvements to localization estimates based on model-based predictions using the BISL method.

We will first focus on recent improvements to the BISL methodology as illustrated in Blom et al. (2020) and Blom et al. (2015), who apply modeling-derived atmospheric models to better account for the atmosphere at the time of an event in order to improve the event localization estimate. Next, we detail the robust set of signal detections for UTTR GT events from a regional infrasonic network in the Western US. Third, we compare location error estimates using the generalized celerity model outlined in Blom et al. (2020) and the monthly path geometry models (PGMs) as outlined in Blom et al. (2015). Finally, we introduce a new data driven empirical celerity model that is also used to assess localization estimates. Results are synthesized in order to provide recommendations of best practices for infrasonic localization at near-regional distances.

4.2. Localization Methodology Overview

Current regional infrasonic location methodologies include a Least Squares inverse approach (Le Pichon et al., 2009), a Reverse Time Migration method (RTM) (Sanderson et al., 2020; Walker et al., 2011), the Bayesian Infrasound Source Localization method (BISL) (Modrak et al., 2010) and the Probabilistic Global Search method (PGS) (Arrowsmith et al., 2020; Koch & Arrowsmith, 2019). The Least Squares Methodology associates infrasonic detections with similar consistent azimuth, trace velocity, frequency and time characteristics
to the same source. A critical simplifying assumption is that the atmosphere is characterized by a constant celerity of 0.3 km/sec that approximates the celerity of infrasonic waves propagating through the stratosphere. The non-linear iterative least squares inversion algorithm then uses this velocity profile to estimate source locations. The assumed static atmosphere with constant stratospheric velocities is limited as it excludes observed seasonal reversals of stratospheric winds.

RTM relies on phase coherence between acoustic waves in order to migrate signals back in time and space to the point where they constructively interfere (Walker et al., 2011). The methodology detects and locates acoustic events by assuming a propagation velocity model and predicted travel times from all possible sources to the infrasonic arrays. Walker et al. (2011) performed RTM using signals recorded by the 2007-08 USArray Transportable Array (TA) and identified infrasonic hot spots in the western United States, while Sanderson et al. (2020) applied RTM to detect and locate remote explosive volcanic eruptions in Alaska utilizing the 2016-2017 TA data.

PGS is a Bayesian approach with a grid search using either/both seismic and acoustic arrival times and backazimuths to produce a set of location estimates that provide a full assessment of errors with no simplifying assumptions (Arrowsmith et al., 2020). Application of this methodology that combines seismic and infrasound arrivals can complement complexities of infrasonic propagation estimates as demonstrated by locations of mining blasts in the state of Utah, United States (Koch & Arrowsmith, 2019).

4.2.1. Updates to the BISL Methodology

The adaptive BISL algorithm utilized in this study combines likelihood equations for backazimuth and travel time observations and associated variances to assign a combined likelihood value for the location parameters (Modrak et al., 2010). Similar to the Least Squares
and RTM approach, the largest assumption inherent in the BISL algorithm as outlined by Modrak et al. (2010) is that it uses a single infrasound celerity value. This single celerity is an approximation to the true time varying celerity for infrasonic propagation through the atmosphere. Marcillo et al. (2013) document improvements to the BISL method by incorporating semi-empirical models based on prior information in the form of modified prior PDFs. Blom et al. (2015) further improved BISL with the implementation of azimuth-deviation and celerity-range prior information, which enhances the precision of source locations. Most recently, developments by Blom et al. (2020) describe a generalized celerity model which weights the probability of thermospheric, stratospheric and tropospheric arrivals in order to more accurately account for signal propagation across a network, thereby producing more precise localization estimates. Here, we evaluate the implications of applying a variety of models to more accurately estimate both arrival celerity and backazimuth deviation for localization purposes. The updated BISL methodology provides the opportunity to improve estimated event location and origin time with accurate distributions of infrasonic phases. Appendix B details the mathematical framework of the BISL method.

4.2.2. Celerity and Backazimuth Deviation Models

4.2.2.1. Generalized Celerity Model

Blom et al. (2020) introduce a generalized celerity model using a three-component Gaussian mixture model (GMM) fit to ray tracing predictions through historical atmospheric specifications. These specifications include contributions from thermospheric, stratospheric and tropospheric waveguides. This generalized celerity model is constructed by combining arrivals predicted by ray tracing across all ranges, azimuths and seasons and defines the distribution of arrival celerities for a source at UTTR as observed at locations within 1000 km. A fixed standard deviation for backazimuth, of 4°, is computed from this analysis taking into
account the effects of cross winds. The full details of the model and its construction appear in Blom et al. (2020). Within this manuscript, the generalized celerity model is referred to as the Blom et al. (2020) model.

4.2.2.2. Seasonal Specific Path Geometry Models (PGMs)

Seasonal changes in the atmosphere affect infrasound source-to-receiver propagation and thus drive errors in epicentral source locations if not properly taken into account. Source location methods that capture temporal variability in the atmosphere (i.e. Marcillo et al. (2013)) correct static models for deviations in backazimuth and celerity. The azimuthal dependence of propagation can be taken into account by grouping propagation paths along similar launch azimuths, assuming a stratified propagation medium. Using this approach, range-dependent celerity estimates are extracted across 16 azimuthal bins. Azimuthal deviation priors are calculated from the mean and variance of the azimuth deviation for the propagation predictions within each azimuthal bin. Spatial and seasonal specific models are then constructed by defining range and azimuth varying weights, means and standard deviations in the GMM as well as means (bias) and standard deviation in the arrival back azimuth. These models are constructed using overlapping 50 km wide bins spaced at 10 km intervals for the same suite of atmospheric specifications utilized to produce the generalized celerity models. The use of a 1D GMM with range and azimuth dependence differs from the methodology outlined in Blom et al. (2015), where celerity probabilities were defined using a 2D GMM with a relatively large number of components. Comparison of localization analyses with both parameterizations have found that the use of a 1D GMM improves numerical efficiency while retaining localization accuracy. For a specific detection, individual azimuth bins are selected by identifying the bin nearest to the inferred launch azimuth of the detection. In each case, we assume a stratified propagation medium and $\phi_{\text{launch}} = \phi_{\text{backazimuth}} \pm 180$
(Blom et al., 2015). Within this manuscript, monthly propagation celerity and backazimuth models are applied and are referred to as PGMs.

4.2.3. Evaluating Errors in Localization Estimates

While infrasonic localization methodologies are similar to seismic localization methods, the dominant errors that impact the location estimates differ, although Arrowsmith et al. (2020) suggest an approach to combine these differences. Both phenomenologies produce location errors resulting from errors in travel times and back-azimuths (particularly important for infrasound), which can be reduced through improvements to the propagation models. The goals in localization are to improve location accuracy by removing model bias while reducing uncertainty, thereby providing a 90% coverage ellipse where 90% of the true locations are found inside the estimated error ellipses (Yang et al., 2004). The crux of this study is that these estimates can be assessed using reliable GT events which have high-confidence epicentral estimates and estimates of errors for arrival time estimates.

Trade-offs between location precision and accuracy are assessed with comparisons between ground truth locations to estimates from the distinct celerity models used for localization. Location accuracy compares the estimated source location with distance in kilometers from the known GT source as well as the difference in estimated source origin time in seconds from the known GT source time (Bondár & McLaughlin, 2009). Location precision refers to error estimates associated with the location procedure, evaluated as either the 90 or 95% confidence ellipses or the arrival time error bounds relative to the known GT source (Arrowsmith et al., 2015; Bondar et al., 2008; Bondár et al., 2004; Flinn, 1965; Yang et al., 2004).
4.3. Regional Infrasound Network and Detection Dataset

4.3.1. Ground Truth Data Set

Signals associated with a set of ground truth events with known source location, origin time and yield are used to assess the capabilities of infrasound networks as well as quantify the performance of event detection and location methods (Green & Bowers, 2010). Within the Western US, the Dugway Testing Ground (DTG), New Bomb (NB) and Utah Test and Training Range (UTTR) regularly generate infrasound signals sourced by explosions (Figure 4.1). UTTR is the largest continuous block of supersonic-authorized restricted airspace, and is frequently used to dispose of explosive ordinance in the Western US (Walker et al., 2011) and is therefore the focus of this study.

Infrasound signals from UTTR are used in this study due to the availability of accurate blast logs which include origin time and explosive yield. Arrivals from these events provide the basis to evaluate localization estimates focusing on observations at near-regional distance. In 2012, 47 missile motor or propellant destruction shots were conducted at UTTR with varying yields (Dannemann Dugick et al., 2020; Park et al., 2014). These GT events are detailed in Appendix A. Origin times were verified using seismic arrivals at the closest seismometer, BGU (black triangle in Figure 4.1), deployed approximately 26 km from UTTR. A lack of seismic arrivals for two of the listed GT events suggests that although blast logs list an event for that time period, there may be either (1) be an error in the blast log or (2) an event so small that it did not generate observable seismic or infrasonic signals. We therefore proceed with a dataset of 45 GT events for analysis.
Figure 4.1: Infrasonic network used for producing event localization estimates, where red star indicates GT source (UTTR) and blue triangles indicate arrays used for localization. Black triangle indicates BGU, the closest seismic station to the source.
4.3.2. Detection Dataset

The Dannemann Dugick et al. (2020) detection dataset consists of detection times and backazimuth estimates, derived from automatic processing utilizing the Adaptive F-Detector (Arrowsmith et al., 2009). The arrival times and backazimuths are refined by an analyst, following a methodology detailed in Dannemann Dugick et al. (2020). The analyst review adds additional GT detections at all stations and particularly increases the robustness of the detection dataset at station NOQ. There are three significant differences between analyst and automatic detections, which are driven by limitations in the power-based automated detector. First, analyst detections generally have lower signal power (low SNR). Second, analyst detections typically have lower signal coherence across the array. Third, analyst arrivals are short duration (<10 sec) compared to automatic detections with durations between 30-90 sec.

Successful detection of GT events varies across both the network and the GT events. Similar to McKenna et al. (2007), there is no indication that detection success is related to source yield, across a limited range of yields. Additionally, although there are documented seasonal trends in infrasonic detections (Green et al., 2010; Le Pichon et al., 2009; Park & Stump, 2015), this detection dataset does not document seasonal detection variability, although it does not cover all seasons. This effect is likely driven by both the uneven distribution of GT events in time and the network geometry, where most propagation paths are either tropospheric or thermospheric and thus may not be as strongly affected by seasonal stratospheric jet reversals.

The number of detections at each station is documented in Figure 4.2. Figure 4.2b details the total number detections across the network per GT event. A variety of detection trends are visible across the network and discussed in detail in Dannemann Dugick et al. (2020). Trends are driven by the network configuration where stations are located between 84-458 km of the source and all but one station is to the S-SE of the source. One station,
HWU, provides azimuthal resolution towards the NE but the network lacks coverage to the west. The resulting detection azimuthal gap is highlighted in Figure 4.3. The observed detection capability varies across the network and is attributed to high noise levels at individual infrasonic arrays combined with signal propagation path effects. Overall the number of detections are driven by source to station distances with the largest number of signals detected at stations between 100-300 km (FSU, WMU and HWU) and the fewest signals at the closest station, NOQ; 84 km and furthest station, LCM; 410 km, from the source.

4.4. Application of the Methodology

4.4.1. Evaluating Localization Improvements

In order to evaluate and compare improvements to localization through the application of various celerity and backazimuth deviation predictions, we utilize the maximum a posteriori (MAP) solution which maximizes the posterior Probability Density Function (PDF) produced by the BISL methodology. This solution corresponds to best spatial and temporal localization, given the set of detections. Utilizing the Blom et al. (2020) generalized celerity model, we successfully locate all 45 GT events, with an average mislocation of 14.34 km, an average 90% error ellipse (EE) area of 3653 km² and an average time residual of -62.37 sec (Appendix B). We note that 44/45 spatial estimates retain the true GT location within the 90% EE bounds. The MAP location estimates (black circles in Figure 4.4) generally trend to the NW of the true source (red star in Figure 4.4) in the spring and summer months and to the NE of the true source in the fall. As seen in Figure 4.5, the major axes of the 90% EEs trend from NW to SE. For all months other than July, temporal estimates produced utilizing the Blom et al. (2020) generalized celerity model have negative origin time residuals, indicating that origin time estimates are early (Figure 4.6). The length of the 90% confidence bounds decreases as the calendar year progresses, ranging from 100-400 sec, with
Figure 4.2: (a) Number of event detections at each station across the network. (b) Total number of detections/GT event across the network.
Figure 4.3: Number of detections, as a function of detection backazimuth measured from source to receiver. Radial bins mark detection number.
an average duration of 173.89 sec. Finally, only 30/45 of origin time estimates contain the true GT origin time within the 90% confidence bounds.

A second set of localization estimates are made using the monthly PGMs in order to test whether these more detailed celerity model can correct for biases within the localization estimates, such as the NW trend observed in MAP solutions and the NW-SE trend observed across 90% EEs. Localization utilizing the monthly PGMs produces results with an average mislocation of 21.49 km (a 50% increase), an average 90% EE area of 2126 km² (a 42% decrease in area) and an average time residual of 14.17 sec (an absolute reduction in error by a factor of 4.4; Table 4.2). We observe that 31/45 localization estimates retain the true GT location within the 90% EE bounds and 34/45 temporal estimates contain the true GT origin time within the 90% confidence bounds.

MAP location estimates using the monthly PGMs (blue squares in Figure 4.4) generally trend to the NE of the true source in the spring and fall months and to the S in the summer months. As seen in Figure 4.5, the major axes of the 90% EEs continue to trend from NW to SE. Temporal estimates produced with the monthly PGMs have positive origin time residuals, corresponding to late origin time estimates; however, in July, a number of estimates have negative residuals (Figure 4.6). Temporal 90% confidence bounds decrease relative to estimates produced with the generic celerity model, ranging from 0-300 sec with an average of 124.11 sec.

Two specific example events illustrate general changes to localization results when the monthly PGMs are applied. Figure 4.5a illustrates an example from 2012-06-04 where application of the June PGM moved the MAP location estimate further from the true source (increased overall mislocation in km). In this example, while the 90% error ellipse area decreased significantly, it no longer contains the known source location. This example il-
Table 4.1: Mean source localization estimates for 45 UTTR events. Mislocations are calculated as the distance between the true location and the MAP location estimate (in km). Origin time residuals are calculated as the true origin time – MAP origin time estimate. The 90% confidence bounds are computed from the marginal spatial and temporal distributions, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Blom et al. (2020)</th>
<th>Monthly PGMs</th>
<th>UTTR Empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mislocation [km]</td>
<td>14.34</td>
<td>21.49</td>
<td>15.49</td>
</tr>
<tr>
<td>EW Variance [km]</td>
<td>14.36</td>
<td>11.24</td>
<td>14.74</td>
</tr>
<tr>
<td>NS Variance [km]</td>
<td>16.48</td>
<td>12.83</td>
<td>17.99</td>
</tr>
<tr>
<td>GT location inside 90% EE [#]</td>
<td>44</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>90% Error Ellipse Area [km²]</td>
<td>3653.47</td>
<td>2126.82</td>
<td>4141.04</td>
</tr>
<tr>
<td>90% Temp. Conf. Dur. [sec]</td>
<td>173.89</td>
<td>124.11</td>
<td>177.70</td>
</tr>
<tr>
<td>GT OT inside 90% Conf. Bnd [#]</td>
<td>30</td>
<td>34</td>
<td>45</td>
</tr>
</tbody>
</table>

Table 4.2: Median source localization estimates for 45 UTTR events. Mislocations are calculated as the distance between the true location and the MAP location estimate (in km). Origin time residuals are calculated as the true origin time – MAP origin time estimate. The 90% confidence bounds are computed from the marginal spatial and temporal distributions, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Blom et al. (2020)</th>
<th>Monthly PGMs</th>
<th>UTTR Empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mislocation [km]</td>
<td>13.36</td>
<td>19.72</td>
<td>14.60</td>
</tr>
<tr>
<td>EW Variance [km]</td>
<td>13.92</td>
<td>10.96</td>
<td>14.80</td>
</tr>
<tr>
<td>NS Variance [km]</td>
<td>11.75</td>
<td>9.49</td>
<td>13.20</td>
</tr>
<tr>
<td>GT location inside 90% EE [#]</td>
<td>44</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>90% Error Ellipse Area [km²]</td>
<td>2493.12</td>
<td>1548.46</td>
<td>2846.39</td>
</tr>
<tr>
<td>Time residual [sec]</td>
<td>-65.01</td>
<td>29.11</td>
<td>-29.71</td>
</tr>
<tr>
<td>90% Temp. Conf. Duration [sec]</td>
<td>139.22</td>
<td>112.03</td>
<td>149.01</td>
</tr>
<tr>
<td>GT OT inside 90% Conf. Bnd [#]</td>
<td>30</td>
<td>34</td>
<td>45</td>
</tr>
</tbody>
</table>
Figure 4.4: Monthly localization comparisons utilizing the two distinct celerity models: Blom et al. (2020), black and monthly PGMs, dark blue. Dashed lines indicate the change in mislocation for an individual event. Filled markers indicate events where the true GT location is contained within the 90% EE and empty markers indicate events where the 90% does not contain the true GT location. The red star marks the UTTR source location. Bottom right panel compares event-to-event changes in mislocation and in 90% EE areas.
Figure 4.5: Spatial localization results for two UTTR (red star) events. (a) marks the first event (date labeled at top), and (b) marks the second, later event (date labeled at top). Results utilizing the Blom et al. (2020) generalized celerity model are shown in black while results utilizing the monthly PGMs are shown in dark blue. Ellipses denote the 90% confidence bounds and the marker denotes the maximum a posteriori solution. Blue triangles represent the arrays with an observable detection for the event and grey lines indicate the measured backazimuth of the detection.
lustrates the trade-off between location accuracy and precision when PGMs are utilized for location, where the 90% EE generally decreases, i.e. precision is improved, but no longer contains the true source, i.e. reduced accuracy (Arrowsmith et al., 2014; Nippress et al., 2014). Figure 4.5b illustrates an example from 07-30-2012 where application of the July PGM moves the MAP location estimate closer to the true source (reduced overall mislocation in km) and reduces the size of the 90% EE area.

This trade-off is further demonstrated in Figure 4.4, where individual MAP location estimates utilizing both the Blom et al. (2020) generalized model (black circles) and the monthly PGMs (dark blue squares) are plotted as a function of month. Application of the monthly PGMs reduces EW and NS variance, resulting in smaller 90% EE areas; however, location estimates for 14/45 events are no longer contained within the 90% EE. In addition, use of the monthly PGMs for location increases mislocation errors by an average of 10 km as compared to results produced with the generalized celerity model. There are clear spatial trends in the results using PGMs where location estimates move south of the true source during the summer months (July and August) and east of the true source during the transitional months (June, September, October; Figure 4.4).

Direct comparisons between localization estimates produced with the two models demonstrate that each give biased results, where spatial localizations utilizing the generalized celerity model are biased towards the NW of the source and the 90% EE axes trend towards the NW-SE. Temporal localization estimates utilizing the generalized celerity model are early, i.e., prior to the true origin time. Although the monthly PGMs were hypothesized to improve localization through an improved characterization of the atmosphere, they instead lead to larger errors in both spatial and temporal location across the GT dataset. Spatial localizations utilizing the generalized celerity model are biased towards the NE of the source while the 90% EEs continue to trend towards the NW-SE. In addition, while the 90% EE
areas decrease, a number of estimates no longer contain the true source within the bounds. Temporal localization estimates utilizing the PGMs are late, i.e., after the true origin time.

4.4.2. Interpreting Localization Errors

Signal characteristics from the network are used to assess the celerity models and backazimuth deviation characteristics applied within the celerity and backazimuth priors for location. Data and model comparisons provide a basis to interpret trends observed in the localization results.

The observed signal celerities are calculated as:

\[
\frac{\text{station distance from source (km)}}{\text{peak signal arrival time-origin time (sec)}} \quad (4.1)
\]

and observed backazimuth deviations are calculated as:

\[
\text{true backazimuth-peak signal backazimuth (calculated from ray-tracing)}, \quad (4.2)
\]

The Blom et al. (2020) generalized celerity model is derived from ray-tracing through seven years of G2S atmospheric specifications and represents all possible arrivals from 0-1000 km and all backazimuths. It is composed of primarily thermospheric arrivals, a significant stratospheric contribution and a small tropospheric component. In contrast, observations across the network are primarily stratospheric, with a large tropospheric component (Figure 4.7a). For location purposes, this comparison suggests that use of the Blom et al. (2020) generalized celerity model allows for slower phase velocities than are actually observed, resulting in artificially longer travel times that produce origin time estimates that are biased earlier than the GT source time. This phenomenon is also documented in Nip press et al. (2014) and is supported by the predominantly negative origin time residuals in Figure 4.7.
Discrepancies between predicted and observed celerities may also drive the bias observed in spatial localization estimates; the location algorithm over-corrects for stations with observations faster than what is predicted in the model by increasing the range to these stations in order to minimize the difference between the observed and predicted travel times.

Observed signal backazimuths are compared to the 4° standard backazimuth deviation utilized in the Blom et al. (2020) model in Figure 4.7b. These comparisons suggest that a slightly larger backazimuth standard deviation of 5.3° may be more appropriate for this dataset, as signal backazimuths deviate more than model predictions. The under-estimation of backazimuth deviation across the network may correspond to spatial localization errors; however, more investigation is needed to concretely support this suggestion since the differences are small.

Figure 4.8a compares observed signal celerities to predicted celerity distributions contained within the PGMs. Although monthly PGMs are used for event localization, for some month/array combinations, there are few, or in some cases no observations, therefore we combine months that are expected to exhibit similar propagation effects (e.g., spring, summer and fall) in order to compare seasonal predicted model results (celerity probability and backazimuth deviation) against seasonal observations. These comparisons are produced for each station (different distances and azimuths). Results from three stations (HWU, FSU and BRP) are used to illustrate the general comparisons (see Figure 4.9 and Figure 4.10 for full comparisons across the network). The seasonal trends in the PGMs are driven by reversals of the stratospheric jet as discussed by Blom et al. (2015). Observations, calculated from seasonal detections at each station, deviate significantly from these predictions. For example, arrivals at HWU are predicted to be primarily tropospheric in the spring, with some thermospheric contributions. Observations are predominately stratospheric in the spring and fall with tropospheric arrivals in the summer. Similarly, arrivals at BRP are predicted to be stratospheric and thermospheric, while observed arrivals are predominantly stratospheric in
Figure 4.6: Monthly origin time comparisons utilizing the two distinct celerity models: Blom et al. (2020), black circles and monthly PGMs, dark blue squares. Filled markers indicate events where the true GT origin time is contained within the 90% temporal confidence bounds and empty markers indicate events where the 90% temporal confidence bounds do not contain the true GT origin time.

Figure 4.7: Comparisons between detection observations (grey histogram) and model predictions for a: signal celerities (in km/sec) and b: backazimuth deviations (in degrees) for the Blom (2020) predictions (dashed black line) and the empirical produced regional UTTR model (dotted green line).
the spring and fall with thermospheric arrivals in the summer. These discrepancies between observed and predicted celerities contribute to the positive origin time residuals where origin time estimates are later than the true GT source time.

Seasonal predicted and observed backazimuth deviations across the network are compared in Figure 4.8b, where the dark blue vertical line indicates the median predicted backazimuth for each distance range and azimuthal bin and the shaded blue region indicates the extent of the predicted deviation. Observed backazimuth deviations at all stations are larger than the predicted deviations; this phenomenon may be driven by the lack of modeled cross-wind deflection for tropospheric propagation over flat ground. Although not analyzed, this discrepancy may contribute to errors in the spatial localization estimates, however a direct relationship between the two has not been developed here.

The discrepancies between the observations and the two atmospheric model-based predictions are likely due to limitations of both the propagation methodology and the atmospheric model used to produce specifications, Ground to Space (G2S) (Drob et al., 2003), combined with the network geometry. Models were produced by ray-tracing through a stratified atmosphere over flat ground out to 1000 km from the source (UTTR) and as discussed in Blom et al. (2015), are dominated by the variability of the stratospheric jet. Because UU/SMU stations are only located to the SE-SSE and NE from 84 to 400 km of the source, signal propagation to these stations may not be driven by seasonal changes in the stratospheric jet. A more in-depth discussion into the propagation dynamics of the network is found in Dannemann Dugick et al. (2020). It is well known that geometric ray tracing for stratospheric arrivals are often not adequately modeled within 200 km of a source (Nippress et al., 2014). Additionally, the actual tropospheric variability may occur on time scales less than the 6 hour resolution of the G2S specifications used in this study (Baldwin & Dunkerton, 2001; Garcés et al., 1998; Sorrells et al., 1997). Finally, propagation predictions may be influenced by the exclusion of terrain from modeling parameters as documented by Blom (2020).
Figure 4.8: Seasonal (rows) and station-level (columns) comparisons between detection observations (grey histogram) and model predictions for a: signal celerities (in km/sec) and b: backazimuth deviations (in deg) for the Blom et al., (2015) PGMs (dark blue line).

Figure 4.9: Seasonal (rows) and station-level (columns) comparisons between detection observations (grey histogram) and model predictions for signal celerities (in km/sec) for the Blom et al. (2015) PGMs (dark blue line).
Figure 4.10: Seasonal (rows) and station-level (columns) comparisons between detection observations (grey histogram) and model predictions for backazimuth deviations (in degrees) for the Blom (2015) PGMs (dark blue line).
which illustrates discrepancies in the timing of tropospheric arrivals between models with flat ground and realistic topography. As seen in Figure 4.1, stations within the UU/SMU are situated across variable elevations and topography. Therefore, both celerity models may need further modification to better capture local-to-regional propagation of tropospheric arrivals across the network. Including terrain effects would increase the predicted azimuth deviation significantly, likely producing predictions that are more in agreement with the observations.

4.4.3. Construction of an Empirically-Derived Celerity Model

Initial results indicate that use of the Blom et al. (2020) generalized celerity model for localization leads to a bias in origin time estimates, along with overly large spatial uncertainty bounds. Application of monthly propagation-derived predictive PGMs designed to better account for signal propagation across the network does not improve localization estimates, for reasons discussed above. Given the general agreement between the generalized celerity model and the observations (Figure 4.7), we elect to replace the generalized model with an empirical celerity model based on the observations in order to assess possible location improvements for events at UTTR observed at stations within the UU/SMU regional network. Due to the uneven temporal and spatial distributions of signal observations, we elect to produce a network-wide model. Further work focusing on station-specific celerity models might be pursued to investigate possible improvements in precision and accuracy of localization estimates.

Following procedures outlined for model construction in Blom et al. (2015), a UTTR empirical celerity model (dashed green line in Figure 4.7a) is computed by applying a kernel density estimate (KDE) to the celerity values in the data set and then performing a curve fit optimization to identify the three weights, means and variances for the GMM. The GMM parameters now best reproduce predicted celerity based on the scatter in the observed celerity values using the ground truth information. All arrivals across the seasons, distance ranges
and backazimuths are combined to define a single empirical estimate of arrival celerities for the source at UTTR, as observed from 84 to 458 km. A fixed standard deviation for azimuth as discussed earlier of 5.3° is used to approximate backazimuth deviation across the network (see dashed green line in Figure 4.7b).

Each of the 45 GT events are re-located using this empirical model. This new model produces an average mis-location of 15.49 km, an average 90% EE area of 4141 km² and an average time residual of -27.36 sec (Appendix B). Most significantly, all spatial and temporal estimates retain the true GT location or origin time within the 90% confidence bounds. Updated spatial and temporal localization results (green triangles) are compared to baseline results produced using the Blom et al. (2020) generalized model (black dots) in Figure 4.11. Origin time residuals are compared against the length of the 90% temporal error bounds in the top right of each monthly panel. Mis-location estimates as compared to the baseline results vary on an individual basis. For most events, spatial location estimates do not dramatically change (demonstrated by small mislocation changes in Figure 4.11). In May, location results trend south of the baseline location results, corresponding to a reduction in mislocation. In July, location results trend east of the baseline location results, corresponding to an increase in mislocation. Temporal trends are more consistent across the full dataset; application of the empirical model corresponds to a reduction in origin time residuals. The duration of the temporal 90% confidence bounds remains constant; however, all these estimates include the true GT origin time within the bounds. These results cumulatively indicates that the empirical model produces both a more accurate and more precise temporal localization estimate.

4.5. Discussion

Infrasonic localization methods that utilize three distinct celerity models and predictions of backazimuth deviation to characterize signal propagation across a regional infrasound network in Utah, USA are tested and compared using data from a series of GT surface
Figure 4.11: Monthly localization comparisons utilizing the Blom et al. (2020) generalized celerity model (black) and the empirical UTTR celerity model (dark green). Dashed lines indicate the difference in mislocation for an individual event. Top right panel indicates the OT residual and length of the 90% error bounds for each event (in km), where filled markers indicate events where the true GT origin time is contained within the 90% EE and empty markers indicate events where the 90% does not contain the true GT origin time.
explosions. Results are evaluated based on a combination of spatial and temporal metrics that include mislocation, size of 90% EE areas, whether or not the true GT source location in contained within the 90% EE areas, origin time residuals and whether or not the true GT origin time is contained within the 90% temporal confidence regions.

Source localization that utilizes the Blom et al. (2020) generalized celerity model and a new UTTR data-based empirical celerity model produce similar spatial results, with mislocation ranging from 0-30 km from the known source location and 90% EEs ranging from 1000-5000 km$^2$ with outliers (Figure 4.12d + e). Despite the discrepancies between predicted and observed signal celerities, the use of an empirical celerity model and an expanded back-azimuth standard deviation does not dramatically improve spatial localization results over the generalized model. These results differ from observations noted by Nippress et al. (2014), where application of a data-based celerity model for localizations of UTTR events utilizing the BISL methodology and a similar network of stations increased the precision of locations. The discrepancy between results of this study and those of Nippress et al. (2014) may be related to both the number and spatial distribution of stations utilized for location in each study. Example events detailed within Nippress et al. (2014) were detected at 10 stations, some of which were arrays of stations and some which were single stations. The stations were located closer to the source (within 200 km) and were more evenly distributed around the source. This denser, closer coverage indicates that network azimuthal distribution and signal detection capability significantly contribute towards localization accuracy, as well as the complementary nature of single station infrasound observations.

The use of the UTTR empirical model does significantly improve temporal localization estimates, measured as a decrease in origin time residuals (Figure 4.12f). This improvement is related to changes in the dominant phases contained within each celerity prior, where the Blom et al. (2020) generalized model predicts primarily thermospheric arrivals, which effectively underestimates the detection travel times, leading to origin time estimates that are
consistently earlier than the true origin time (black histogram in Figure 4.12f). In contrast, the monthly PGMs generally overestimate detection travel times, assigning tropospheric celerities to primarily stratospheric arrivals. This leads to origin time estimates that are consistently later than the true origin time (blue histogram in Figure 4.12f). The data-centric UTTR model accurately predicts primarily stratospheric arrivals, leading to reduced origin time residuals that are centered around zero seconds (green histogram in Figure 4.12f). Additionally, the tropospheric celerity contribution in the empirical model is centered at 0.340 km/sec, while the tropospheric celerity contribution in the Blom et al. (2020) model is centered at 0.327 km/sec, which likely impacts temporal estimates.

Localization utilizing the PGMs increases precision (90% EE areas decrease; Figure 4.12e) with the trade-off of reduced accuracy (EEs no longer contain the true GT source location). This phenomena is noted in Nippress et al. (2014), Marcillo et al. (2013) and Blom et al. (2015) and is related to the incorrect weighting of celerities within the monthly celerity PDFs. Unlike observations in Nippress et al. (2014) and Blom et al. (2015), the exclusion of the true GT source location within the 90% EE areas is limited to localization estimates produced only by the monthly PGMs.

Seasonal trends in spatial localization utilizing the PGMs illustrate that there are months when celerity estimates predicted from the PGMs differ by 0.1-0.2 km/sec from observations, suggesting that the models do not capture observed atmospheric dynamics across the network. These discrepancies may be related to both the limited availability of atmospheric specifications from models such as G2S (Drob et al., 2003), limitations in the geometric approximation utilized in ray-tracing (Blom & Waxler, 2012) where small-scale atmospheric variabilities are difficult to resolve and generally observed inconsistencies between predicted and observed infrasonic phases, particularly related to near-source stratospheric arrivals (Blixt et al., 2019; Nippress et al., 2014). Results of this study suggest that while application of monthly PGMs were shown to successfully improve event location estimates across
Figure 4.12: Summary statistics (a, d: mislocation in km; b, e: 90% EE area in km$^2$; c, f: origin time residual in sec) for localizations produced utilizing the Blom et al. (2020) generalized celerity model (black circles), UTTR empirical model (green triangles) and monthly PGMs.
IMS-scale networks, higher resolution models that account for tropospheric contributions and gravity waves may be needed to successfully apply the methodology to near-source regional distance networks.

The regional network used for this study is comprised of seven infrasound arrays located between 84-458 km from the source. Six of the seven arrays are located between 309-15° backazimuth of the source, with one array, HWU, offering additional azimuthal resolution. Detection capability varies across the network, therefore there are several GT events with reduced azimuthal coverage. Analysis of the error metrics as a function of detection azimuthal gap (Figure 4.12a-c) demonstrates that regardless of celerity model, spatial localization errors are driven by detection azimuthal gaps and thus must be taken into account in assessing location estimates. A similar relationship is noted in Blom et al. (2015). These comparisons document that spatial localization accuracy at near-to-source distances is driven more by detecting station distribution and the resulting detection backazimuth rather than signal travel times. The lack of improvements in spatial localization accuracy and precision utilizing the data-based empirical model may indicate that for non-optimally designed networks with sparse observations, predictions for expected backazimuth deviation are more important than celerity models. Backazimuth deviations observed across the network (grey histogram in Figure 4.7b) exceed predictions from ray-tracing and demonstrate that deviation in the Western US region may be higher than other regions with deployed infrasound sensors where azimuth residuals for arrivals from GT sources are quite small (Blixt et al., 2019; Smets et al., 2015). Increased backazimuth deviations may be related to elevation differences between array elements (Edwards & Green, 2012) or the presence of strong cross-winds in the lower atmosphere. Both of these factors could be reduced, following methods outlined in Koch (2020) with the goal of producing more accurate results.

4.6. Conclusions
A robust set of signal detections from 45 GT surface explosions at the Utah Test and Training Range (UTTR) are used to evaluate recent improvements to the Bayesian Infrasonic Source Localization (BISL) methodology. This work represents an expansion of previous studies that used a few select or optimal events for location, such as results documented in Nippress et al. (2014), Marcillo et al. (2013), and Blom et al. (2015). The extended analysis identifies an optimum operational procedure for use across a suite of events observed by a regional infrasound network with variable spatial and temporal signal detection. Signal arrivals are predominantly stratospheric and tropospheric and backazimuths deviate from -15 to 15° of the true backazimuth.

We successfully locate all 45 GT events, improving upon localization rates produced through automatic processing in Park et al. (2014). This significant improvement is related to several factors that may have contributed to the lack of UTTR events in the Park catalog. First Park et al. (2014) focused on bulletin production, thereby using automatic detection times and back-azimuths. Dannemann Dugick et al. (2020) demonstrate that across the UU/SMU network, an analyst review is necessary to identify many of the UTTR detections. The BISL methodology requires a minimum of three detections to produce a localization estimate, suggesting that a number of GT events were not included in the Park catalog because an insufficient number of arrivals were identified by the automatic detector. Similarly, Dannemann Dugick et al. (2020) illustrate differences between automatic and analyst arrival times on the order of 60-180 sec. These differences, particularly within an automatic processing pipeline, can lead to significant errors in both event association and location. Evaluation of the association methodology is beyond the scope of this study but may need further investigation. Finally, as noted in the introduction, the automatic processing pipeline in the previous study utilized the Modrak et al. (2010) BISL iteration with a singular velocity model, which may have introduced error into localization processing as signal celerities were not accurately predicted.
We apply three sets of celerity predictions to produce comparable localization estimates: (1) a generalized celerity model, derived from ray-tracing; (2) monthly PGMs for celerity and backazimuth based on station range and azimuth; and (3) a data-based empirical celerity model. Results indicate that the use of the monthly PGMs is inappropriate for localization utilizing stations at the SMU/UU network distance range and azimuthal distribution, as propagation characterized within the PGMs is optimized for stations further from the source. Application of the PGMs both underestimates and overestimates signal celerities, leading to large errors in both spatial and temporal location across the GT dataset. These errors correspond to perceived improvements in spatial accuracy, where the 90% EE areas decrease, coupled with reductions in spatial precision, although the 90% EEs do not contain the true GT location for 14/45 or 33% of the events in this dataset. Use of the generalized celerity model and the empirical UTTR celerity models produce spatial location estimates with similar accuracy, while use of the empirical model improves localization precision. We demonstrate that spatial localization accuracy is strongly dependent on the azimuthal gap in detecting stations. This result indicates that the BISL localization methodology weighs detection backazimuth observations more heavily than travel time estimates and further suggests that additional research into backazimuth corrections may offer the opportunity to improve spatial localization results. Use of the empirical celerity model reduces origin time residuals, providing the most accurate origin time estimates. This result suggests an updated procedure to improve source time estimates and indicates that station-specific celerity GMMs may further improve localization estimates.

Finally, we note that SMU/UU network utilized in this study leads to heavily biased location estimates, driven by the geometry of the network where six out of seven arrays are located to the SE-SSE of the source and only one station offers azimuthal resolution in the NE. Ideally, this study should be repeated utilizing a synthetic or real data network comprised of azimuthally distributed sensors with a repeating GT source to best evaluate location dynamics over time and space.
4.7. Author Contributions

FD led the analysis and writing of the manuscript under the supervision of BS. PB assisted with the interpretation of the Blom et al. 2015;2020 infrasound celerity models and development of the UTTR empirical celerity model. All authors provided critical feedback and helped shape the research, analysis and structure of the manuscript.

4.8. Acknowledgments

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 & 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.
Chapter 5
Conclusions

Within this dissertation, a network of regional infrasound arrays in the western United States is used to evaluate automatic processing pipelines through the successful detection, identification and location of surface explosions. These evaluations are motivated by high volumes of acoustic waveform data that are constantly recorded worldwide. The large data volume drives a need within the monitoring community to understand the applicability of algorithms and processing methodologies across regions of interest, as well as establish baseline performance metrics. By refining automated processing and assessing results, a broader understanding of network dynamics and atmospheric modeling limitations are established. The extended analysis of algorithms across a suite of events provides an enhanced basis and justification for operational use.

Chapter two establishes novel methodologies for producing direct comparisons between two automated signal detectors utilizing both real and synthetic signals. The first detector, the Adaptive F-Detector, accounts for coherent noise across an array 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. The second detector, the multivariate adaptive learning detector, applies three distinct time windows are used to adaptively alter detection thresholds in order to account for the changing background noise environment. Detection thresholds are evaluated based on the semblance across the array. Direct detector comparisons are needed in order to justify operational use as well as understand limitations associated with automated processing and evaluate associated analyst burden.
Both realistic and synthetic signals are used as a means of developing comparisons across ideal and not-ideal noise environments, utilizing a combination of short and long period signals. Automated detection catalogs are compared against analyst picks to evaluate detector performance in terms of precision and recall. Comparisons using real waveform data document similar precision and recall rates for the two detectors, indicating that performance across both detectors is nearly equal. Comparisons against the synthetic catalog indicate that both automated detectors identify long period signals with the same success rate, while successful identification of short period signals varies based on the detector methodology and background noise level. An implicit assumption in this assessment is that the analyst has superiority over the algorithm. Results demonstrate that multiple detector comparisons illuminate additional ‘true’ signals over the analyst, resulting in increases in detector precision and recall of between 10-50%. The hybrid approach applied within this work utilizing both an analyst and independent detection catalogs seems to enhance results particularly with regards to low SNR signals. Documented limitations of current infrasound detection algorithms include detection of short duration signals (< 30 sec), detection of signals in high noise environments (SNR <10) and identification and classification of repeating noise sources that originate from the same, or similar backazimuths. Additionally, without a standard set of parameters across signal detectors, it is difficult to produce direct comparisons between results. Comparisons using synthetic data highlight that using consistent p-values as thresholds for detection across multiple detectors does not produce comparable results. For example, in order to produce comparable detection catalogs with the synthetic data p-value thresholds were lowered for the multivariate adaptive learning detector. Future research efforts towards automated detection capability should focus both on parameters optimization as well as the highlighted limitations.

Chapter three utilizes a single detector from chapter two to demonstrate that across the UU/SMU regional infrasound network, an analyst review is necessary to identify a robust number of known event detections over the exclusive use of the automated Adaptive
F-Detector. The number of automatic detections varies as a function of array distance from the source where the arrays closest and furthest from the source detect very few events, while arrays between 140 – 300 km identify 60 to 80% of the GT events. Analyst review adds additional GT detections at all stations, and particularly increases the number of detections within the dataset at stations near the source of interest, increasing the percentage of successful detection to between 50-90% of GT events. Results illustrate that across this particular regional network, detection capability is neither driven by expected seasonal reversals within the atmosphere, nor by event yield. Detection capability is instead related to station proximity to the source, driven by the atmospheric propagation of tropospheric and thermospheric infrasound signals. Limitations in the power-based automated detector drive three significant differences between analyst and automatic detections. First, analyst detections are generally low in signal power (low SNR). Second, detections typically have low signal coherence across the array. Third, arrivals are short in duration of 10 sec. Results indicate that automatic and analyst arrival times vary on the order of 60-180 seconds, likely related to window length and parameter choices for automated processing, and detection algorithm triggers. These differences, particularly within an automatic processing pipeline, may lead to errors within eventual event association or location.

Chapter four represents a comprehensive assessment of infrasound event location algorithm applicability and measures of precision and accuracy over previous studies that used a few select or optimal events. Three variable celerity models are used to evaluate improvements to localization estimates through more accurate predictions of the changing propagation medium. The first, a singular celerity and backazimuth deviation model is constructed using ray tracing analysis based on an extensive archive of historical atmospheric specifications. The second model is a set of multi-variate, season and location specific models, or Path Geometry Models (PGMs) based on similar ray tracing analysis using historical atmospheric specifications. The third is an empirical celerity model based on signal arrival observations across the infrasound network. Application of the PGMs both underestimates
and overestimates signal celerities, leading to large errors in both spatial and temporal location across the GT dataset. These errors correspond to perceived improvements in spatial accuracy, where the 90% EE areas decrease, coupled with reductions in spatial precision, although the 90% EEs do not contain the true GT location for 14/45 or 33% of the events in this dataset. Use of the generalized celerity model and the empirical model improves both spatial localization precision and accuracy; all estimates retain the true GT location within the 90% confidence bounds while average mislocation is reduced to 15.49 km and average 90% error ellipse areas reduce to 4141 km$^2$. The empirical model additionally reduces origin time residuals; origin time residuals from the other location models were in excess of 160 seconds while residuals produced with the empirical model were within 20 seconds of the true origin time. This result suggests an updated procedure to improve source time estimates and indicates that station-specific celerity GMMs may further improve localization estimates. Results improve upon localization rates produced through automatic processing in prior studies and demonstrate that spatial localization accuracy is strongly dependent on the azimuthal gap in detecting stations. This result indicates that the Bayesian Infrasonic Source Localization methodology weighs detection backazimuth observations more heavily than travel time estimates, and further suggests that additional research into backazimuth corrections may offer the opportunity to improve spatial localization results.

As a whole, these three chapters advance the field of infrasound by providing an in-depth analysis of signal characteristics from surface explosions, summary statistics on detection and localization capabilities, and direct comparisons between predicted and observed signal dynamics across a regional network. Further refinement of the initial methodologies introduced in chapter two for direct comparisons will continue to promote quantitative assessments of algorithm performance. Chapter three demonstrates that a through analyst review is necessary for identifying and verifying a suite of signal detections; further detector developments should lead to high rates of successful automated detection as well as increased confidence in results. Results from the first two chapters combine to suggest that while ad-
vanced automated detectors often successfully identify events of interest within correlated noise, additional investigation is needed to understand background noise environments. The consistent mis-identification of coherent noise sources, as demonstrated in chapter two, indicates that methods to adaptively characterize background noise distributions should improve identification of transient signals of interest within changing noise environments by more accurately accounting and adapting for known noise sources.

Similarly, results within chapter four indicate that station-specific celerity probabilities may further improve localization estimates over network-based models by directly accounting for source-to-receiver propagation. Propagation interpretations are limited by the resolution of atmospheric models, leading to gaps in our understanding of signal propagation dynamics. Location results from chapter four demonstrate that sparse propagation predictions derived through modeling can be evaluated as well as supplemented by waveform data to provide a more-accurate understanding of spatial and temporal atmospheric propagation variability. The network utilized in this study offers poor azimuthal resolution where six out of seven arrays are located to the SE-SSE of the source of interest, and only one station is located towards the NE. The resulting signal detection and eventual location results demonstrates that network design and propagation dynamics may introduce bias.
## Appendix A

Ground Truth Events and Associated G2S Profile using for ray tracing

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4/9/12 11:59</td>
<td>17:59:23</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-04-09T18:00</td>
</tr>
<tr>
<td>4/9/12 13:33</td>
<td>19:32:52</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-04-09T19:00</td>
</tr>
<tr>
<td>4/16/12 11:34</td>
<td>17:36:47</td>
<td>3697</td>
<td>1676.93</td>
<td>2012-04-16T17:00</td>
</tr>
<tr>
<td>4/16/12 12:59</td>
<td>18:53:35</td>
<td>3697</td>
<td>1676.93</td>
<td>2012-04-16T18:00</td>
</tr>
<tr>
<td>4/23/12 11:57</td>
<td>17:55:58</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-04-23T17:00</td>
</tr>
<tr>
<td>4/23/12 13:10</td>
<td>19:09:46</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-04-23T19:00</td>
</tr>
<tr>
<td>5/14/12 11:55</td>
<td>17:56:34</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-05-14T17:00</td>
</tr>
<tr>
<td>5/14/12 13:25</td>
<td>19:25:43</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-05-14T19:00</td>
</tr>
<tr>
<td>5/21/12 11:51</td>
<td>17:50:57</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-05-21T17:00</td>
</tr>
<tr>
<td>5/21/12 14:56</td>
<td>20:55:28</td>
<td>17401</td>
<td>7892.96</td>
<td>2012-05-21T20:00</td>
</tr>
<tr>
<td>6/4/12 12:00</td>
<td>18:01:25</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-06-04T18:00</td>
</tr>
<tr>
<td>6/6/12 13:33</td>
<td>19:33:32</td>
<td>17401</td>
<td>7892.96</td>
<td>2012-06-06T19:00</td>
</tr>
<tr>
<td>6/6/12 14:42</td>
<td>20:43:18</td>
<td>17401</td>
<td>7892.96</td>
<td>2012-06-06T20:00</td>
</tr>
<tr>
<td>6/11/12 12:03</td>
<td>18:05:11</td>
<td>35000</td>
<td>15875.73</td>
<td>2012-06-11T18:00</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>6/13/12 14:16</td>
<td>20:15:17</td>
<td>38714</td>
<td>17560.38</td>
<td>2012-06-13T20:00</td>
</tr>
<tr>
<td>6/19/12 11:45</td>
<td>17:45:05</td>
<td>3746</td>
<td>1699.16</td>
<td>2012-06-19T17:00</td>
</tr>
<tr>
<td>6/25/12 11:41</td>
<td>17:44:35</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-06-25T17:00</td>
</tr>
<tr>
<td>6/25/12 13:34</td>
<td>19:36:03</td>
<td>38914</td>
<td>17651.09</td>
<td>2012-06-25T17:00</td>
</tr>
<tr>
<td>6/26/12 13:19</td>
<td>19:19:04</td>
<td>35000</td>
<td>15875.73</td>
<td>2012-06-26T18:00</td>
</tr>
<tr>
<td>6/26/12 14:50</td>
<td>20:50:13</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-06-27T00:00</td>
</tr>
<tr>
<td>7/9/12 11:58</td>
<td>17:58:20</td>
<td>35000</td>
<td>15875.73</td>
<td>2012-07-09T17:00</td>
</tr>
<tr>
<td>7/10/12 11:41</td>
<td>17:42:19</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-07-10T17:00</td>
</tr>
<tr>
<td>7/23/12 11:45</td>
<td>17:42:01</td>
<td>38714</td>
<td>17560.38</td>
<td>2012-07-23T17:00</td>
</tr>
<tr>
<td>7/23/12 13:18</td>
<td>19:17:47</td>
<td>35000</td>
<td>15875.73</td>
<td>2012-07-23T19:00</td>
</tr>
<tr>
<td>7/24/12 12:34</td>
<td>18:34:56</td>
<td>38714</td>
<td>17560.38</td>
<td>2012-07-24T18:00</td>
</tr>
<tr>
<td>7/30/12 13:00</td>
<td>19:05:00</td>
<td>37614</td>
<td>17061.42</td>
<td>2012-07-30T19:00</td>
</tr>
<tr>
<td>7/31/12 11:28</td>
<td>17:28:36</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-07-31T17:00</td>
</tr>
<tr>
<td>7/31/12 13:24</td>
<td>19:24:25</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-07-31T19:00</td>
</tr>
<tr>
<td>8/6/12 11:55</td>
<td>17:55:06</td>
<td>3697</td>
<td>1676.93</td>
<td>2012-08-06T12:00</td>
</tr>
<tr>
<td>8/6/12 13:36</td>
<td>19:36:26</td>
<td>3697</td>
<td>1676.93</td>
<td>2012-08-06T12:00</td>
</tr>
<tr>
<td>8/7/12 11:56</td>
<td>17:55:59</td>
<td>38914</td>
<td>17651.09</td>
<td>2012-08-07T00:00</td>
</tr>
<tr>
<td>8/13/12 13:43</td>
<td>19:43:00</td>
<td>38914</td>
<td>17651.09</td>
<td>2012-08-13T19:00</td>
</tr>
<tr>
<td>8/13/12 15:25</td>
<td>21:26:07</td>
<td>35000</td>
<td>15875.73</td>
<td>2012-08-13T21:00</td>
</tr>
</tbody>
</table>
Appendix A – continued from previous page

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8/14/12 13:43</td>
<td>19:44:13</td>
<td>38914</td>
<td>17651.09</td>
<td>2012-08-14T19:00</td>
</tr>
<tr>
<td>8/14/12 15:09</td>
<td>21:09:48</td>
<td>38914</td>
<td>17651.09</td>
<td>2012-08-14T21:00</td>
</tr>
<tr>
<td>8/15/12 12:16</td>
<td>18:14:18</td>
<td>3873</td>
<td>1756.76</td>
<td>2012-08-15T18:00</td>
</tr>
<tr>
<td>8/20/12 11:45</td>
<td>17:45:27</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-08-20T17:00</td>
</tr>
<tr>
<td>8/20/12 13:21</td>
<td>19:22:29</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-08-20T19:00</td>
</tr>
<tr>
<td>8/27/12 11:47</td>
<td>17:46:56</td>
<td>17500</td>
<td>7937.87</td>
<td>2012-08-28T00:00</td>
</tr>
<tr>
<td>8/27/12 13:22</td>
<td>19:22:29</td>
<td>38714</td>
<td>17560.38</td>
<td>2012-08-28T00:00</td>
</tr>
<tr>
<td>9/11/12 14:30</td>
<td>20:26:26</td>
<td>3671</td>
<td>1665.14</td>
<td>2012-09-11T20:00</td>
</tr>
<tr>
<td>9/26/12 11:46</td>
<td>17:45:48</td>
<td>6594</td>
<td>2990.99</td>
<td>2012-09-26T17:00</td>
</tr>
<tr>
<td>9/26/12 14:16</td>
<td>20:11:50</td>
<td>6594</td>
<td>2990.99</td>
<td>2012-09-26T20:00</td>
</tr>
<tr>
<td>10/3/12 10:54</td>
<td>16:53:57</td>
<td>6594</td>
<td>2990.99</td>
<td>2012-10-03T16:00</td>
</tr>
<tr>
<td>10/3/12 13:02</td>
<td>19:02:55</td>
<td>6594</td>
<td>2990.99</td>
<td>2012-10-03T19:00</td>
</tr>
</tbody>
</table>
Appendix B
Bayesian Infrasonic Source Localization Mathematical Framework

The adaptive BISL algorithm (Blom et al., 2015) combines likelihood equations for back-azimuth and travel time to assign a combined likelihood value for the location parameters. The posterior probability density function is defined from Bayes’ theorem,

\[ P(H|D) = \prod_j P(D_j|H) \int \prod_j P(D_j|H') (P_0(H')dH') P_0(H) \]  

(A1)

where \( H \) is the source hypothesis, \( D_j \) is a list of infrasound detections, \( dH' \) contains the variational elements of the hypothesis parameter space and \( P_0(H) \) is the prior estimate of the space containing all possible source characteristics. The likelihood function, \( P(D_j|H) \), quantifies the probability of the observed infrasound detection data, given that hypothesis \( H \) is true. The prior PDE incorporates a priori information while the likelihood function incorporates real time data. The final a posteriori function \( P(H|D) \) describes the probability that the source hypothesis \( H \) is true, given \( D \).

In the initial BISL iteration (Modrak et al., 2010), the likelihood function, \( P(D_j|H) \), was specified by the normal distributions for the azimuth and travel time residuals. Modifications begin with identifying the celerity of an infrasound detection based on the source-receiver separation and propagation time,

\[ v_j = \frac{\sqrt{(x_s - x_j)^2 + (y_s - y_j)^2}}{\tau_j - \tau_s} \]  

(A2)

where a signal propagating from a proposed source to the \( j^{th} \) detecting array can be computed from \( x_j, y_j, \tau_j, x_s, y_s, \) and \( \tau_s \). Next, the hypothesis parameter \( (H) \) is defined to only contain
the source location and time,

\[ H = x_s, y_s, \tau_s. \]  \hspace{1cm} (A3)

The detection \((D_j)\), is defined by the location of the detecting array \((x_j, y_j)\), the detection backazimuth \((\varphi_j)\), detection arrival time \((\tau_j)\) and a Fisher statistic of the detection \((F_j)\). The Fisher statistic is used to weight higher quality detections more heavily in the source location estimate. The likelihood function \(P(D_j|H)\) serves to replace the normal celerity distributions outlined by (Modrak et al., 2010) with appropriate distributions for azimuth and travel time. The function describes the probability of observing a detection \(D_j = x_j, y_j, \tau_j, \varphi_j, F_j\), under the assumption that the source location and time are specified by \(H = x_s, y_s, \tau_s\),

\[ P(D_j|H) = \frac{e^{\kappa \cos(\phi - \varphi_j)}}{2\pi I_0(\kappa)} \rho_v \left( \frac{r}{\tau_j - \tau_s} \right), \]  \hspace{1cm} (A4)

where \(I_0(\kappa)\) is the modified Bessel function of zeroth order, and \(\kappa\) is a von Mises width that is used to define the azimuthal dependence of the likelihood. Following the updated derivation in (Blom et al., 2020), the beam uncertainty component of \(\kappa\) is calculated as,

\[ \kappa = 2(N - 1)F \]  \hspace{1cm} (A5)

and is combined with \(\Delta\), the uncertainty due to propagation effects, to produce a final \(\kappa\) value for use in the likelihood:

\[ \kappa = \frac{\ln(2)}{(1 - \cos \left( \frac{\phi_1}{2} \right) + \Delta)} \]  \hspace{1cm} (A6)

where the beamed halfwidth is calculated from the \(F\) values as,

\[ \phi_\frac{1}{2} = \cos^{-1}(1 - \frac{\ln(2)}{2(N - 1)F}). \]  \hspace{1cm} (A7)
Finally, the azimuthal direction between detection station and hypothetical source, \( \phi \), and \( r \), the source to station range, are defined to be:

\[
\phi = \tan\left( \frac{y_s - y_j}{x_s - x_j} \right) \quad (A8a)
\]
\[
r = \sqrt{(x_s - x_j)^2 + (y_s - y_j)^2}. \quad (A8b)
\]

A posterior PDF for the source location and time can be computed by combining Eq. 1 and Eq. 4. However, this PDF does not account for the distribution of the signal celerity \( \rho_v \). Due to the uncertainties in infrasonic propagation, \( \rho_v \) cannot be referenced from a table. A distribution must be calculated that measures the probability of observing specific celerities for particular source-receiver geometries. In the case of localizations produced with the Blom et al. (2020) generalized celerity model and the UTTR empirical model, \( \rho_v \) consists of a Gaussian mixture model (GMM) defined in terms of inverse celerity, \( v^{-1} = \frac{\delta r}{\delta s} \). Table A1 illustrates the means, \( \mu_j \), variances, \( \sigma_j \), and weights, \( w_j \), for each respective GMM.

For localizations produced with the Blom et al. (2015) path geometry models (PGMs), the distribution of celerities in the likelihood function \( P(D_j|H) \) is estimated using propagation path distributions calculated from ray-tracing (Blom & Waxler, 2012) through 7 years (2007-2013) of G2S (Drob et al., 2003) atmospheric specifications. In this study, eight az-

**Table A1: Means, \( \mu_j \), variances, \( \sigma_j \), and weights, \( w_j \), for the two Gaussian mixture models (GMM) defining infrasound celerity.**

<table>
<thead>
<tr>
<th>Model</th>
<th>( j )</th>
<th>( \mu_j ) s km(^{-1})</th>
<th>( \sigma_j ) s km(^{-1})</th>
<th>( w_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blom et al. (2020)</td>
<td>1</td>
<td>1.0 / 0.327</td>
<td>0.066</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.0 / 0.293</td>
<td>0.080</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.0 / 0.260</td>
<td>0.330</td>
<td>0.856</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.0 / 0.430</td>
<td>0.07</td>
<td>0.15</td>
</tr>
<tr>
<td>UTTR Empirical Model</td>
<td>2</td>
<td>1.0 / 0.280</td>
<td>0.20</td>
<td>0.704</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.0 / 0.235</td>
<td>0.30</td>
<td>0.146</td>
</tr>
</tbody>
</table>
imuth bins are used, centered every 45° with the azimuthal bins each covering 60°. Using this technique, ρv estimates are extracted for each azimuthal bin, which introduces both range and azimuth dependence. Azimuth deviation priors are calculated from the mean and variance of the azimuth deviation for the propagation predictions within each azimuthal bin. Azimuthal deviation is defined as δϕ(r) = δtr − δpr where δtr is the true azimuth from the source to the receiver and δpr is the backazimuth of the predicted arrival. Spatially and seasonally specific models are constructed by defining range and azimuth varying weights, means, and standard deviations in the GMM as well as means (bias) and standard deviation in the arrival back azimuth. These models are constructed using 50 km wide bins spaced at 10 km intervals for the same suite of G2S atmospheric specifications utilized to produce the generalized celerity models.

Now, the likelihood distribution is defined as,

\[ P(D_j|H) = \frac{e^{\kappa \cos(\phi - \psi - \delta \phi)}}{2\pi I_0(\kappa)} \rho_v\left(\frac{r}{r_j - r_s}\right) \]  \hspace{1cm} (A9)

Where \( \kappa(\Delta, F) \) remains as defined in Eq. 6, and \( \phi \) and \( r \) remain as defined in Eq. 8b. In this approach, \( \delta \phi \) and \( \kappa \) vary with propagation range. The azimuthal difference in the exponential of the von Mises distribution is modified to include the azimuth deviation mean, \( \Delta \), and the width parameter is modified to include the increased width due to the variance in the backazimuth, both derived from azimuthal deviation priors produced above. Additionally, the celerity distribution, \( \rho_v \), azimuth bias, \( \delta \phi \), and the azimuth variance, \( \Delta \), are indexed by propagation azimuth to account for uncertainties within atmospheric propagation.


135


Averbuch, G., Assink, J. D., Smets, P. S., & Evers, L. G., 2018. Extracting low signal-to-noise ratio events with the hough transform from sparse array data, Geophysics, 83(3), WC43–WC51. 6, 14, 27


Blandford, R. R., 1982. Seismic event discrimination, Bulletin - Seismological Society of America, 72(6B), S69—S87. 4


Bondar, I., Mclaughlin, K., & Israelsson, H., 2008. Improved event location uncertainty estimates. 95


Denny, M. D. & Johnson, L. R., 1991. The explosion seismic source function: Models and scaling laws reviewed, pp. 1–24, American Geophysical Union (AGU). 77


Evers, L., 2008. The inaudible symphony: on the detection and source identification of atmospheric infrasound, Tech. rep. 66


140


Hayward, C., 2010. Infrasound from earthquakes: Signal characteristics and depth discrimination. 15


143


144


Webster, J. D., Blom, P. S., & Dannemann Dugick, F. K., 2021. Lanl-seismoacoustics/infrapy v0.3.4.  10, 13


