A SOLUTION TO THE PROBLEM OF INTERPRETING

VERY LONG TIME SERIES OF AMBIENT NOISE

AS MEASURED AT VERY SPARSE NETWORKS

by

Kathryn Teresa Decker

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Geophysics

Boise State University

August 2013

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Kathryn Teresa Decker

Dissertation Title: A Solution to the Problem of Interpreting Very Long Time Series of Ambient Noise As Measured At Very Sparse Networks

Date of Final Oral Examination: 01 February 2013

The following individuals read and discussed the dissertation submitted by student Kathryn Teresa Decker, and they evaluated her presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Paul Michaels, Ph.D.   Chair, Supervisory Committee

John Bradford, Ph.D.   Member, Supervisory Committee

Leming Qu, Ph.D.   Member, Supervisory Committee

Michael West, Ph.D.   External Examiner

The final reading approval of the dissertation was granted by Paul Michaels, Ph.D., Chair of the Supervisory Committee. The dissertation was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

For my family: both one I was given, and the one I found along the way

ACKNOWLEDGMENT

I am happy to acknowledge several individuals and organizations that contributed immensely to my success as a graduate student at Boise State University. Without the help of the community here at BSU and the funding provided by both internal and external sources, I never would have accomplished my goals.

I would like to thank my committee members, Dr. John Bradford and Dr. Leming Qu, who have been enthusiastic and reliable throughout my time at Boise State University. I am especially grateful for my former and current committee chairs: Dr. Matt Haney, currently with the USGS, and Dr. Paul Michaels. Dr. Haney was responsible for securing funding from ION Geophysical and the ARRA Grant for my research, and his enthusiasm for new projects has introduced me to the world of seismology. Dr. Michaels has been a remarkable source of reassurance and inspiration and I would like to thank him for countless hours and words of advice on programming, writing, and completing my degree. I would also like to express my appreciation for his enthusiasm for taking over the duties of committee chair in the middle of a project and his ability to direct my research in a way that has allowed for a timely completion of a work of which I can truly be proud.

I would also like to thank the institutions who provided my funding for the duration of my time at Boise State University: the Inland Northwest Research Alliance, for providing full scholarship for my first year, the Department of the Interior and US Geological Survey

ABSTRACT

Ambient seismic noise techniques are an excellent choice for imaging the subsurface in areas that are seismically quiet or otherwise unsuitable for active source experiments due to geographic isolation or environmental sensitivity. Recently, decades-long time series were made available for download from the Incorporated Research Institutions for Seismology (IRIS) from permanent network installations, allowing access to long, uninterrupted recordings from seismometers around the world. This has spurred the development of an entire field of applications for passive seismic noise analysis. Over the continental United States, the USArray project has advanced to provide station coverage in relatively dense and regularly spaced arrays, but along the Aleutian Island arc in Alaska and other geographically isolated but seismically active locations, the hazards associated with volcanic eruptions and the difficulty of accessing stations for repair or replacement throughout most of the year has allowed only for sparse coverage.

The analysis of ambient seismic recordings generally suits one of two purposes. The first involves the parameterization of the source of each component of the ambient seismic noise spectrum and focuses on both the spatial locations and mechanisms of generation. The second purpose of looking at ambient seismic noise is to create velocity models of the subsurface below the array. The assumptions required for the traditional approach to analysis of ambient seismic noise, namely beamforming and the spatial autocorrelation

coefficient method, involve specific requirements for array geometry and density that often are not met when only limited permanent network coverage is available. Furthermore, most studies focused on mapping the subsurface generally assume that the source of ambient noise is diffuse and azimuthally homogeneous. Here, principal component analysis (PCA) is used to show that the source of ambient noise is often highly directive. Methods of incorporating the direction of approach are introduced to allow one to correct the apparent velocity calculated by the consideration of an omnidirectional source when the energy is instead strongly directive. To ensure that the energy measured at a remote network is indeed pervasive energy that travels through the ground rather than local noise, a quality control algorithm based on the results of PCA is also introduced. The benefits of improving the reliability of velocity measurements for the subsurface and reducing the size of the dataset by the introduction of the quality control parameter will greatly enhance the accuracy and ease with which scientists may determine subsurface velocity profiles at especially sparse networks with particularly long recording times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Data Accessibility and New Challenges

Over the past 40 years, the field of seismology has changed dramatically due to the introduction of digital recording technologies and incredible increases in the memory storage capabilities of personal computers. The science is now on the verge of a second revolution: data storage is no longer limited by the finite memory of one's personal computer or external storage device, but instead can be stored in the computing cloud and accessed almost instantly through the internet. The primary publically accessible source of seismic data is a massive online database of recordings from seismometers installed at permanent or semipermanent arrays around the world, organized and maintained by the Incorporated Research Institutions for Seismology (IRIS). In addition to the historic data holdings, IRIS also collects data in real-time from stations in networks established in particularly seismically active regions.

The International Association for Seismology and Physics of the Earth's Interior's (IASPEI) adoption of an internationally recognized format for digital seismic data in 1987 was a crucial element in simplifying the exchange of data (Incorporated Research Institu-

tions for Seismology, 2010). The continually increasing volume of continually recorded and archived data through the IRIS SeismiQuery system with free access through the IRIS breq_fast request utility has spurred the further development of both active and passive seismic data analysis and processing of data from very remote locations around the world. The SEED format is used for unfiltered data that is sampled at regular intervals. It was designed so that a user may request only one channel and receive a file with all of the necessary information for processing: the header includes information on station location, sampling interval, the transfer function as described by a list of the poles and zeros for the instrument used, the type of signal recorded (velocity or ground motion), and the date and time of the start and endpoint in the request.

The length of available time series is continually increasing for each permanent installation since many of these seismic stations are linked to the IRIS database and add new data every few minutes, and the number of networks where recordings are made has increased rapidly as well (Figure 1.1).

Despite all of the benefits of having an almost inexhaustible supply of data, most users reach RAM limits when attempting to process very long time series. Historically, the challenge of having years or sometimes decades of recordings has been addressed through time stacking and resampling to reduce the number of data points. In this dissertation, a new way to select smaller subsets of data by implementing a quality control algorithm to isolate data with specific spatial characteristics is introduced as an alternative to a blind approach to reducing data set size.

Although coverage with permanent seismic network stations has improved and continues to increase each year, there remain portions of the world where the installation of expensive broadband seismometers is impractical or impossible. Hazardous environments,

IRIS DMC Archive

as of 6/1/12
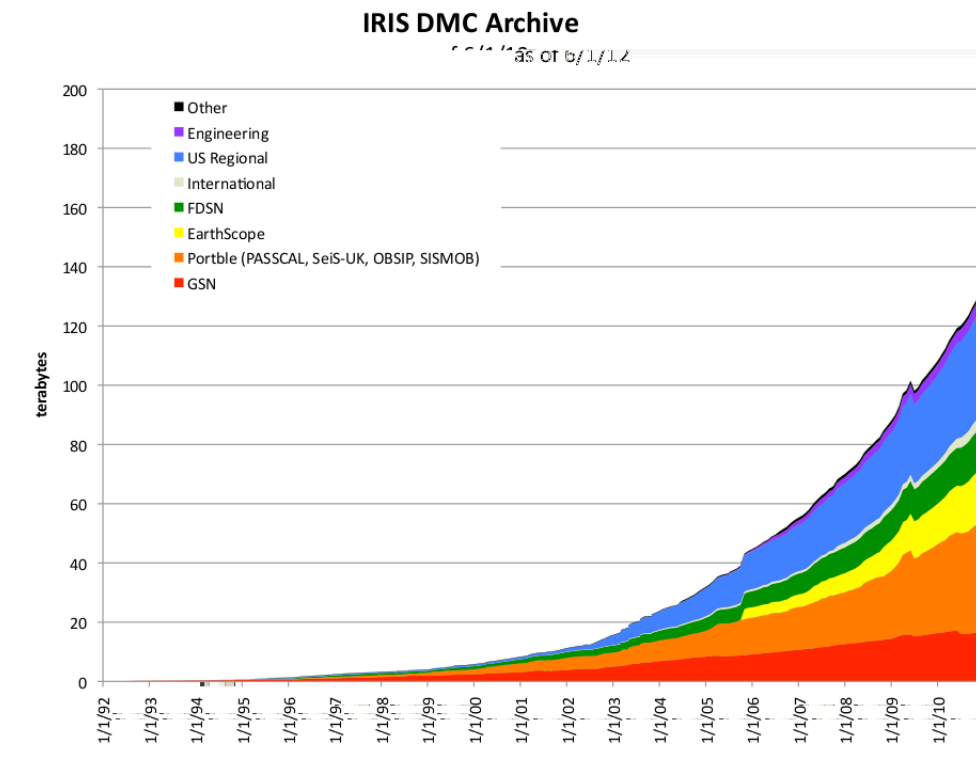
Figure 1.1: The IRIS database has expanded exponentially over time as assorted projects involving long- and short-term installations of seismometers are completed or underway.

such as sites in close proximity to or upon active volcanoes, are excellent and interesting candidates for seismic observation, yet the safety of the instruments cannot be guaranteed. In other areas, extreme weather or geographic isolation can prevent easy year-round site access. In these areas, seismic arrays generally have very few instruments, although each instrument may have very long recording times. The new methods presented in this dissertation provide geophysicists with strategies for working with data collected over sparse networks, where lack of control over survey geometry precludes the use of traditional methods of data analysis.

1.2   Passive Seismic Methods

One of the major benefits of having a network of new permanent stations designed to record continually is the new possibilities for analyzing very long time series of data recorded at times both with and without earthquake activity. Although the modern methods used in the study of seismic surface waves from earthquake events for investigating subsurface structure date back to the late 1950s and early 1960s (see, for example, Press *et al.* (1956), Press (1956), Press (1957), Alexander (1963), Ewing *et al.* (1959), and others), the majority of the applications for using ambient seismic surface waves to produce velocity profiles of the subsurface came decades later. Primarily, the limiting factor for the earlier studies was computer memory. In Landisman *et al.* (1969), for instance, the research team reports that the earliest attempts at digital experiments for moving window analysis of surface waves in 1958 required a full 60 hours to analyze data from just one hour of recordings with a sampling rate of 1 Hz. The team considered the evolution of what was then modern equipment as a major advance because the same analyses could be performed in "only a few minutes." Today, performing sliding window calculations for processing hour-long

time series sampled at 1 Hz takes only seconds, making practical the analysis of very long time series of ambient noise recorded by permanent network stations.

The use of ambient seismic noise presents several advantages in areas that are geographically isolated or delicate due to the presence of infrastructure, human activity, or wildlife. First, although scientists have used the signals generated by earthquakes to study the subsurface structure in these areas for decades, the analysis of ambient noise is perfect for determining structure in areas that are seismically quiet because no sources are needed. Second, ambient noise is pervasive. Since the majority of the permanently installed seismic networks were established for detection of earthquakes and eruptions, yet record continuously so as not to miss an event, the analysis of the surface waves generated by ambient sources is easy because the volume of data when events do not occur is magnificent and continually increasing. This allows one to study very long time series without the requirement of isolating the portions of the record with an earthquake source. Third, the method of processing ambient seismic noise relies primarily on the technique of cross correlation, which allows one to measure the similarity of a signal received at any two seismic stations (Bendat and Piersol, 1971). Thus, for cases with permanent arrays of seismometers, any one station can be treated as a source so the deleterious effects of distance on resolution that arise when analyzing teleseismic earthquake signals is avoided (Snieder and Wapenaar, 2010). The study of ambient seismic noise is often referred to in the literature as the microtremor method (Okada, 2003), Green's function retrieval (Snieder and Wapenaar, 2010), or seismic interferometry (Schuster, 2009). Any one of these terms can be used to search for more information on the subject.

1.3 Motivation

In general, ambient seismic noise studies treat the noise as a stationary stochastic process. That is, the spectral characteristics of ambient seismic noise are assumed not to vary in time or space. However, the characteristics of ocean surf are known to change throughout the year with variations in seasonal weather patterns. On a smaller time scale, variations also occur in signal strength over the course of the tidal cycle on a daily basis. As one examines the surf over a period of minutes, other variations may be noted due to ship traffic and sea life. The literature dedicated to studying spatial and temporal changes in the characteristics of ambient seismic noise is vast, yet studies that focus on using the ambient seismic noise records for surface wave analysis generally do not accommodate ideas of signals that may change in frequency content or direction. This study seeks to investigate the possibility that the stationary, stochastic assumption is invalid on time scales commonly used in surface wave analysis. A reconceptualization of ambient seismic noise as a signal that is often highly directive rather than pervasive and omnidirectional, allowing for the potential consideration of a "source" of the noise that varies based on the frequency of observation, is presented. Following this departure from the more traditional assumption of omnidirectional energy, the first contribution of this dissertation is the development of a new, simple way to identify the direction of arriving noise from an ambient seismic source that requires no specific station array geometry and a minimum of only three stations. Secondly, the effects of avoiding the common assumptions in favor of conceptualizing a source as a directional signal that may vary with frequency and in time are condsidered. An application of the new method, which allows the observer to determine the dominant angle of approaching energy, is presented. Information on angular approach is then used to correct the time delay for a wave passing from one station to another, ensuring an accurate

calculation of velocity.

After developing a way to identify the signal direction, the presentation continues with an outline of the second contribution of this dissertation: an algorithm to quickly isolate portions of time series when ambient noise is highly directional in order to determine a correction factor based on the angle of approaching energy that should be applied before surface wave dispersion analysis continues. The advantage of using the algorithm is twofold. First, it allows the user to isolate portions of time in very long records where the signal is highly directional. By use of the plane wave assumption, one may then choose to focus on these highly directional windows to analyze very distant station pairs as the effective distance "seen" by the seismic waves may be significantly shorter, allowing analysis of the higher frequencies in the signal. Secondly, the length of the time series available at permanent network stations generally requires one to resample at larger time intervals, reducing the signal bandwidth and the Nyquist frequency. By working with subsets of the recordings that occur with known direction, one may be fully confident that the travel time between station pairs is based on the signal characteristics. With the rapid growth of the IRIS database, the preference for working with smaller sets of data is quickly becoming an imperative. With the increase in interest in ambient seismic noise methods, the importance of working with signals where changing spatial and temporal characteristics are accommodated rather than neglected is worthy of attention.

Of particular interest to organizations like AVO, where the focus is on monitoring active volcanos and predicting eruptions, is the ease with which the new methods and procedures presented in this dissertation can be automated. The availability of continually updated data with IRIS suggests that an almost real-time monitoring system can and should be built at some point in the near future. The extension of this research to the production of

continually updating inversions of phase velocity curves is not covered in this dissertation, but this application is an area of interest for future research.

1.4 The Island of Akutan

Prior to the 1964 M_w = 9.2 earthquake near Prince William Sound, the state of Alaska had only two operational seismographs. As a result of the earthquake, the University of Alaska and the U.S. Coast and Geodetic Survey installed almost 40 additional stations from 1966 to 1972 (Plafker and Berg, 1994). Since then, the Alaska Volcano Observatory (AVO) office of the United States Geological Survey (USGS), the West Coast and Alaska Tsunami Warning Center (WCATWC), and the University of Alaska have established and maintained hundreds of additional seismic stations. Many of the stations are organized in networks with capabilities for real-time and archival storage of digital data with IRIS, all organized in the international system of Standard for the Exchange of Earthquake Data (SEED) format.

The island of Akutan is located in the eastern portion of the Aleutians. It is small and can be accessed only by boat. It is located to the west of the larger and more populated Unalaska Island (home of the king crab fishery and filming location for Discovery Channel's Deadliest Catch documentary) and east of the currently uninhabited Akun Island, which forms the eastern margin of Unimak Pass. Both the Northern Sea Route and the polar Great Circle Route for freight shipping go through Unimak Pass. In general, the pass is densely traveled, even for local shipping in and out of the port of Unalaska (Morton, 2010). The major geological feature on the island is Akutan Volcano, in the northwest portion of the island. The city of Akutan is the only populated place on the island and is located on the easternmost point, where effects of volcanic activity are historically limited to ashfall

from the larger eruptions. Since a larger, caldera-forming eruption is unlikely in the near future, the primary hazard for Akutan is the emission of ash in eruption plumes that can affect commercial and freight flight patterns over the Aleutians. Monitoring the velocity structure of the subsurface may provide scientists with clues about impending eruptions as magma moves below the surface of the islands and could lead to better eruption predictions and therefore earlier warnings for residents and air traffic controllers.

Akutan is one of the most active volcanoes in the Aleutians, despite its small-scale eruptions, and is therefore heavily monitored with seismometers. The majority of these instruments are (by IRIS classification) extremely short period seismometers, with corner periods of less than ten seconds. There are also several permanent broadband seismometer installations on the island and, because the passive seismic energy of interest for this dissertation comes from the ocean surf at frequencies below 1 Hz, recordings from only the broadband instruments are used. It was further found that there were a total of four instruments with long, overlapping recordings located at opposite ends of the island. Arrays of this size are likely to become more common in isolated, hazardous areas, and the methods developed in this dissertation focus on how to best use the information from these arrays since the traditional techniques of beamforming and the spatial autocorrelation methods fail for irregularly spaced arrays with very few components.

1.5   Organization of Dissertation

Chapter 2 of the dissertation introduces the reader to the concept of a source for passive seismic noise by highlighting discoveries of specific features in the spectra of ambient noise recordings and exploring their origins. The array geometry of the broadband stations on Akutan is discussed in greater detail and a brief tutorial on data access for each station is

also included.

In Chapter 3, an overview of the established practices for ambient seismic noise analysis is given and the ways in which the methods used in this dissertation differ from the current methods are reviewed. Special attention is paid to examining the assumptions that must be made to use the traditional statistical methods — the spatial autocorrelation coefficient (SPAC) and the frequency-wavenumber (f-k) method. Care is taken to make the case for avoiding the constraints of these assumptions in the methods developed in the next chapter.

Chapter 4 develops the methods and further describes the motivation for this dissertation by the development of specific derivations and examples from data sets recorded by broadband seismometers on Akutan.

Chapter 5 presents the results of the methods described in Chapter 4, and Chapter 6 finishes with a discussion of the results.

Chapter 7 contains the author's concluding remarks.

The instructions for accessing the specific dataset used in this dissertation and a general set of steps for using the IRIS database search tools are provided in Appendix A. Appendix B contains the annotated code used to generate figures and results for this dissertation. Appendix C contains a comprehensive review of the local and regional geology, while Appendix D provides explicit (although still abbreviated) derivations of two of the classic techniques for ambient seismic noise analysis to supplement the overview of the methods in Chapter 3.

Chapter 2

SOURCE PARAMETERS AND ARRAY GEOMETRY

2.1   Introduction

An extensive body of research exists on analysis of patterns and occurrences of microseisms produced by magmatic systems beneath the surface at volcanoes (e.g., Lauro *et al.* (2005), Arciniega-Ceballos *et al.* (2003), Tolstoy *et al.* (2002), and many others). Sources of the microseisms produced at volcanoes typically include vibrations of volcanic conduits and fracturing in the subsurface from the movement of magma and gas. Although the data set for this dissertation comes from Akutan, a volcano in the Aleutian Island Arc (a full report on the local and regional geology of the area is provided in Appendix C), the research presented here concerns microseisms generated by the interaction of ocean waves with the seafloor. The goal is to produce a new and nonstatistical method of analyzing the source of ocean-origin microseisms that may be used at any remote location with severely limited permanent seismic network coverage and the methods developed in this dissertation are not limited to volcanic environments. The data set from Akutan was chosen with the early goal of the project in mind — to create an early-warning system based on real-time, continually updated images of the velocity structure of the subsurface — but in the attempt to achieve

this goal, it was soon found that there were significant challenges to working with sparse but very long recordings that must be addressed before subsurface imaging can begin. As these challenges will become commonplace in the future when more and more stations in IRIS arrays in inaccessible or highly hazardous regions of the world come online, it was crucial to develop this dissertation to provide a way forward for researchers interested in working in areas with limited station coverage.

Studies on ambient seismic noise from ocean waves generally fall into one of two categories. The first focuses on determining the origins of different components of the noise. This generally allows one to consider a specific source for the ambient seismic noise. Sources include waves that form in particularly stormy parts of the world as well as more local interactions between waves and the seafloor. The idea of a directive source and a particular origin is explored and analysis techniques of beamforming (see Chapter 3 and Appendix D for description and derivation) are often applied. The second category assumes that the source of ambient seismic noise is diffuse and homogeneous. Studies where subsurface velocity models are produced from ambient recordings generally fall into the second category. The research for this dissertation was conducted with the hope of exploring the area where the two categories overlap: by disallowing the assumption of diffuse and uniformly distributed noise, the spatial and temporal characteristics of the signal are first observed and then actually used to define a quality control parameter and construct corrected velocity dispersion curves for the subsurface. The procedure requires an array consisting of a minimum of two stations, far fewer than the number of array elements required for beamforming.

2.2   Sources of Noise

The frequency bandwidth for ambient seismic noise may be divided into frequencies above 1 Hz, caused by pedestrian or machine traffic on land (Okada, 2003) or, at sea, marine life, ship traffic, raindrops, and bubbles (Urick, 1984). Noise from oceanic and atmospheric waves occurs at frequencies below 1 Hz (Okada, 2003). The work presented here will focus on the natural sources of noise below 1 Hz.

The current body of research on source parameters for microtremors has identified several components of the ambient noise spectrum that show up in recordings taken at stations around the world. Each source is briefly described below, starting with the lowest observed frequencies in the ambient noise bandwidth and ending with the highest. Following this, an amplitude spectrum plot of a time series recorded at one of the permanent seismic network stations on Akutan is examined for evidence of each commonly observed peak of the ambient spectrum. The time series used for the following examples was recorded with a Streckeisen STS-2 seismometer. The STS-2 instrument is commonly used to measure extremely low frequency signals as the velocity response corners are at 0.00833 Hz, or a period of approximately 120 s, and >50 Hz (Rhie and Romanowicz, 2006; Incorporated Research Institutions for Seismology, n.d.).

2.2.1   Infragravity: 0.005-0.02 Hz

The ambient noise in the infragravity bandwidth arises due to changes in atmospheric pressure with the passage of storms. Analysis of data from permanent network stations revealed that noise in this frequency band was pervasive and, since the changes in air pressure due to weather systems apparently excites the Earth's spheroidal elastic normal modes, this phenomenon is often called the Earth's "hum." It has been reported that the average back-

ground noise in this range of frequencies is equivalent to a moment magnitude of M_w=5.75, although the level of excitation has a seasonal dependence with slight but distinct maxima during January and July (Ekstrom, 2001). The maxima correspond to the winters in the northern and southern hemispheres.

Interestingly, the source of the ambient seismic noise in the infragravity bandwidth is still debated in the literature. Earlier studies reported that the source was atmospheric forcing (Kobayashi and Nishida, 1998, for example), whereas later studies favor generation by an interaction of changing atmospheric pressure with the ocean surface, which generates short period waves that then interact in a nonlinear fashion to generate the long-period infragravity waves that couple with the seafloor to produce the Rayleigh waves observed at seismic stations (Rhie and Romanowicz, 2006; Webb, 2008).

The 26-second Microseism: 0.038 Hz

A peak in the microseism spectrum at a period of 26 s (approximately 0.038 Hz) is found in ambient seismic recordings taken at stations in the United States, Africa, and Europe. When the source of ambient seismic noise is not homogeneously distributed around an array of seismometers, asymmetries arise in the cross correlations between station pairs. By observing the asymmetry in cross correlations of signals from stations along the same great circle paths and inverting the apparent travel times that were measured from these cross correlations, it was found that the source of the 26 s noise is located in the Gulf of Guinea, off the west coast of Africa (Shapiro *et al.*, 2006). The amplitude of the 26 s microseism varies with time of year and is at its maximum during the stormy winter months in the southern hemisphere (Holcomb, 1980).

A signal with a slightly broader peak has been recorded at the station in East Asia

and the western Pacific, as well. Using the same methods of analysis to locate the source, Shapiro *et al.* (2006) determined that this signal originates in the North Fiji Basin. Although this signal is also located at a specific point, there are no seasonal variations in signal amplitude (Holcomb, 1998). It is difficult to say whether or not the signal coming from the North Fiji Basin is actually the antipodal expression of the source from the Gulf of Guinea, or whether it is of different origin (Shapiro *et al.*, 2006).

2.2.2   Microseisms: 0.04-1.0 Hz

Ambient noise in the microseismic bandwidth has been found in all seismic recordings, starting from the earliest days of seismology. As with the efforts to define the source of infragravity waves, determining the source of the signal has been of interest to researchers ever since. Microseisms consist of mostly surface waves with some short-period *P* waves. The surface waves tend to be Rayleigh waves, although Love waves may arise in simply layered structure without lateral heterogeneity (Cessaro, 1994). The generation of micro-seisms is thought to arise from the interaction of the ocean waves with the seafloor, but the relationship between storm systems that generate ocean swell and variation in microseism amplitude and source location with time is complicated and poorly understood (Cessaro, 1994; Tanimoto, 2007; Rhie and Romanowicz, 2006). Within the microseismic bandwidth, which is generally reported as being over the frequencies from 0.04-1.0 Hz for natural sources, there are a few distinct features that one may observe in the spectrum of record-ings taken at any station (Haubrich *et al.*, 1963; Holcomb, 1980; Webb, 2008, and many more).

The Single Frequency: 0.06-0.07 Hz

The single frequency (SF), also known as the primary microseism or primary multiple, arises from the direct interactions of the surf with the ocean bottom in shallow water (Brooks *et al.*, 2009). The frequency of the primary microseism is the same as the frequency of ocean swell. The SF is generated by the pressure of the ocean waves directly on the sea floor. Overall, the amplitude of the SF microseism is generally lower than that of double frequency microseisms and, since it arises only in shallow water, the SF originates near the shore.

The Double Frequency: 0.12-0.15 Hz

Also called the secondary microseism or secondary multiple, the double frequency (DF) microseism is generated by the same nonlinear second-order pressure changes caused by the interaction of short-period ocean waves of similar wavenumbers traveling in opposite directions with the sea floor that generates the Earth's hum (Longuet-Higgins, 1950; Tanimoto, 2007), although the DF microseism is both a near-shore and deep-water effect and is directly correlated to wind activity (Haubrich *et al.*, 1963; Cessaro, 1994). The terms "double" and "multiple" are used because the frequency of the secondary microseism is twice that of the ocean waves (Brooks *et al.*, 2009). DF microseisms tend to have a larger amplitude than the SF microseisms. The DF microseism as measured in the northern hemisphere is at its minimum during summer months, when weather is generally calm, and has been observed to have much higher amplitude fluctuations with seasons in comparison to the hum signal (Rhie and Romanowicz, 2006).
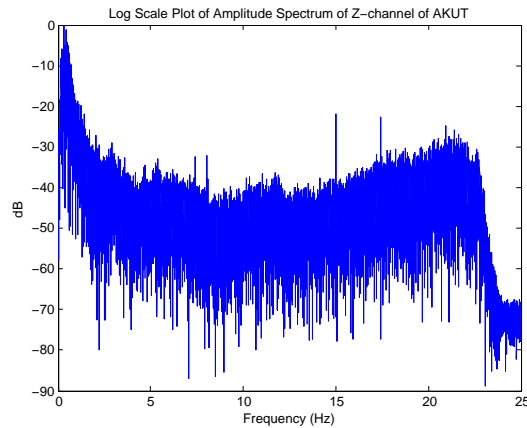
Figure 2.1: The full spectrum for a twenty-minute long signal recorded at station AKUT on an Streckeisen STS-2 broadband seismometer with sampling interval of 0.02 s. Note that the majority of the signal amplitude occurs for frequencies below 2 Hz. This recording was obtained during a seismically quiet time period and therefore includes energy from only the ambient seismic noise sources.

2.3   Spectrum of Ambient Recordings at Akutan

To assess the frequency content of an unfiltered signal recorded at station AKUT, the signal was Fourier transformed and the amplitude spectrum was plotted. The entire spectrum, from DC to the Nyquist frequency, is plotted in Figure 2.1.

Zooming in on the infragravity frequencies mentioned above, it is found that there is spectral amplitude for frequencies that are very close to DC. This is in keeping with the expectations for the frequency spectrum of Earth's hum (Figure 2.2).

In Figure 2.3, the 26 s microseism is observed. Note that this particular time series was recorded on August 4, 2008, and therefore one would expect to see that the signal has a relatively high amplitude when compared to recordings at the same station taken during the northern hemisphere winter.
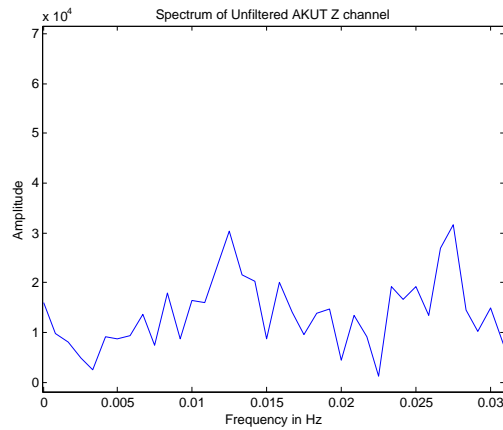
Figure 2.2: The frequencies from DC to approximately 0.02 Hz are pictured here. Spectral amplitude is found even as one moves towards zero. The very low frequency components of the amplitude spectrum arise from the excitation of lower frequency normal modes by ambient seismic sources.
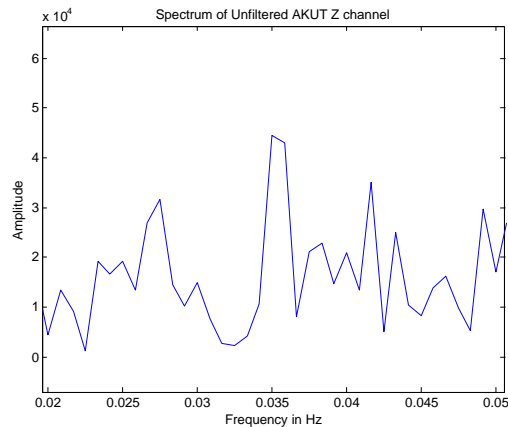
Figure 2.3: A spectral peak is evident at approximately 0.036 Hz (period of approximately 27.8 s) in the spectral amplitude. This is in keeping with what one would expect for a recording during the northern hemisphere summer as signal amplitude for the 26 s microseism is highest during the winter months in the southern hemisphere.
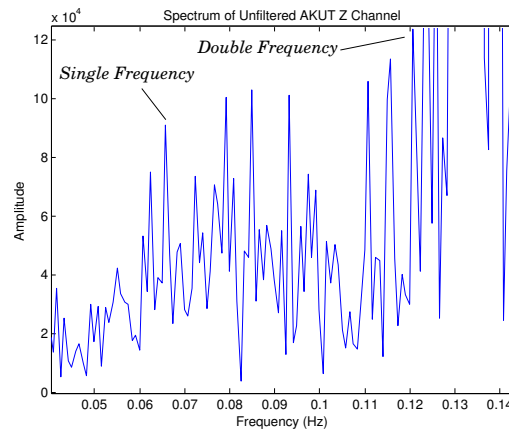
Figure 2.4: The single and double frequency microseisms are evident in the plot of the spectrum for the recording at AKUT at approximately 0.06 Hz and 0.12 Hz, respectively. The SF amplitude is less than that of the DF. The lack of a particularly evident individual peak for the two features is in agreement with the relatively broad power spectral density peaks published elsewhere (see, for example Lee *et al.*, 2011).

Finally, Figure 2.4 shows the primary and secondary multiples for the recording at station AKUT. The amplitude of the SF is lower than that of the DF, as expected.

To give an idea of the relative amplitude of each of the spectral peaks in the microseism bandwidth mentioned above, Figure 2.5 shows a closeup of the portion of the bandwidth that covers the frequencies from the lower range of the Earth's hum through that of the DF microseism.

The remaining portion of the spectrum above the DF is undifferentiated strong signal. In this research, the focus is on frequencies within the portion of the spectrum that has the largest amplitude, both for each individual recording (as in Figure 2.1), and for the frequency content found in the cross correlations of station pairs.
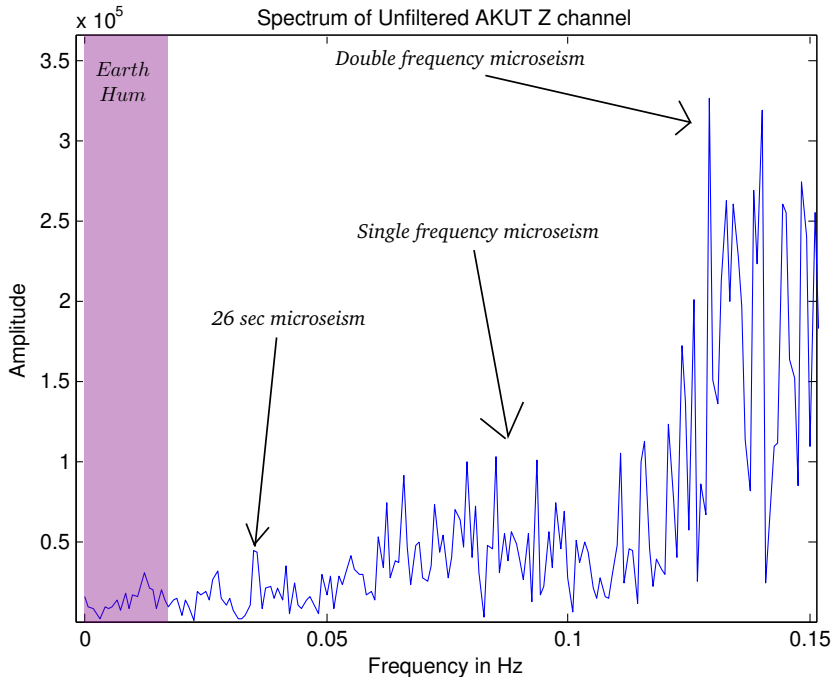
Figure 2.5: Plot of the Earth's hum, the 26 s microseism, and the SF and DF for a 20-minute long time series recorded at station AKUT. Note the relative amplitude of each of the signals.

2.4   Survey Geometry

The IRIS database holds records of time series from several broadband seismometers stationed on Akutan. The survey geometry is thus imposed: unlike in typical field applications, there is no opportunity to install and configure array elements in a geometry that will yield high resolution results. Although the deployment of IRIS instruments in remote locations and the storage of data has created a wealth of information on current and past seismic events that is accessible by any interested individual, there exist no current comprehensive methods for selecting and interpreting which portions of the recordings should be used for modeling the subsurface when the only available option is data from a very sparse and randomly arranged seismic array. This project serves to fill that void.

There are two networks containing broadband instruments that are in use on Akutan. Data obtained at stations from both will be used in the application of the theories presented in this dissertation. The first, abbreviated AV for the Alaska Volcano Observatory, is operated by the USGS-Anchorage and the Geophysical Institute at the University of Alaska. The second network, from which only one station is available, is abbreviated AT for the Alaska Tsunami Warning Seismic System and is operated by the West Coast and Alaska Tsunami Warning Center. The broadband recording from the AT network spans the period from 2002 - present (Incorporated Research Institutions for Seismology, 2012a). There are 13 additional broadband stations in the AT network ($\geq$ 10 to $>$ 80 Hz and corner period of $\geq$ 10 s) spread throughout the southern portion of the state, more than 200 km from Akutan. The AV network currently has 128 extremely short period ($\geq$ 80 Hz sampling rate with corner period of $<$ 10 s) seismometers and has had, at various points since 2005, eight broadband seismometers spread over the island. Unfortunately, only three broadband stations in the AV network have decent continual recordings over a number of years (Incor-

Figure 2.6: Stations AKGG, AKLV, AKRB, and AKUT were positioned using the GPS coordinates of each station as recorded in the header of the data files from IRIS.

porated Research Institutions for Seismology, 2012a). The remaining five AV stations are either unavailable in the database or were online for only very short periods of time. A plot of the island of Akutan with the array geometry is given in Figure 2.6.

2.4.1   Selection of Recordings and Data Extraction

Selecting and Retrieving Data from IRIS

Although the available recordings for each network span a number of years in total, there tend to be gaps in the records at various times and instruments in the networks throughout the recording periods. The selection of stretches of time where data was recorded at a maximum number of instruments was desirable so that analysis could be performed with the largest number of stations for relatively long uninterrupted time series. Because the goal is to look at microseisms, which are assumed to have periods of 1-10s, only data from broadband stations were analyzed. The instrument noise for both the Streckeisen STS-2 in use for the AT network and the Guralp CMG-3T seismometers in the AV network increases in magnitude at significantly lower periods, on the order of 1000 s citepRinglerHutt2010, and at frequencies above 10 Hz, so it is assumed that the instrumentation is appropriate for the frequency ranges of interest. Out of curiosity over the characteristics of signals in the lower end of the bandwidth of the instruments, analysis proceded for all frequencies of significantly nonzero amplitude in the power spectra of the broadband recordings at periods of greater than 1000 s. To determine data availability at each station of interest over a certain timespan, the index of IRIS data holdings was accessed (Incorporated Research Institutions for Seismology, 2012d). The full description of the IRIS data access and request procedures may be found in Appendix A. Here, only the details relevant to the extraction of the specific time series used in this analysis are given.

The IRIS Timeseries Request utility can be used to generate graphs of the data holdings for a user-defined range of time and stations (or entire networks of stations) to show where gaps exist in the records of each station over time. Figure 2.7 provides an example of the display in response to an inquiry for data holdings for the year 2009 for all stations in the

AV network.

In this example, it can be seen that gaps exist in the records for several stations. Other stations, such as AKLV, were apparently turned off for the first or second half of the year but have full coverage for the times when they were in use. The search format described in Appendix A, Section A.2, can be used to generate a graph for the network and time period of choice, up to one year in duration. Time ranges longer than one year apparently overwhelm the plotting capabilities of the system.

To obtain data for this project, a bounding box search (see Appendix A) was conducted over the island of Akutan to determine which stations were available. The graphic display from the Timeseries Request utility was used to find stretches of time with the best coverage. Out of all of the instruments at the AV network, three broadband stations were found to have long, uninterrupted recordings in the years 2005, 2008, and 2009. A fourth station in the AT network also had decent coverage in these years. A breq_fast request, formulated as detailed in Section A.5 of Appendix A, was then sent to IRIS to obtain the seismic record in the SEED format for these stations over a period from May 16, 2008 to February 16, 2009. This stretch of time was chosen for two reasons: first, it covers both summer and winter conditions. Second, all four stations were recording for the majority of the time during this period (see Figure 2.8).
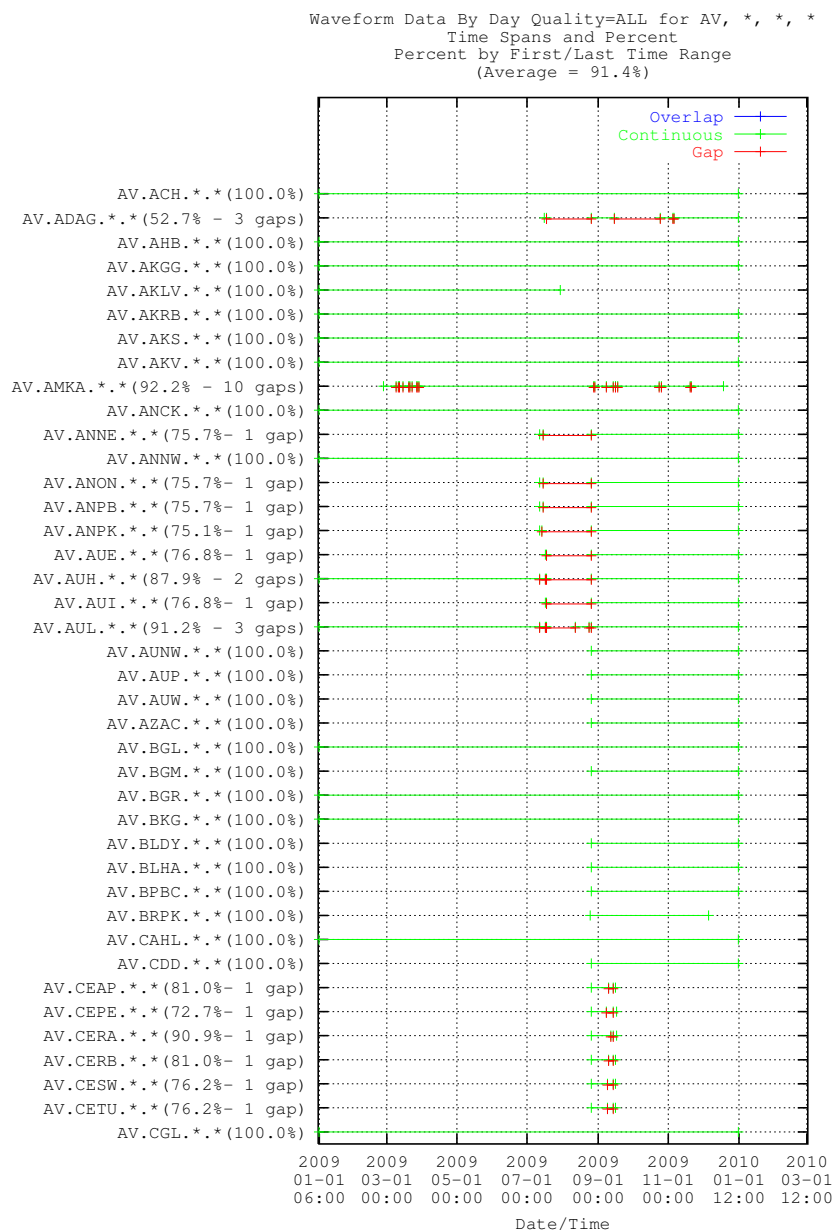
Figure 2.7: IRIS database station coverage for the year of 2009 for the AV network. The percentages in parentheses on the vertical axis indicate the amount of coverage during the time of inquiry, in this case ranging from January 1 of 2009 to March 10, 2010. The number of substantial gaps in the records are also listed on the vertical axis. Although the time series request eventually used for this project did not cover this particular time window, it is worth noting that stations AKGG, AKLV, and AKRB all have continuous coverage here, as well. This suggests the analysis presented here could be adapted for investigations of data sets recorded over several years.
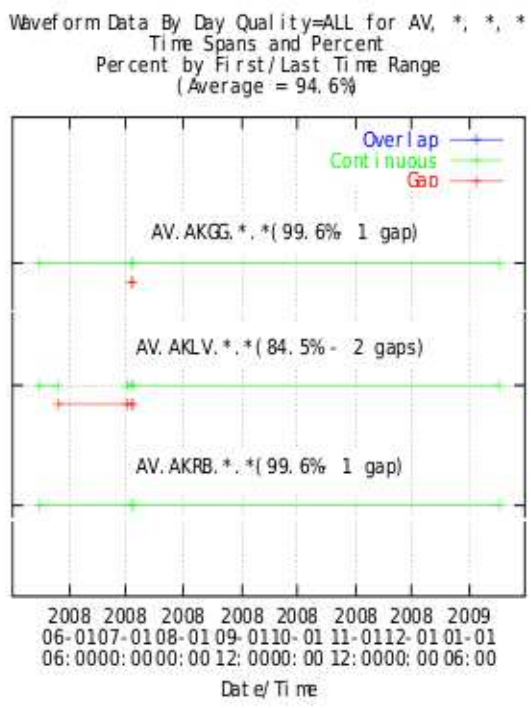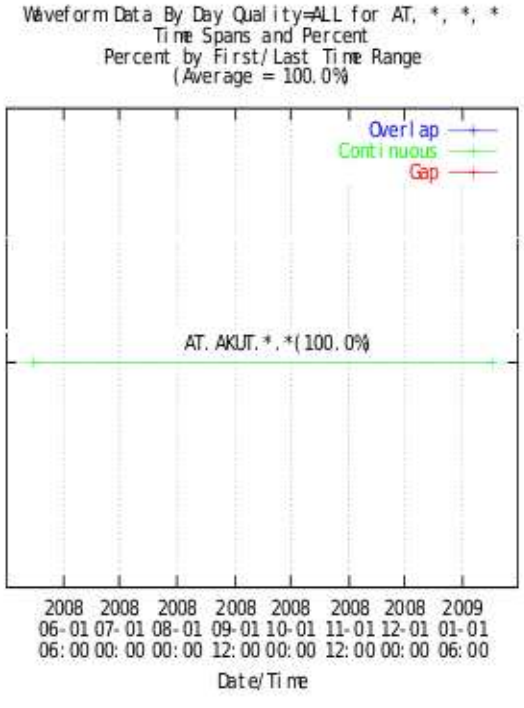
Figure 2.8: IRIS database station coverage for the broadband stations on Akutan from (a) the AV network, and (b) the AT network, over the period of May 16, 2008 - January 16, 2009. Although some gaps exist in the records from the AV network, all stations had 100% coverage during the period of August 2008 - January 2009.

Chapter 3

AMBIENT SEISMIC NOISE METHODS: AN HISTORICAL PERSPECTIVE OF STATISTICAL TECHNIQUES

3.1   On the Origin of the Term "Ambient Seismic Noise," Consideration of Its Source, and Its Applications

Microtremors, sometimes also called microseisms, are the constant low amplitude vibrations of the earth's surface and include both body and surface waves (Okada, 2003). The amount of dispacement of the surface from microtremors is far below human perception, and yet the vibrations remain a source of noise in earthquake studies. Especially when trying to study very distant or small earthquakes, microseisms often obscure the desired signal as their amplitude increases whenever amplifier gain is increased. One of the main focuses of earthquake seismology research through the 1970s was the removal of the microseism signal from the earthquake signal. In this way, the phenomenon has earned the name "noise."

The "ambient" portion of the term is used to convey the pervasive nature and lack of impulsive source for the signal — ambient seismic noise has been observed at stations all around the world, and possible origins include anthropogenic sources, such as ship or foot traffic, motion of water in waves in the oceans and rivers, and changes in atmospheric pressure. In the literature, many researchers focus on using ambient seismic noise data for the calculation of subsurface velocity profiles. For these studies, the source of ambient seismic noise is always considered to be diffuse. However, as was discussed in Chapter 2, there is an entire body of work devoted to determining whence the ambient signals arrive and discussing the implications of the temporal and spatial variations in the source. That is, rather than considering only omnidirectional and pervasive noise, there is instead *always* a source and it is likely that the source may be highly directional. The work presented here differs distinctly from the usual assumption that any one station used as the first term in a cross correlation to compare the waveforms at two stations within an array can serve as a virtual source because the ambient energy is coming from all directions at all times. Here, it is assumed that there is an actual source of the noise and that energy does not necessarily always arrive from all azimuths. For this case, it is crucial to understand which array elements are closest to the actual source in order to correctly determine travel times and, thus, velocity structure.

In recent times, the advent of digital recording and increased capacity of data storage media has made feasible the interpretation of ambient noise for subsurface exploration. Statistical analysis of microtremor has proven that the noise is generally a stationary stochastic process. In Okada (2003), a review of the methods of analysis is provided. It was found that, regardless of time, the microtremor signal for signals lasting less than three hours demonstrated constant values of average, variance, and autocorrelation function in ampli-

tude, implying that the microtremor signal is a stationary stochastic process. However, it was also observed that the stationarity did not hold for signals measured more than three hours apart from each other. Stationarity in space has also been observed over distances of 1-3 km. Beyond this distance, it appears that the stationarity in space is assumed rather than proven. A particular advantage of the work presented in this dissertation is that the assumption of stationarity in space and time is unnecessary as the method is equally capable of handling non-stationary, non-stochastic, and highly directional signals. Once the noise was proven to be stationary in time (when measured over appropriate lengths of time) and space (over limited volumes) and could be represented as a stationary stochastic process, one of two primary statistical methods of detecting the surface waves in the noise is typically applied, depending on the survey geometry. The first is the spatial autocorrelation coefficient method by Aki (1957). The second is the frequency-wavenumber (f-k) method. The f-k method uses only the signal measured by the vertical component of the seismometer or geophone and, by the assumption that surface waves dominate the body waves recorded as microseismic noise, only measures Rayleigh waves. The SPAC method is also capable of using information from the horizontal channels and can thus analyze Love waves if they arise (Okada, 2003). The number of studies that use one of these two methods to produce images of the subsurface from recordings of the noise signal at two or more stations has increased dramatically since the first successful studies by Okada *et al.* (1990), Shapiro *et al.* (2005), and Sabra *et al.* (2005). Although each method has its advantages, disadvantages also exist that limit the feasibility of using seismic noise data in particular circumstances as described in Sections 3.2 and 3.3 below.

For the case of Akutan, even without the assumption of directional noise, it is apparent that the geometry requirements of either method are not met by the sparse array available

on the island. It is the pursuit of using seismic noise for determining velocity models of the subsurface for both arrays and recording times that are unsuitable for either method that forms the motivation of this study.

3.2   Spatial Autocorrelation Coefficient Method

One of the earliest and most commonly cited publications on the stationary stochastic nature of ambient seismic noise and the possibilities for attainment of dispersion curves thereof comes from Aki (1957), where the stochastic method now known as the spatial autocorrelation coefficient (SPAC) technique was first explored in detail. The paper attempts to provide a method of analysis for time and space spectra of seismic noise where it is too complicated to perform phase analyses. This is the earliest example in the literature of representing both the temporal and spatial components of microseismic noise as stationary stochastic processes. The product of this investigation was the first step towards a solution to the problem of pervasive microtremor noise in seismic signals found in recordings made with the intent of detecting earthquakes in Japan that reveals the statistical character of the noise as well as the subsurface through which it travels via the production of phase velocity dispersion curves. The investigation of microtremor noise was motivated by the study by Kanai *et al.* (1954), among others, where it was observed that the spectral peaks for microtremor noise were identical to those produced by earthquakes and were affected by the local geology where the signal was observed. Unfortunately, due to the lack of digital recording methods, the analysis of signals with frequencies above 1 Hz and the attempt to determine the phase velocity curves was unsuccessful because the signals were not stationary over the limited length of time that they could be recorded.

A full derivation of the theory of SPAC is not necessary for the purposes of this project

as the method is not used for analysis, but a rough outline is provided in Appendix D, Section D.2, to inform the discussion of the limits of the method for the particular case of the array at Akutan.

From the outline of the derivation in Appendix D, it is obvious that the method of SPAC is unconcerned with determining the source direction of the noise since the results arise from spatial averaging of signals from all angles. Instead, the SPAC method seeks to find the scalar velocity of the signal. The technique is only viable for omnidirectional sources of noise due to the requirement of a circular or, at the least, a particularly precise array of three sensors at the points of an equilateral triangle and one sensor located at its center (Okada, 2003), and the spatial autocorrelation coefficient itself is a function of the array radius as well as the frequency and phase velocity.

One of the primary benefits of the SPAC method as compared to the f-k method discussed in the next section is that the array size required for SPAC is much smaller and requires fewer components. The method also allows one to observe Love waves in addition to Rayleigh waves if three component seismometers or geophones are used. The ability to isolate the Love wave signal and the simplicity of extending the Rayleigh wave theory to the case of Love waves is particularly attractive because it allows one to determine the s-wave velocity model for the subsurface, which is not possible when using the f-k method (Okada, 2003). Aki's 1957 suggestions for applications for microtemor analysis included locating the source of volcanic tremor, studying the coda of waves generated in earthquakes, and discovering the location of the epicenter for small earthquakes. In light of the research suggesting that many components of the ambient seismic spectrum are not homogeneously distributed (see, for example, the references in Chapter 2), the SPAC technique may be inappropriate because the case of directional noise cannot be modeled using

this technique.

Subsequent to the publication of the paper by Aki in 1957, the field of research involving SPAC was mostly limited to studies in Japan as scientists in the west focused more on finding the direction of energy, possibly due to a desire to locate signals within the microseism bandwidth generated by nuclear tests during the Cold War, and limits in the amounts of data that could be recorded prevented the collection of stationary data sets. Even in recent years, the requirements for specific survey geometry limits the number of cases in which SPAC can be applied. Several studies do exist, however. Besides the examples cited in the SEG monograph by Okada (2003), other studies have since come to completion (e.g., Haney *et al.*, 2012). It is difficult to overstate the influence of Aki's seminal publication in convincing scientists of the potential for using rather than attempting to eliminate seismic noise.

3.3 Frequency-Wavenumber Spectral Method and Beamforming

The power spectrum of a random time series is a description of the frequency content of that signal. The expansion of this concept to a series that is also stationary in space requires the addition of a spatial dimension to the power spectrum, as well. This new spectrum is the frequency-wavenumber (f-k) spectrum, also referred to as the frequency-wavenumber power spectral density (PSD). The information contained in the f-k PSD describes, in the case of signals consisting of spatially and temporally stochastic microseismic noise, both the frequency content and the vector velocity of the signal. The f-k spectrum is thus a statistical parameter that allows one to detect both the frequency and directional, or vector,

velocity of a high-energy signal in the microtremor bandwidth. As in the SPAC method, the information used to build dispersion curves comes from the spectra of the signals received. Velocity models of the subsurface may be produced by inversion of either phase- or group-velocity dispersion curves.

Historically, the theory for the f-k technique was first developed and then applied during the Cold War because of its suitability for detecting the directional component of the velocity vector of waves generated by underground nuclear testing (Okada, 2003; Capon, 1969). The focus on the direction of arrival is the most elemental difference in the functionality of the SPAC and f-k methods. Other significant differences between the two methods concern the type of energy that the method is capable of measuring and the requirements for array geometry. Although one may assume that the majority of the wave energy in the microseismic signal comes from surface waves, the f-k method does not discriminate against strong signals from body waves. Researchers may opt to use only the vertical component of a sensor to focus on Rayleigh waves. Like the SPAC method, the array size for the f-k method is chosen based on the depth to the target of interest. Although the requirements for survey geometry are much less specific for the f-k method, resolution of vector velocity increases with the number of array elements because larger arrays contain larger numbers of station pairs.

There are two ways of determining the f-k power spectrum for a signal that may or may not be a stochastic process. The first way consists of executing the normal steps used to find the frequency-only power spectral density of a time signal — find the autocorrelation function, then apply a Fourier transform to it — but for a signal with both a time and space component. The second way consists of first applying a finite Fourier transform or fast Fourier transform (FFT) to the signal and then taking the average of the square of its

absolute value. Okada (2003) suggests comparing the PSDs from at least two stations to determine if a process is stochastic in space. If the f-k spectrum is the same for the stations, the signal is considered stationary in space. In practice, a signal may be broken down in to windows of a managable length and the PSD of each block may be determined. This reduces the cost of computation, although the blocks must be long enough to capture the temporal stationarity of the signals to avoid the problems encountered in the earliest studies when recording time lengths were limited by system memory.

Nonstationary temporal effects are often noted in actual microseismic records, and arise from sources including but not limited to the passage of cold and warm fronts of weather that mark transitions to different atmospheric pressure regimes and lead to the development of atmospheric infragravity waves, as discussed in Chapter 2. Nonstationary space effects will also exist when the assumption of a plane-layered model of the subsurface beneath the array is invalid (Okada, 2003). To deal with these nonstationary effects, and to accommodate the physical limits of obtaining continuous spatial coverage that are encountered when establishing an array geometry, two statistical routines may be used to interpret the f-k PSD. The first is called beamforming. It is also referred to as the "common method" because it is generally the method of choice for analyzing the f-k PSD of microseismic signals. The second routine is called the maximum likelihood estimator, or high-resolution, method (MLM or HRM, respectively), first introduced by Capon (1969).

3.3.1   Beamforming Method

Of the two methods, the beamforming method is mathematically simpler and more intuitive. For an array of seismometers or geophones, there exists some delay time τ, sometimes called the station residual, for a wave traveling from one of the stations in the array to

another. Beamforming is a way of enhancing the microseismic signal by first correcting the arrival times by this τ for the same wave at all stations so that the wave appears to arrive simultaneously everywhere in the array and then summing the adjusted records for each time sample in the series. The time correction is done by expressing the time shift t_i at each of the i stations as

$$t_i = t_0 + \frac{\cos\phi}{c}(x_i - x_0) + \frac{\sin\phi}{c}(y_i - y_0) + \tau_i. \tag{3.1}$$

where c is the apparent velocity model, ϕ is the azimuthal direction of the signal expressed as an angle from North, t_0 is the arrival time for the station in the array that is chosen as the reference point, and the subscript i represents the values at each of the i stations in the notation of Okada (2003). After the beamforming is performed on each station, one may calculate the beam output. By calculating the beam output, or energy that exits one station and enters the next, for each possible combination in the array, the direction that produces the signal with the best SNR can be identified. The beam output is defined as a time series:

$$b(k_x/\omega, k_y/\omega, t) = \frac{1}{N}\sum_{i=1}^{N} n_i(t + t_i), \tag{3.2}$$

where k_x and k_y are the wavenumbers in the x and y directions, respectively, ω is some angular frequency, and $n_i(t)$ is the time series of the microseisms at the ith station. The power spectrum of the beam output is obtained in the usual way by first autocorrelating the timeseries and then applying a Fourier transform. A weighting function, sometimes called the array response because it is dependent on the array geometry used, is then introduced. One may prove that the estimated power spectrum of the beam output is a weighted average of the true f-k power spectrum. Details of the derivation can be found in the textbook by Aki and Richards (1980). To obtain the best estimate of the power spectrum, one must

set up a survey geometry so that the array response will approach two-dimensional Dirac delta functional, which requires a large number of stations over an area appropriate for the intended depth of exploration. If the array generates broader sidelobes in the weighting function, this will increase the error in the estimated power spectrum (Okada, 2003). For the perfect case, where the array geometry produces a weighting function that is exactly a two-dimensional Dirac delta functional, the estimate of the power spectrum for the beam output will be identical to the actual f-k power spectrum.

Overall, beamforming is traditionally used for higher frequency radio communications and avionics, but the narrowband nature of communications signals allows for much simpler processing techniques that allow the observance of phase differences between stations (Chen *et al.*, 2002). For broadband signals, beamforming focuses on time delays as waves pass from station to station in the array and requires an algorithm for least squares analysis for finding the source position. Three dimensional localization requires at least six sensors to produce an overdetermined least squares solution (Chen *et al.*, 2002). Even without the minimum station number requirement for three dimensional analysis, the array response from a four element array will have very large sidelobes and thus very low resolution.

The maximum likelihood estimator method (MLM) requires a more complex derivation. The interested reader is referred to Capon (1969) for the full details. The abridged version of the derivation is given in Appendix D, Section D.3. It is based upon the theory of the maximum likelihood estimation presented by Aster *et al.* (2005), the presentation by Aki and Richards (1980), and the summary of Aki and Richards (1980) presented by Okada (2003), but notation and the evolution from the simplest to most complex case has been constructed to suit the author's preferences. The maximum likelihood estimation is, in principle, the method of determining the model from which an observed data set would

most likely be generated when the statisticial properties of the signal are known and the forward model is understood (Aster *et al.*, 2005).

After the development of the MLM, resolution for the estimates of power spectra was limited only by signal-to-noise ratio and, of course, the ability to set up an ideal survey geometry to generate the desired beam pattern. Although the accuracy of the estimated power spectrum produced by the MLM does not depend on the particular spatial arrangement of each station in relationship to the others in the array, it does depend on the number of stations available, as proven in Okada (2003). When stations are few and far between, as in the field site of Akutan, where stations are irregularly spaced at distances on the order of hundreds of meters apart, the usefulness of statistical methods is greatly constrained. A major contribution of this dissertation is the advancement of a non-statistical method for isolating and analyzing Rayleigh waves in the seismic record and performing apparent phase velocity corrections without the need for densely spaced or specifically arranged sensor arrays.

The textbook by Aki and Richards (1980) is an excellent reference for the theoretical development and derivations of the f-k and SPAC methods. For an example of results of a practical application of the f-k method using the MLM and the detailed derivation thereof, one may refer to the 1969 publication by Capon, as well as later publications by the same author, where case studies are presented for the MLM used for interpretation of directional source of waves arriving at the large aperture seismic array (LASA) located in eastern Montana. Research on LASA was especially focused on determining the difference from signals generated by small earthquakes and those generated by nuclear bomb testing below ground, as well as on determining the direction of arriving signals, but the value of the innovation and contributions from scientists working with data from this array has long

outlived the period of the Cold War.

3.4   A General Overview of Traditional Processing Steps for the Passive Seismic Method

Whether the f-k method or the SPAC method is chosen to obtain dispersion curves, passive seismic data is generally prepared and processed in the same way. Bensen *et al.* (2007) identify and describe the four principle steps in the processing flow for ambient seismic noise data. They are presented here so that the reader may compare the author's own methods to the current standard. The steps are as follows:

- Prepare the data from a single station.

- Perform cross correlation to obtain an estimate of the Green's function and apply temporal stacking for data from two or more stations.

- Measure the surface wave dispersion curves using frequency-time analysis (FTAN) for phase and/or group velocity.

- Produce error analyses and determine quality control standards.

The steps provided in the paper by Bensen *et al.* (2007) are used throughout the body of literature on studies of applications for ambient seismic noise. After processing, many studies invert the dispersion curve data for shear wave or surface wave velocity to produce tomographic images of the subsurface. Since the inception of ambient noise surface wave tomography (ANSWT) in the 2005 studies by Shapiro et al. and Sabra et al., where data from stations in Southern California were analyzed to produce maps of the subsurface

that correlated remarkably well with the known geology of the area, the field has expanded rapidly and new applications are increasingly common. Exploration of the subsurface using the ANSWT method is greatly aided by the massive increase in data storage capabilities, number of permanent seismic stations, and improvements in the bandwidth of seismic instruments over the past two decades. The steps for processing are detailed below.

3.4.1 Data Preparation for a Single Station

For the first step, impulsive signals from earthquakes or explosions and non-stationary signals from air, land, and sea traffic are avoided using a method of time-domain normalization (see Bensen *et al.* (2007) for a review of different techniques) to isolate the ambient signal because the methods of analysis for passive waves are usually statistical and require the assumption that the energy under consideration is stationary in time and space. Even though catalogs with current and past earthquakes are maintained, it is still necessary to perform time-domain normalization if one wishes to remove the signals of smaller earthquakes that may be missing from records. Elimination of earthquakes from the ambient seismic recordings prevents what appear to be precursory arrivals in the cross correlations between two stations, which is important for the second step of the processing flow (Bensen *et al.*, 2007). Instrument effects, including instrument response, mean, and trend, are also removed.

The peak in spectral amplitudes for ambient noise lie within the microseism bandwidth, which covers periods ranging from approximately 5 to 17 seconds, but several recent studies have included or focused on longer period noise (e.g., Yang *et al.* (2007), Yao *et al.* (2006)). Signals are generally bandpass filtered to eliminate the higher frequency parts of the signal that are not used. Spectral whitening of data is also common. It is used to reduce noise from the 26 second microseism and to broaden the spectrum of cross correlations so

as to reduce the dominance of the primary and secondary multiples. Whitening is preferred over elimination of the 26 s noise with a narrow bandstop filter because the gap produced in the bandwidth by the use of a bandstop filter causes problems in the FTAN portion of the processing flow (Bensen *et al.*, 2007).

3.4.2   Temporal Stacking of Cross Correlations to Obtain the Green's Function for Station Pairs

Frequently, studies take advantage of the linearity of the cross correlation function and stack cross correlations for each day of recordings to cover longer periods of time in an effort to improve signal-to-noise ratio (SNR) when completing the second step in processing ambient data. The waveform that comes from the stacking of the cross correlations is an estimate of the Green's function for each of the station pairs and is affected only by structure within the immediate vicinity of the raypath between the pairs (Snieder, 2004). Success with extracting the Green's function for station pairs was shown in earlier studies on the use of ambient seismic noise, including the works of Sabra *et al.* (2005) and Shapiro and Campillo (2004), among many others. The longer time series can then be examined in the frequency domain to look for frequency dependencies in SNR. Cross correlations of ambient noise are generally asymmetric around the zero lag point when plotted from acausal (negative lag) to causal (positive lag). Asymmetries arise due to inhomogeneous azimuthal source distribution around the receivers and differences in "source process" (Bensen *et al.*, 2007).

Commonly, a "signal window" is created from the ambient records by first estimating the minimum and maximum arrival times for a station pair over a period band of interest in the time domain. Then, a spectral SNR can be determined by evaluating a series of narrow

bandpass filters over the signal window to obtain a filtered signal in the time domain. The peak in the time domain for each filtered is compared with the rms average of a subsequent time window of "noise" consisting of components of the record that were filtered out. The ratio between these two amplitudes is then a measure of the SNR in the time domain. Because SNR is estimated for each of the period bands of interest, this SNR is referred to as "spectral SNR" because it is a measure of signal level at different frequencies. Note that longer time series produce higher SNRs and therefore are preferable for analysis. The spectral SNR will vary depending on the frequency of interest, location, station spacing of the array, time of year, and recording length. The examination of variations in the spectral SNR for time series measured over several years may be an interesting way to use data from permanent seismic stations for remote monitoring of weather.

In this step, it is assumed that any one element of the array at a site can be used as a virtual source and taken as the first term in the cross correlation between station pairs. Although the temporal stacking will reduce the effects of transient directional noise, the attempt to accurately calculate the travel time between two stations by measuring the lag found in the peak of the cross correlation between station pairs will fail if there is persistent directional noise, as was suggested in the study by Shapiro *et al.* (2006). An incorrect calculation of traveltime results in an incorrect calculation of the Green's function. The methods developed in the next chapter offer a way to correct the apparent traveltimes to reflect the actual traveltime determined by the geology beneath the ray path connecting each station pair when directional noise violates the assumption of an azimuthally homogeneous source.

3.4.3 Frequency-Time Analysis (FTAN)

The third step in the processing algorithm focuses on measuring the phase and group velocities as a function of period by applying the FTAN method to the estimates of Green's functions found in the second step. A calculation of surface wave dispersion allows for a description of surface waves in terms of phase angle, amplitude, and group arrival times (Feng and Teng, 1983). FTAN was originally used to develop dispersion curves for group velocities. Phase velocity dispersion curves are obtained simultaneously, though, and may also be analyzed. Dispersion curves bring us closer to the final goal of most surface wave analysis studies: the construction of a 2- or 3-dimensional velocity model of the subsurface.

The evolution of FTAN as a distinct method came from two different representations of the same idea: first, the moving window analysis of Landisman *et al.* (1969), and second, the multiple filter technique of Dziewonski *et al.* (1969). Both methods were developed as a solution to the problem of determining dispersion data for frequencies with low SNR as prior methods had been limited to only the frequencies with higher SNR (Feng and Teng, 1983). The moving window analysis is a time-domain convolution of the signal with a filter of the form

$$g_\omega(t) = h_\omega(t)exp(i\omega t), \tag{3.3}$$

where ω is frequency, h_ω is a symmetric window function, and the convolution produces a complex function, $C(\omega, t)$. The multiple filter technique instead calculates $C(\omega, t)$ by applying a symmetric bandpass filter to the spectrum and then taking its inverse Fourier transform, using the center frequency of the filter bandwidth. It is easy to show that the two methods are equivalent (Feng and Teng, 1983), and this proves that the filter application in the frequency domain or the time domain produces the same result. Since both methods are interchangable, they are collectively referred to as FTAN.

FTAN is applied to signals to first calculate the group arrival time as a function of frequency by finding the point where the absolute value of the signal, $C(\omega, t)$, as determined by either filtering in the time or frequency domain, reaches its maximum value (Nyman and Landisman, 1977). Then, the value of the phase angle and amplitude for the particular frequency of interest are obtained from the phase and amplitude of $C(\omega, t)$. The group velocity dispersion is measured from the amplitude, whereas the phase velocity dispersion is measured from the progression of the phase of the signal as it travels, which is comprised of a propagation term, a phase ambiguity term, and an initial source phase. The phase ambiguity term may be resolved through the use of a global 3-D model, although the accuracy of phase velocity dispersion measurements is affected by geographical distance between the stations and frequency (Bensen *et al.*, 2007).

The FTAN operation smooths the dispersion data and thus enhances the resolution of the dispersion, even for frequencies where SNR is low. A disadvantage of FTAN is that it is poorly suited to areas where the amplitude or phase spectrum behavior changes rapidly over a narrow bandwidth and, in this region, systematic errors are introduced (Feng and Teng, 1983). One solution to the problem is to apply the residual dispersion method (RDM), where a theoretical signal designed to have a dispersion curve that approximates the curve for the actual data is generated. The original signal is cross correlated with the synthetic signal to produce what is called the "residual" signal, upon which FTAN is applied. The method of measuring the residual signal is called match-filtering. Since match-filtering eliminates the sidelobes in the cross correlation, narrowband signals cannot be resolved by this method. The match-filtering option is well-suited only to broadband measurements (Bensen *et al.*, 2007). The outcome of the RDM is a reduction of errors, as the residual signal is less dispersive than the original signal (Feng and Teng, 1983).

The method of selecting filter bandwidth has varied throughout the literature to suit the purpose of investigation and preference of the researchers. The most common type of filter is the constant relative bandwidth filter (CRBF) of Dziewonski *et al.* (1969), where bandwidth of the filter is the same at all frequencies. A second choice is the display-equalized filter (DEF) of Nyman and Landisman (1977), where the signal is averaged over an elliptical region of group velocity and period where the area of the ellipse is allowed to vary in size to prevent the disparity between the region of high time resolution and low frequency resolution at short periods and low velocities and low time resolution but high frequency resolution at long periods and high velocities in dispersion plots. This method defines a bandwidth that is dependent on both period and group velocity.

Both the CRBF and the DEF require no prior knowledge of the behavior of the dispersion. The third choice for the definition of the bandwidth parameter of the filter is known as the optimum bandwidth filter (OBF) of Inston *et al.* (1971) and Cara (1973). The benefit of this method is the optimization of the temporal resolution of the dispersion curve across the entire frequency band, but the implementation is based upon an understanding of the approximate dispersion behavior of the signal to be analyzed, which must first be obtained by a simpler method such as CRBF (Feng and Teng, 1983).

Part of the appeal of FTAN is the ease with which the calculations may be automated. By using the formulation of FTAN by Levshin *et al.* (1972), it is simple to create a series of narrowband filters that can be applied (with bandwidth determined according to the method that suits the user's preference) one at a time to the signal. Then, the generation of dispersion curves is almost instantaneous as the maximum of the signal at each frequency is plotted on a frequency- or period-traveltime graph. Based on the distance between array elements, which is always known (either from GPS positioning of the sensors in permanent

IRIS arrays or by the investigator's own choice survey geometry), the traveltime axis can then be converted to velocity. The characterization of these plots as a matrix of elements was particularly eloquently expressed by Nyman and Landisman (1977): "Consecutive rows [of the array] correspond to equal increments in...velocity, while consecutive columns correspond to equal increments in log period." Examples of applying FTAN with more specific steps and plots from earthquake seismology studies are included in Bensen *et al.* (2007).

Given that FTAN is a method of interpreting the Green's function obtained in the second processing step for determining phase and group velocity, the accuracy of the FTAN results relies on the accuracy of the extracted Green's function. When the Green's function is obtained for station pairs under the assumption that the source of the noise recorded at each is omnidirectional for times when this assumption is invalid, the resultant phase and group velocity measurements will be inaccurate. To improve the reliability of subsurface velocity profiles for cases where there exists persistent directional noise, one must incorporate the azimuth of the noise source when calculating the travel times between stations. As is demonstrated in Chapters 4 and 5, strongly directional noise will produce an offset in the peak of the cross correlation at positive or negative lag, depending on whether the first term in the cross correlation is closer to or farther away from the source of the incoming directional energy. It is easy to see why a directional source will contaminate the calculation of velocity profiles if the azimuthal approach is not incorporated into the calculations.

3.4.4 Error Analysis and Quality Control

The final step in the general processing flow for ambient seismic noise data is implementing a quality control scheme. Ambient seismic noise studies usually use arrays that are rela-

tively dense so the ray paths between stations are numerous and often exceed the amount that would be practical for user-based quality control, providing motivation for the development of automated methods that can be applied to all ray paths under inspection.

The first quality control standard is easy to implement. Experiments have proven that there is a minimum distance over which stations may be spaced to guarantee the reliability of the dispersion measurements. For group velocity, this distance is generally assumed to be three times the wavelength, λ, of the signal of interest (Bensen *et al.*, 2007). Using the relationship between wavelength, phase velocity, c, and frequency, ω, it is then possible to identify an upper limit to period, T, as follows:

$$distance > 3\lambda, \tag{3.4}$$

$$\lambda = c/\omega, \tag{3.5}$$

$$\lambda = cT, \tag{3.6}$$

and, as a consequence,

$$T = \frac{distance}{3c}. \tag{3.7}$$

Fewer phase velocity interpretations of ambient noise are available in the literature and so the upper limit for period may be more flexible than it is for group velocity calculations (Bensen *et al.*, 2007). Applying these conditions when analyzing data from stations in the USArray, for example, will serve to limit the number of station pairs.

The second quality test is of the reliability of the velocity data obtained from analysis of the ambient noise between station pairs meeting the period and distance requirements. Ideally, the way to do this would be by comparing the velocity profiles obtained from the

ambient seismic recordings with those from analysis of earthquake signals. Since the odds of an earthquake originating at one of the stations are low, a more practical assessment of quality of results from ambient noise involves the determination of the stability of the measurement, both spatially and temporally. The spatial resolution is possible only with dense station spacing. This so-called "cluster analysis" involves a comparison of the dispersion curve between each of several stations in a closely spaced array with a much more distant station. A high degree of similarity indicates spatial robustness. Since there are few places with close enough spacing for cluster analysis, the quality tests tend to focus more on temporal repeatability. The temporal repeatibility test takes advantage of the seasonal changes in the level of ambient seismic noise by evaluating the dispersion curves for the same station pairs at different times of the year over three-month periods and comparing these plots to the plot of the dispersion curve for an entire year or more of data. Examples of testing temporal stability may be found in Bensen *et al.* (2007), Lin *et al.* (2007), Yang *et al.* (2007), and others.

For the case of arrays that are unsuitable for either cluster analysis or temporal repeatability assessments due to wide station spacing and recordings over shorter stretches of time, plots of spectral SNR v. standard deviation expressed in terms of velocity from a set of measurements of SNR compiled from cross correlations of many stations across the continent of investigation over a full year can be used to estimate the robustness of the dispersion curve. At low spectral SNR (typical for periods below 10 s), the measurements are unreliable. Stacks of longer observational periods provide better results, as one might expect, but it is worth noting that SNR is dependent on the length of the time series of recording or cross correlation stacking according to a power law so there is little to be gained by producing extremely long stacks (Bensen *et al.*, 2007).

The final quality check involves comparing the results from a particular study with a smooth tomographic map of an area. If the tomographic inversion of ambient noise dispersion curve measurements over a region generates a smooth map, it is assumed that the measurements from each station pair are in agreement with each other (Yang *et al.*, 2007).

The goal of the research for this project was not to produce tomographic maps of the subsurface, but rather to provide a way to correct the difference between apparent and actual velocity that arises when noise is directional rather than homogeneously distributed at all azimuths. Therefore, the last processing step is not a focal point of the dissertation. The research instead is focused on exploring a way to assess quality when requirements for station density are not met and wide variety is found in dispersion curves produced at different times in the recordings.

Chapter 4

METHODOLOGY AND APPLICATION

4.1 Introduction

The longest continually-recording seismic networks that contribute to the IRIS database have recorded data for every year since 1982 (Incorporated Research Institutions for Seismology, 2012a). On Akutan, too, many years of data are available from most of the permanent seismic network stations. One primary objective of this research was to determine a way to define a quality control algorithm that could be implemented in a preprocessing step in the evaluation of ambient seismic noise. Currently, publications on ambient seismic noise studies generally use all available data regardless of the type of wave and azimuthal direction of approach of the signal, and time series are down sampled to accommodate memory limitations, even with modern computers. The reduction in the amount of data allows for more manageable file sizes and, thus, shorter processing times for the later steps in the algorithm presented here or additional user-selected processing techniques 3, all without sacrificing the finely spaced time sampling interval used at most locations.

The second primary objective of this study was to develop a method for determining apparent velocity corrections for the case where the source of ambient seismic noise is highly

directional. The beamforming technique is useful for large arrays when one wishes to determine the direction of the source of the noise, but the requirement of a minimum number of stations described in Chapter 3 and Appendix D prevents the use of this technique on Akutan. The station geometry for Akutan was set up for the purposes of observing volcano-related earthquake activity and is not appropriate for using the SPAC method. Even with perfect station geometry, SPAC requires the assumption of omnidirectional energy. As neither of the two traditional methods are appropriate for analysis of data from the station array at Akutan, there is a clear need for a new method that does not depend on statistical assumptions about signal behavior over time. As the permanent seismic network station coverage increases to areas that are geographically remote or unstable enough to make the installation of many broadband seismometers impractical, the need for a new non-statistical method of determining direction of approaching signal becomes paramount. Incorporating information on the direction of signal arrival into surface wave analysis is necessary for finding the correct traveltime because it allows the user to identify the distance seen by the seismic waves rather than the geographic distance between station pairs. In keeping with the first primary objective, windows described as "good", or "high quality" (as opposed to "bad," or "poor quality") are windows with strongly directional energy arrivals. The author does not claim that windows with truly omnidirectional energy should be removed from analysis when computing power allows longer time series, but instead suggests that seeking directional arrivals may act as a tool to cut down on the amount of data that must be processed since the spatial behavior of these windows is well-understood.

4.2   Methodology

4.2.1   Principal Component Analysis

Principal component analysis (PCA) is commonly used in fields other than geophysics. In statistics, PCA is used to reduce the number of variables in a dataset through eliminating variables that correlate with each other by transforming the data set into a smaller number of variables, or principal components, that are linearly uncorrelated with, or orthogonal to, each other. The majority of the variance in the data set is captured in the first few principal components. Thus, a very large matrix of observations may be re-represented as a significantly smaller matrix when redundancies are found within the data (Jolliffe, 2002). In other fields, PCA is known as the Karhunen-Loeve transform, the Hotelling transform, and the Eckart-Young theorem. Here, since the horizontal channels of a seismometer record the same signal simultaneously, the data from the two channels is clearly correlated. PCA reduces the signals recorded over the two channels to a principal component signal that is orthogonal to both, and by simple vector analysis one may see that the principal component thus points to the direction from which the signal originates, with an ambiguity of 180 degrees. The method of using PCA to determine directionality from two horizontal vectors has been used elsewhere to determine tool orientation in downhole seismic surveys (Michaels, 2001). Eigenvector/eigenvalue analysis, of which PCA is a flavor, has also been used to investigate the source of volcanic tremor (Ereditato and Luongo, 1994). However, the application to the case of determining ambient signal directionality is new. PCA also offers a way to directly detect and image off-path signal bounces that contribute to the energy recorded at seismic stations. Eliminating the off-path contributions avoids contamination from phase ambiguities and allows one to focus on noise traveling between each station

pair in the ground. The Matlab code written by the author to conduct PCA and implement quality control parameters is given in Appendix B.

The broadband seismometers installed at stations AKGG, AKLV, AKRB, and AKUT on Akutan are all three-component instruments. PCA uses the recordings from the north and east (horizontal) channels of each instrument. In general, PCA is set up by arranging the data recorded over the time series by the north and east channels into a matrix, ξ, of the form

$$\xi_n = \begin{pmatrix} E_1 & E_2 & \cdots & E_n \\ N_1 & N_2 & \cdots & N_n \end{pmatrix}^T \tag{4.1}$$

where E_n are the n samples in the time series recorded on the east channel, N_n are the n samples in the time series recorded on the north channel, and T is the matrix transpose. Each $(1,...,n)$ 2-dimensional element of ξ_n is, physically, a position in the east-north plane of particle motion such that each two-component (east,north) element of ξ_n is a vector variable of the particle motion throughout the period of observation. To accommodate analysis over a very long time series, the matrix is analyzed over a certain number of time samples in a window as:

$$\xi_{win} = \begin{pmatrix} E_1 & E_2 & \cdots & E_{win} \\ N_1 & N_2 & \cdots & N_{win} \end{pmatrix}^T, \tag{4.2}$$

where the variable *win* is the length of the window. An idea that emerged over the course of this study concerned the possibility of some frequencies offering more directional signals than others. To investigate, each window of data was narrowband filtered with a zero-phase 8th order Butterworth bandpass filter created with the Matlab FILTFILT command. Center frequencies of the bandpass filter ranged from the lowest frequency of interest, 0.01 Hz,

to the highest frequency of interest, 1 Hz, and the bandwidth of the pass band was set to ± 0.001 Hz on either side of the center frequencies. The window length was set to a number of samples that covered one wavelength for a frequency of 0.01 Hz. The larger window size provides a wider aperture, allowing one to look at more closely-spaced frequencies, and also ensures that each frequency may be present in each window. The functions bpfilt3.m and pca2staGen.m in Appendix B define the filters and show the implementation. The choice of a set window length rather than a window length definition that depends on the frequency under investigation and the consequences of this choice are discussed further in Chapter 6. For a general investigation of the variation in signal arrival direction with frequency and to speed runtime, a wider bandpass filter was used for the generation of Figures 4.1 and 4.3 in this section. An advancement of a number of points equivalent to 200 samples was chosen to assure overlap of a decent number of samples as the analysis slides along the time series, but this may be increased to reduce runtime of the codes provided in Appendix B because larger advances will reduce the number of windows generated. In keeping with the decision to avoid the assumption of temporal signal stationarity, the mean of the signal was removed from each window individually instead of from the entire signal. This allows one to accommodate times with high seas or increased ship traffic. Since PCA focuses on defining a new set of orthogonal variables, the next step in the derivation takes the definition of the scalar inner product of the north and east components as:

$$\langle E, N \rangle = \sum E_{win} N_{win} = E^T N, \tag{4.3}$$

where the superscript T indicates the vector transpose, and defining the mean for the

east component as:

$$E_{win}^- = \frac{1}{win} \sum_{i=1,...,win} E_i,$$ (4.4)

and similarly, for the north component as:

$$N_{win}^- = \frac{1}{win} \sum_{i=1,...,win} N_i.$$ (4.5)

Then, the covariance matrix is defined by taking the outer product:

$$Cov = \begin{bmatrix} \langle (E - \bar{E}), (E - \bar{E}) \rangle & \langle (E - \bar{E}), (N - \bar{N}) \rangle \\ \langle (N - \bar{N}), (E - \bar{E}) \rangle & \langle (N - \bar{N}), (N - \bar{N}) \rangle \end{bmatrix}.$$ (4.6)

The principal component directions are the eigenvectors of the covariance matrix. The azimuthal direction for the approaching ambient seismic noise is then identified by taking the four-quadrant arctangent of the eigenvectors. The eigenvector associated with the largest eigenvalue is, physically, the long dimension of the particle motion ellipse. In the quality control component of the PCA algorithm, the ratio between the maximal and minimal radii of the ellipse traced by the particle motion is used to determine whether a data window is classified as good or bad as noted below.

4.2.2   Quality Control

A new set of eigenvalues and associated eigenvectors is calculated for each window in the sliding window algorithm as defined in Section 4.2.1. The maximum eigenvalue is found for each two-dimensional covariance matrix and then the ratio between this maximum eigenvalue and the minimum eigenvalue is calculated. When the ratio is high, the ratio between the maximal and mimimal radii of the particle motion ellipse is also high,
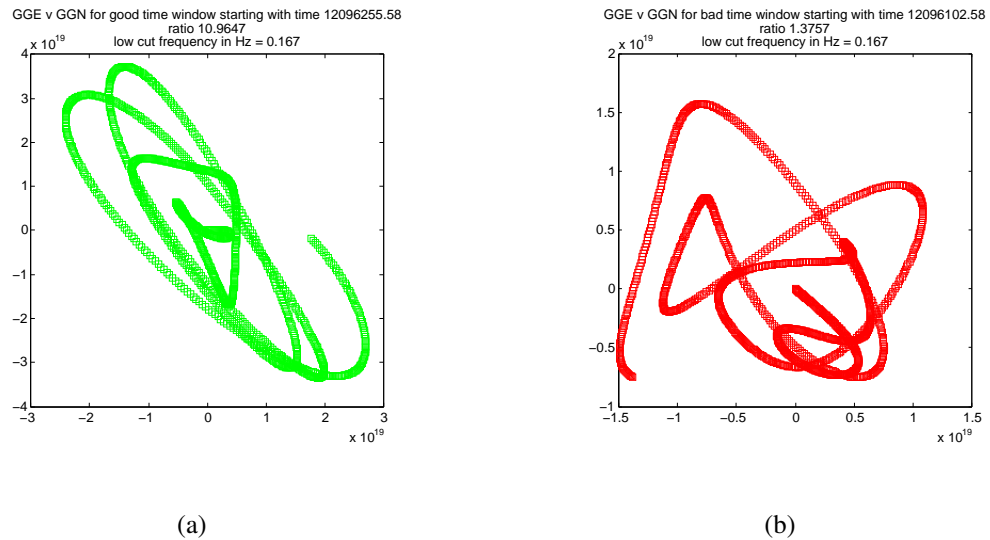
Figure 4.1: Plot of particle motion for a window with (a) high maximum-to-minimum eigenvalue ratio and (b) low maximum-to-minimum eigenvalue ratio for data filtered over a pass band of 0.167-0.334 Hz. Note that the motion in (a) is distinctly elliptical and the minor axis of the ellipse is significantly smaller than the major axis. In (b), the particle motion is more circular than elliptical because the ratio between the maximum and minimum eigenvalue is closer to 1.

producing a more linear trace of particle motion over time. This is easily visualized by plotting particle motion (Figure 4.1). The plots in Figure 4.1 are of filtered data, which is apparent because the motion appears smooth. For comparison, particle motion is plotted again for a different, arbitrary good window of unfiltered data in Figure 4.2. The motion appears much more sporadic, although the linear direction and amount of spread in the y- and x-directions are similar in both plots.

It should be stated that the images plotted in Figure 4.1 reflect the extremes in the data set. Although many windows did show particularly linear or non-linear particle motion, many others were just somewhat linear or somewhat non-linear. To select particularly linear data for further analysis, a cutoff criterion for the maximum-to-minimum eigenvalue
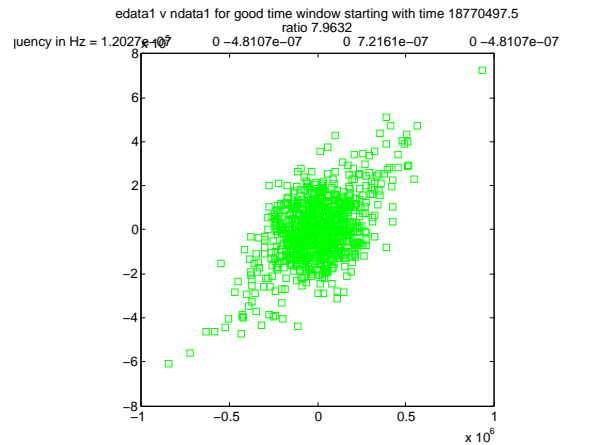
Figure 4.2: Particle motion for unfiltered good window is much less smooth than the motion plotted for filtered windows.

ratio was established so that windows with higher maximum-to-minimum eigenvalue ratios were retained and windows with lower ratios were discarded. Two approaches may be used for the selection process, and options for both are written into the function ratioanalysis.m (provided in Appendix B) so that the user may select the approach of his or her choice. The first sets the cutoff criterion as a simple percentage of the largest eigenvalue ratio for any of the windows in the time series. The default value is 60%, but this can be changed to include or exclude more data according to the user's preference. The second approach takes the mean value of the ratios of all time windows and sets the cutoff to 1.5 times one standard deviation above the mean ratio value. Cutoff criteria were chosen as a suggestion and not a rule and the potential need for an alternative cutoff scheme based on the results found when analyzing real data is explored in Chapter 6. When a good window is identified, the start time of that window, the ratio of the maximum-to-minimum eigenvalue, and the azimuth are printed to a file. Analysis of good windows with the same start time that are

identified in time series from both stations in each station pair continues along the track of the typical analysis procedure presented in Chapter 3 to produce phase velocity curves, as detailed below.

In addition to the selection of good data, it is also important to ensure that the recordings suit the assumption made for the rest of this analysis: that is, that the energy recorded is mostly Rayleigh waves. After identifying good windows with PCA, plots of particle motion in the Z direction versus the North and East direction during good windows were created to ensure that the energy observed did indeed come from a signal mostly composed of Rayleigh waves. The idea of plotting particle motion to establish wave types present in a seismic record has also been approached by Vidale (1986). The plot in Figure 4.3 is an example of one of these plots for a good window in the time series recorded at AKLV. The retrograde ellipse motion is obvious. This is in keeping with the genearlly accepted assumption that the ambient seismic noise is mostly composed of surface waves (i.e., Harmon *et al.*, 2008) and, in this case, seems to be predominantly composed of Rayleigh waves.

4.3   Application

4.3.1   Production of Phase Velocity Curves

As is typical for ambient noise analysis, the cross correlation step is performed to measure the similarity of the energy received at each station in the array. One of the four stations is first selected as the "source," based on its proximity to the shore that the user believes is most likely to be closer to the source of ambient seismic noise. Comparisons are then made for time delays associated with signals that pass first through the "source" station and onto the rest of the stations, just as one would for a controlled experiment with an impulsive
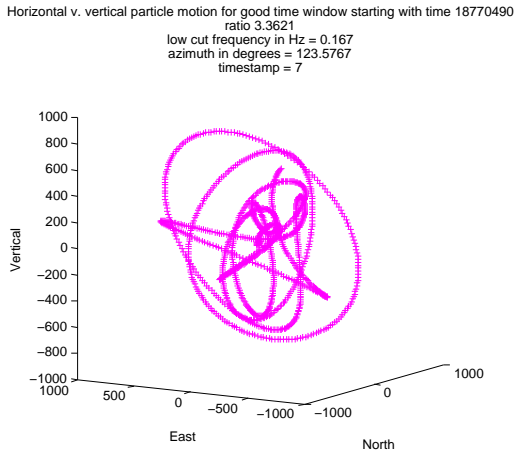
Figure 4.3: Example of elliptical particle motion for a portion of recordings on the vertical, north, and east channels of station AKLV over a good window of time. Signal has been filtered over a pass band of 0.167-0.334 Hz.

signal. The cross correlation, XC_{AB}, is calculated by averaging the product of two signals obtained at different positions A and B over the length of an observation time:

$$XC_{AB}(\tau) = \frac{1}{T} \sum_{i=1,\dots,n} x_i(t)y_i(t+\tau), \tag{4.7}$$

where τ is the lag time between the two stations or geophones, $x(t)$ is the waveform received at station A, and $y(t+\tau)$ is the time-delayed waveform arriving at station B. T is the length of the observed time series in units of time and n is the number of samples in the time series (Bendat and Piersol, 1971). By calculating first the autocorrelation of the station chosen as the virtual source, which will peak at zero, and then the cross correlation of the virtual source with all of the other stations, one can observe the delay time associated with travel from the virtual source station to the remaining stations in the array. The time delay is determined by station distance and the velocity structure of the subsurface, but it

is affected by the directivity of the incoming ambient energy. The 180-degree ambiguity introduced by the calculation of the eigenvectors can then be resolved by observing the location of the peak of the cross correlations between the virtual source station and the rest of the stations: if the virtual source station is truly closest to the actual source of directive ambient noise, all cross correlations with other, more distant stations will have a negative (causal) lag. If the virtual source station is, in actuality, the farthest away, the lag will be positive (acausal).

Preserving the large aperture to produce tightly spaced frequency steps, the raw signals from two of the available stations were cross correlated according to Equation 4.7 to produce plots similar to Figure 4.4 for each station. Each signal was zero padded by the number of points in the total signal to the next power of two before transforming into the frequency domain to avoid circular correlation. A plot of the amplitude spectrum indicates the range of frequencies over which phase information is relevant (Figure 4.5). Figure 4.6 shows the unwrapped phase changing over the range of high-amplitude frequencies found in Figure 4.5. It is worth noting that phase angle for the entire frequency range, from DC to Nyquist, plots erratically beyond the high amplitude portion of the bandwidth because the phase of almost zero-amplitude frequencies is physically insignificant.

4.3.2 Calculating Phase Velocity Curves

Given the phase spectrum for the cross correlations of each station pair, one may obtain the phase velocity dispersion curve for energy traveling from the "source" station to every other station in the array. This is done by calculating the time shift, δt, between stations at each cyclic frequency, f, found in the high amplitude portion of the spectrum of the cross correlation in Figure 4.5, using the phase angle ϕ measured at each frequency in the plot in
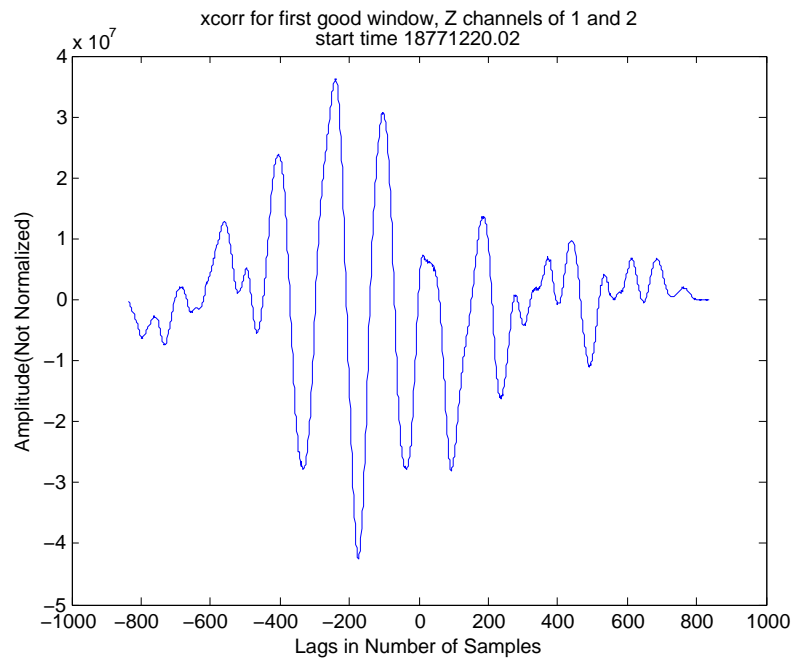
Figure 4.4: Plot of cross correlation between station AKLV and AKUT for the 20-minute period beginning at 6:00 pm on August 4, 2008. Note that the peak occurs near zero, indicating a very brief delay time between the two stations. This is in agreement with the directional arrival determined in the PCA analysis, which indicates an almost broadside arrival of energy for the station pair AKLV-AKUT.
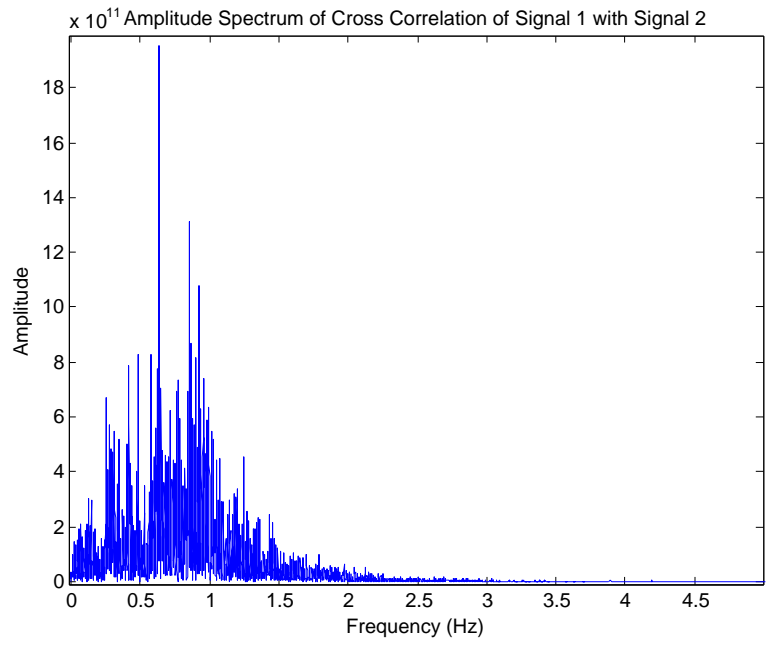
Figure 4.5: Plot of the amplitude spectrum for the FFT of the cross correlation displayed in Figure 4.4. At frequencies higher than approximately 1 Hz, the amplitude quickly drops off and then remains close to zero.
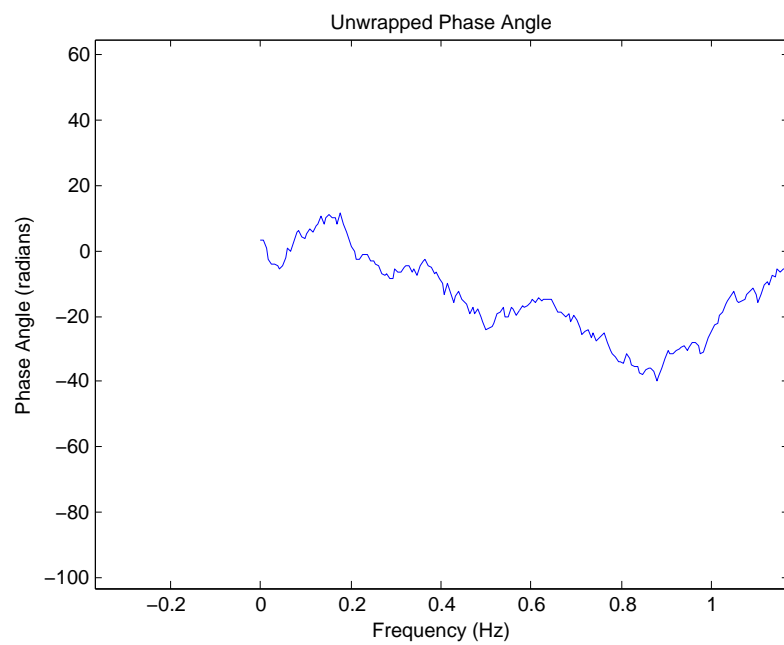
Figure 4.6: Plot of the unwrapped phase spectrum for the FFT of the cross correlation displayed in Figure 4.4. This figure shows the unwrapped phase result for the portion of the frequency bandwidth with the highest amplitude, as indicated in the plot in Figure 4.5.

Figure 4.6 according to the formula:

$$\delta t = -\frac{\phi}{2\pi f}.$$ (4.8)

Then, using the known station separation to calculate the phase velocity, v_p, for each f:

$$v_p = \frac{\Delta A}{\delta t},$$ (4.9)

where ΔA is defined on Figure 4.7.

Alternatively, the phase velocity dispersion curve can be calculated by first filtering the two signals, z_1 and z_2, by a cascade of narrow bandpass filters with center frequencies within the portion of the amplitude spectrum of the cross correlation that shows the most significant amplitude in Figure 4.5. Here, a fourth order Butterworth bandpass filter was used with a low cut frequency that was 0.001 Hz less than the center frequency and a high cut frequency that was 0.001 Hz larger than the center frequency. Note that the analysis could be done with a wider frequency bandwidth for the filters, but velocities determined by this approach will be closer to the measurement of group velocity rather than phase velocity. Next, the cosine of the angle between two filtered signals recorded on the vertical components, z_1 and z_2, was calculated as:

$$\cos\theta = \frac{z_1^T z_2}{||z_1||||z_2||}.$$ (4.10)

The inverse cosine is calculated to obtain the angle θ between the two signals. The time shift is then found by dividing θ by each center frequency, ω, used for the filtering opera-

tion:

$$\delta t = \frac{\theta}{\omega}. \tag{4.11}$$

An angle, α, is computed based on the angle of approaching energy, known from PCA, and the station geometry as pictured in Figure 4.7, by using what is known about the station geometry relative to the North-East coordinate system used for UTM to find the distance that an incoming ray "sees" between stations. In Figure 4.7, this distance is labeled ΔA. First, one of the two stations is taken as a local origin. The angle ϕ of incoming energy is measured relative to North in the PCA algorithm. The values Δx, Δy, and Δr are calculated based on the stations' UTM coordinates in the function delr.m found in Appendix B. The angle ψ is then calculated by taking the arctangent of the Δx value divided by the Δy value. Adding together ϕ and ψ gives the angle β. Subtracting β from 180° if it is larger than 90° ensures that the value is always an acute angle. Finally, the angle α is equal to 90°-β and this allows one to find the distance ΔA by Equation 4.12. The calculations of the angles are found in the function stationgeom3.m in Appendix B.

$$\Delta A = (\Delta r)(\sin \alpha). \tag{4.12}$$

Using the calculated value of α for each good window, the apparent velocity for a wave that passes from one station to the next at a particular frequency is corrected for the azimuthal angle of the directional noise by:

$$v_{true} = \frac{\sin \alpha \Delta r}{\delta t}. \tag{4.13}$$

The calculation of v_{true} may then be plotted against the bandwidth. In a dispersive medium, the lower frequencies are expected to arrive before the higher frequencies. Figure
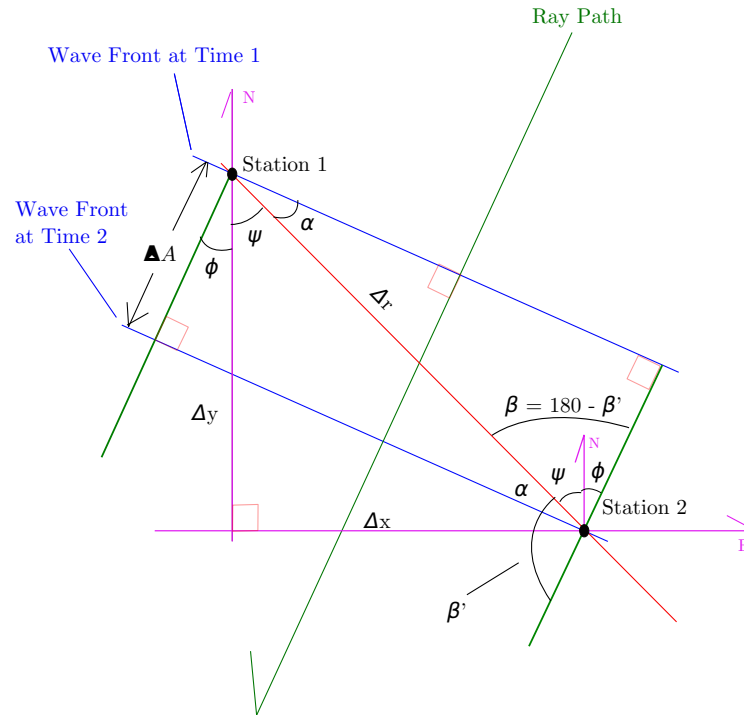
Figure 4.7: Schematic of a wave front passing through two stations of arbitrary orientation. Note that the difference in time between the time 1 and time 2 noted in the figure is the δt of Equation 4.8, and this represents the time lapse for the same wavefront to pass from one station to another. The distance labeled ΔA is the distance that a wavefront must travel from Station 1 to Station 2, and is dependent on the direction, measured as ϕ degrees from North, of incoming energy.

5.7 in Chapter 5 shows the results of this analysis.

4.4 The Importance of Directive Energy for Sparse Networks

One of the primary considerations for the calculation and analysis of phase curves is the assumption that stations are spaced less than one wavelength apart for the frequency of interest. Assuming a plane wave model, if the majority of the ambient seismic noise signal was traveling so that the stations were arranged in a "down the line" configuration, a new method would have to be used to work with frequencies where more than one cycle will pass before the energy travels between the first and second station to avoid incorrect measurement of phase angle. On the other hand, if the energy hits the station pair from an angle slightly off broadside and if the source of the energy is far enough in the distance that one may assume a relatively planar wavefront, as in Figure 4.7, the apparent distance that the wavefront must travel between the first and second station (labeled ΔA on Figure 4.7) becomes much shorter. It is this that allows us to include the higher frequencies in the ambient noise signal in the analysis of the resulting phase curves. Note that in the case of Akutan, the maximum station-to-station distance is on the order of over 10,000 m between stations AKGG and AKUT. Assuming a velocity of 3,000 m/s and frequencies over the range of 0.01-1 Hz for microseisms from natural sources, the limit imposed by the wave relationship indicates that only frequencies below 0.3 could be used for determining phase velocities in the case of a down-the-line angle of incidence. If frequencies above 0.3 Hz are of interest, one would have to determine a method of compensating for the completion of more than one phase cycle over the distance between the stations. With the "apparent"

distances on the order of 800-2000 m for the angles found for many "good" windows in the PCA analysis, a much larger portion of the available bandwidth can be used without any modifications.

Although this situation should not arise with good survey design in an active source experiment, for passive recordings where the user has no control over the station geometry the so-called "good" windows of directive energy where the direction is from an angle that shortens the apparent path to a permittable distance may be the only portions of the signal that can be used in standard analyses.

Chapter 5

RESULTS

For application of the methods developed in Chapter 4, a twenty minute time period beginning at 6:00 am on August 4, 2008 was chosen. The selection of this time period was based on the limited amount of reported earthquake activity around the world and especially nearby at this time, as determined by consulting the events catalog from IRIS (Incorporated Research Institutions for Seismology, 2012b). The results of the IRIS search for events of any magnitude occurring over the time period between August 1, 2008 and August 5, 2008 are shown in Figure 5.1.

The bandwidth of the ambient noise signal generally ranged from 0.01 - 2 Hz, as can be seen in Figure 5.2. All of the signals had similar frequency content to this example. It should be noted that the ambient noise power spectrum has the characteristics observed of power spectra of ambient noise from any location: the microseism signal ranges over a frequency of approximately 0.2 - 1 Hz and, below this, there is evidence of signal from the earth's hum (Rhie and Romanowicz, 2006).

IRIS Earthquake Browser

Magnitude (Size)
9 8 7 6 5 4 3 2 1

Events Displayed: 16 of 16

Map | Satellite | Terrain

Bering Sea

North Pacific

Depth (km)
0
33
70
-150
-300
-500
-800

POWERED BY Google   200 mi / 200 km

Map data ©2012 GIS Innovatsia, Google, MapLink, Tele Atlas - Terms of Use

Display at most:
4000 Events

Warning: Some browsers may become unresponsive when large numbers of events are displayed.

Prioritize:
Newer Events

Time Range:
Earliest Available
2008/08/01
Latest Available
2008/08/05

Magnitude Range:
Depth Range (km):
Advanced Event Filter:

Apply

Selected Lat/Lon Range:
60.22
169.37   -150.38
49.98

Magnitude	Depth (km)	Time	Lat	Lon	Mag. Type	Mag. Contrib.	Event Source	Catalog
3.91	26.2	2008/08/05 14:04:31.5600	51.2652	175.2085	MB	ISC	ISC	ISCCD
3.0	112.2	2008/08/04 21:26:45.9300	59.645	-153.002	MG	AEIC	NEIC	MHDF
3.0	18.6	2008/08/04 18:17:05.5600	56.706	-156.007	ML	AEIC	NEIC	MHDF
3.91	30.2	2008/08/03 22:18:53.3000	51.2436	-174.0599	MB	ISC	ISC	ISCCD
3.32	17.7	2008/08/03 18:38:50.5800	54.062	-163.5818	MB	ISC	ISC	ISCCD
2.7	72.7	2008/08/03 10:26:38.5800	58.683	-153.78	MG	AEIC	NEIC	MHDF
3.3	5.8	2008/08/03 08:37:43.1200	51.624	176.153	ML	AEIC	NEIC	MHDF
3.5	25.4	2008/08/03 08:25:40.7100	51.472	176.094	ML	AEIC	NEIC	MHDF
3.83	25.8	2008/08/02 19:40:04.7000	51.2582	-174.0493	MS	ISC	ISC	ISCCD
2.6	25.6	2008/08/02 07:54:27.4100	52.548	-166.739	ML	AEIC	NEIC	MHDF
4.6	224.0	2008/08/01 18:37:24.9000	52.67	-173.05	MB	NEIC	NEIC	QED
3.98	199.3	2008/08/01 18:37:18.5700	52.6299	-173.9982	MB	ISC	ISC	ISCCD
2.6	113.3	2008/08/01 16:51:19.3600	55.255	-162.671	MG	AEIC	NEIC	MHDF
3.91	38.8	2008/08/01 13:15:11.5300	52.8366	-166.9364	MB	ISC	ISC	ISCCD
3.3	26.2	2008/08/01 13:15:05.5900	52.358	-166.573	ML	AEIC	NEIC	MHDF
3.65	130.6	2008/08/01 02:50:47.0600	59.6921	-153.3754	MB	ISC	ISC	ISCCD

Figure 5.1: Plot of earthquakes of all magnitudes recorded from August 1, 2008 to August 5, 2008 near the Aleutian Islands, Alaska. Note that no earthquakes occurred in or around the period from 6:00 - 6:20 am, the time of measurement for the signals used here for analysis.
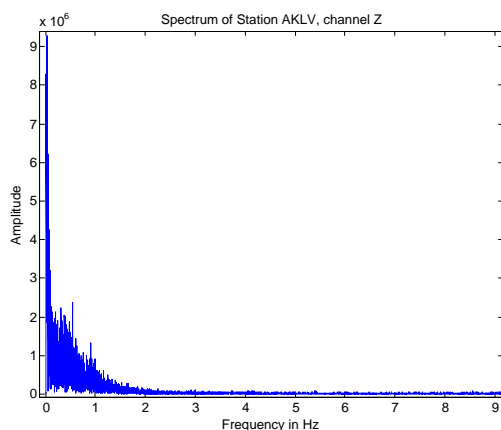
Figure 5.2: Example of the power spectrum for one channel of one station used in analysis. Frequency content of all time series was about the same. The majority of the spectral amplitude occurs within the microseismic bandwidth, as expected, and spectral amplitude above 2 Hz is almost zero.

5.1 PCA

The results from the PCA analysis and quality control algorithm were recorded for Butterworth bandpass filtered signals with center frequencies ranging over the bandwidth of the signal. The chart in Table 5.1 shows the number of good windows for each frequency where the filter bandwidth covered one decade. This was done to explore whether or not the number of good windows changed dramatically for each decade of frequency. Interestingly, although not much variation was seen on the decade bandpass scale, once the signals were filtered by much narrower bandpass filters, large differences in the numbers of "good" windows began to become apparent.

To demonstrate the effects of applying PCA, the plot in Figure 5.3 shows the distribution of azimuths for each good window and bad window for stations AKLV and AKUT

Filter Parameters in Hz			AKLV			AKUT		
Low Cut	High Cut	# Windows	# Good Windows	# Bad Windows	% Good Windows	# Good Windows	# Bad Windows	% Good Windows
0.05	0.10	73	1	72	1.4	2	71	2.7
0.10	0.20	153	4	149	2.6	9	144	5.9
0.15	0.30	233	17	216	7.3	19	214	8.2
0.20	0.40	312	21	291	6.7	22	290	7.1
0.25	0.50	393	28	365	7.1	25	368	6.4
0.30	0.60	473	33	440	7.0	29	444	6.1
0.35	0.70	554	42	512	7.6	38	516	6.9
0.40	0.80	631	45	586	7.1	49	582	7.8
0.45	0.90	707	49	658	6.9	61	646	8.6
0.50	1.00	793	67	726	8.4	58	735	7.3
0.55	1.10	875	61	814	7.0	63	812	7.2
0.60	1.20	945	54	891	5.7	67	878	7.1
0.65	1.30	1045	62	983	5.9	69	976	6.6
0.70	1.40	1104	54	1050	4.9	78	1026	7.1
0.75	1.50	1193	71	1122	6.0	107	1086	9.0
0.80	1.60	1269	83	1186	6.5	101	1168	8.0
0.85	1.70	1355	99	1257	7.3	94	1262	6.9
0.90	1.80	1421	100	1321	7.0	110	1311	7.7
0.95	1.90	1531	98	1433	6.4	108	1423	7.1
1.00	2.00	1572	116	1456	7.4	114	1458	7.3

Results From PCA for Filtered Signals

Table 5.1: Example of good and bad window statistics for two stations, AKLV and AKGG, over different bandwidths.

when a fourth order Butterworth bandpass filter is applied to the data at each station for a pass band of 0.3-0.6 Hz before conducting the PCA. Figure 2.6 in Section 2.4 shows the geographical distribution of the stations over the island.

Keeping in mind the 180-degree ambiguity introduced by calculation of the eigenvectors, it appears that the majority of the good windows have azimuths within a range of 210 or 30 degrees from North for stations AKLV and AKUT over the frequency band of 0.3-0.6 Hz. At higher frequencies, it was noted that the dominant direction of incoming energy shifted to azimuths of around 170 or 350 degrees from North (Figure 5.4), but this analysis was over a frequency band that is above that which is commonly accepted for microseisms (Okada, 2003). Stations AKLV and AKUT are the most distant stations on the island. Given the small differences in azimuthal approach of energy from the source, one may assume that the source is located some significant distance offshore.

The suggestion of signal direction changing with frequency in Figures 5.3 and 5.4 mo-

Angle histogram of good azimuths for station 1

(a)

Angle histogram of good azimuths for station 2

(b)

Angle histogram of bad azimuths for station 1

(c)

Angle histogram of bad azimuths for station 2

(d)
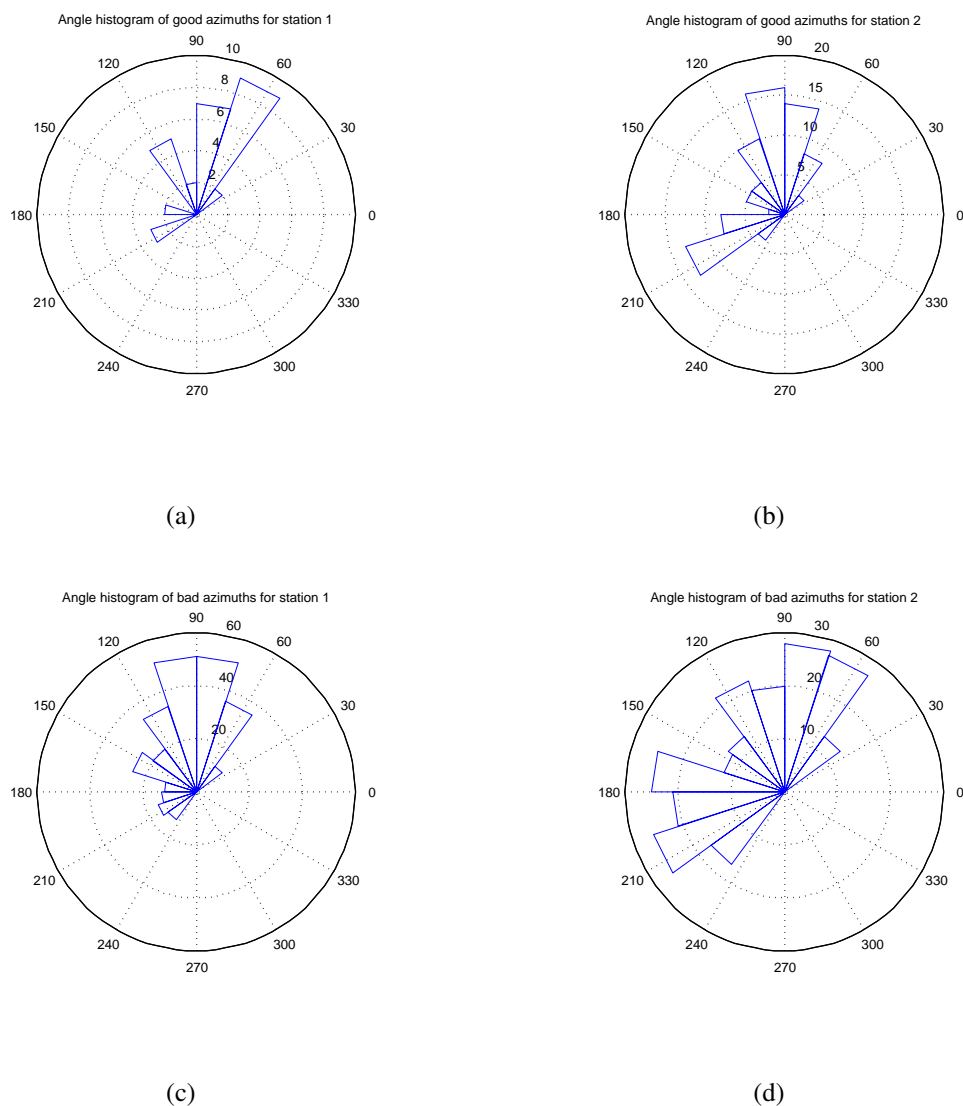
Figure 5.3: Histogram rose plots of azimuths at good windows for (a) AKLV and (b) bad time windows for stations AKLV and AKUT. Note that the analysis to produce these figures was conducted for data that had been filtered with a fourth order Butterworth bandpass filter with a passband of 0.1-0.2 Hz. This bandwidth covers portion of the spectra for each station where the largest spectral amplitudes are found. One may observe that performing PCA and adding a quality control has selected very few good azimuths. The bad windows show energy arriving from many sources. Compared to the plots in Figure 5.4, it is clear that the isolation of a few good azimuths no longer occurs as one moves to frequencies above the point where the spectral amplitude begins to decay. This could be a suggestion of the signal source being mainly local noise above the microseism bandwidth.

(a)

(b)

(c)

(d)
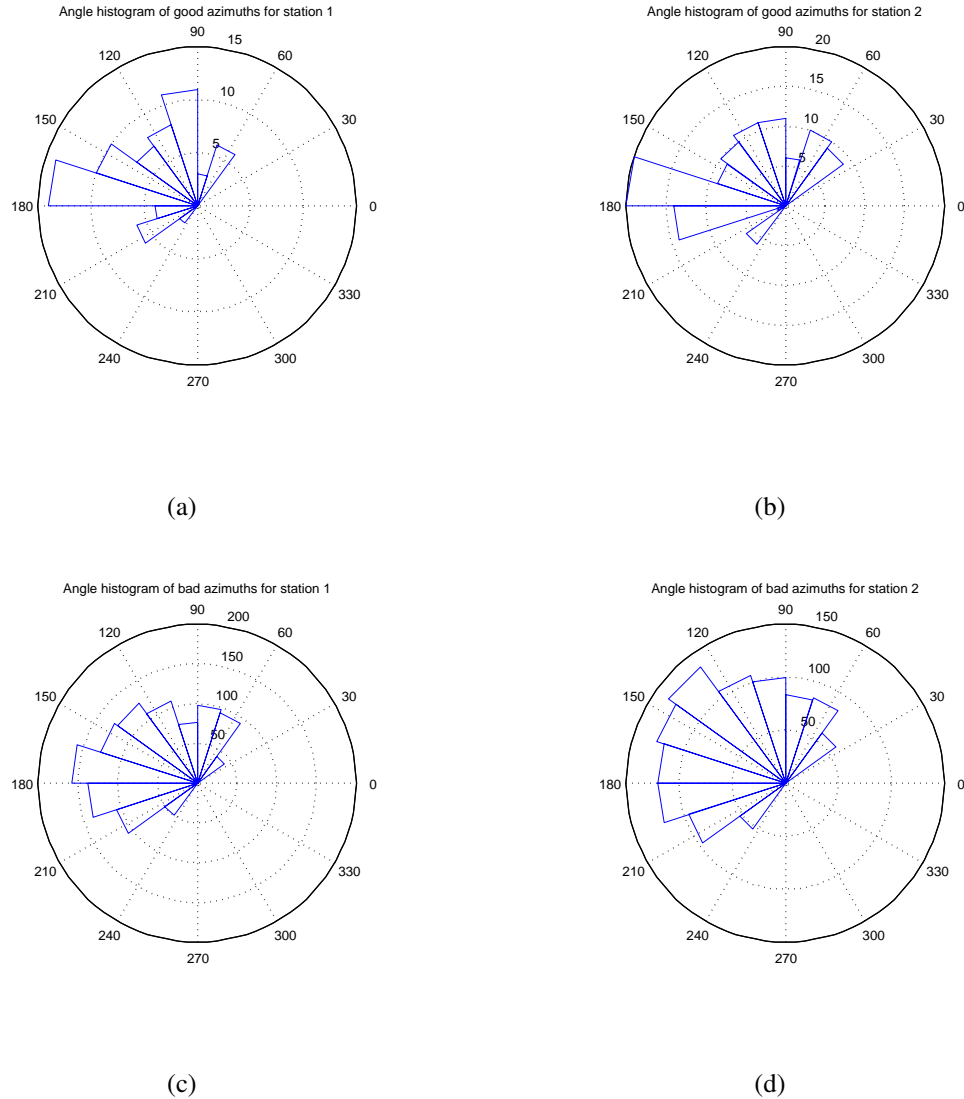
Figure 5.4: Histogram rose plots of azimuths at good and bad time windows for stations AKLV and AKUT. The fourth order Butterworth bandpass filter used to produce these figures had a passband of 0.75-1.5 Hz, above what is commonly accepted as the bandwidth for microseisms. The cutoff point for the largest spectral amplitudes is around 1 Hz so this bandwidth overlaps into the area with less signal strength. This shows that the higher frequency noise is less likely to come from a dominant direction and may indicate that higher frequency noise is a local effect.

tivated changing the code to allow one to implement filters with much narrower pass bands of $\pm$ 0.001 Hz around center frequencies. To ensure that evaluation would occur over the same portion of data, the window definition was also changed from its original version, where the length was defined as five times the number of samples in one wavelength of each frequency, to the inflexible length of the number of samples in one wavelength of the lowest frequency of interest, as described in Chapter 4. One of the primary consequences of this decision was that the maximum ratios from PCA at the higher frequencies were reduced, so the relatively strict cutoff criteria was also relaxed to allow windows with ratios that were at least 40% of the maximum. In Chapter 6, this is discussed further. After the narrowing of the pass band, it was discovered that several specific center frequencies seemed to have many good windows whereas other center frequencies had very few good windows or even no good windows.

5.2   Quality-Controlled Cross Correlations

The first insight gained from examining the cross correlations between station pairs concerns the resolution of the 180-degree ambiguity introduced by the determination of the eigenvectors in the PCA algorithm. Given the azimuth of approach, one may surmise that the energy is directed either from the northeast or the southwest. To determine which of the two directions is correct, the cross correlation between stations on opposite (north-south) ends of the island was produced. Figure 5.5 shows the results for the input of AKRB as the first station and AKLV as the second. It is noted that the lag appears to be acausal. This implies that, instead of traveling from station AKRB to AKLV, the energy passes first through AKLV. In this way, the direction of approaching energy is resolved: the energy is approaching from the northeast. Given the average angle of approach for good windows in
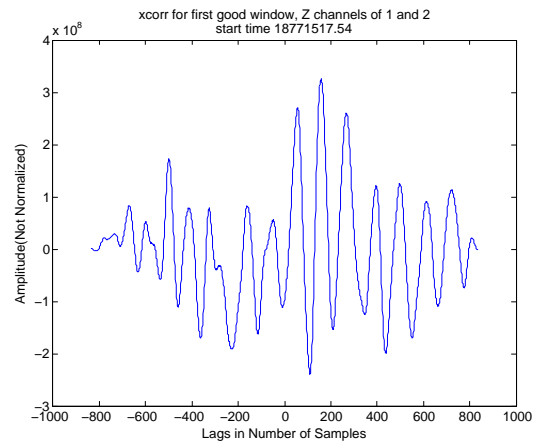
Figure 5.5: Cross correlation for AKRB (southern station) and AKLV (northern station). The positive (acausal) lag indicates the energy passes from AKLV to AKRB rather than from AKRB to AKLV.

stations AKGG and AKLV, it may be seen that the energy hits this particular station pair at an angle of intersection of almost 90 degrees with the raypath. Indeed, the cross correlations for good windows shared by these two stations shown in Figure 5.6 reveal that the peak is just on the positive lags side of zero. This indicates that AKLV is slightly closer to the noise source.

5.3 Construction of Phase Curves

Phase curves may be constructed in either of the ways documented in Chapter 4. To ensure that the code written by the author was, in fact, performing as it should, an impulsive test signal with predetermined dispersion was generated and the code was able to successfully reproduce the phase velocity dispersion curve. An example of the resultant plot from one good window of station pair AKLV-AKUT at 0.7576 Hz as calculated by the cross correla-
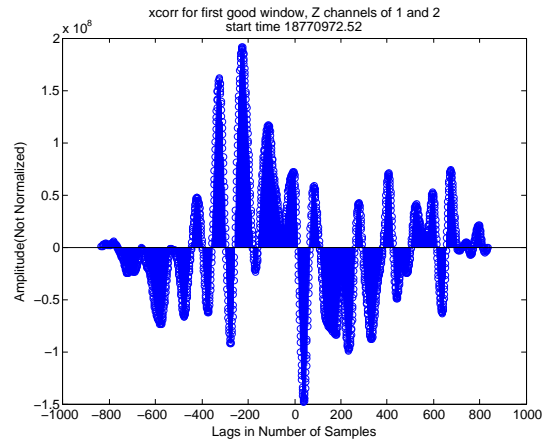
Figure 5.6: Cross correlation for the two northernmost stations, AKGG and AKLV.

tion method discussed in Chapter 4 is shown in Figure 5.7. The inversion of resultant phase velocity curves may be performed in the manner favored by the user.

To ensure that the results shown in Figure 5.7 were repeatable, a second data set consisting of another seismically quiet period of time was tested using the code. For the twenty minute period beginning at midnight on July 6, 2008, using stations AKRB and AKLV, dispersion curves were extracted for shared good windows of time. As with the station pair AKUT-AKLV, it was observed that some of the dispersion curves were not physically realistic, whereas others were reasonable. A plot of a reasonable dispersion curve for the second set of data is shown in Figure 5.9. Some of the irregularity in the unwrapped phase is captured in the very low frequencies of the curve, but then the plot is more regular and there is the suggestion of a higher mode at higher frequencies. It is worth noting that dispersion curves from any set of real data are generally irregular at the lowest frequencies, but to test the robustness of the results a physically realistic forward model was also developed to compare to the extracted results. Values of layer thickness and shear velocity were
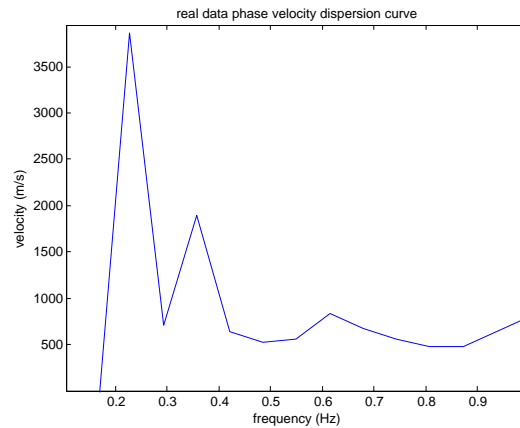
Figure 5.7: Phase velocity dispersion curve from the first of the shared good windows in the time series from station AKUT and AKLV. The filter used for the PCA analysis had a center frequency of 0.7576 Hz and produced several good windows. Note that there is strong dispersion. At frequencies below about 0.2 Hz, there is little signal, which explains the steep dropoff to unreasonable slow velocities.

chosen based on the published values of compressional velocity from the Alaska Volcano Observatory (Dixon *et al.*, 2002). The plot of the forward model was for the fundamental mode only, but produced a shape quite similar to the curves extracted for both data sets.

Despite the existence of sensible phase velocity dispersion curves for several "good" windows produced by measuring the angle of energy arriving at 0.7576 Hz, it was apparent that there was a need for additional selection criteria because there also existed several windows at this frequency with nonsensical dispersion curves. An example of an unrealistic dispersion curve is seen in Figure 6.1. A discussion of ideas for improvements to the selection criteria is provided in Chapter 6, Section 6.5.
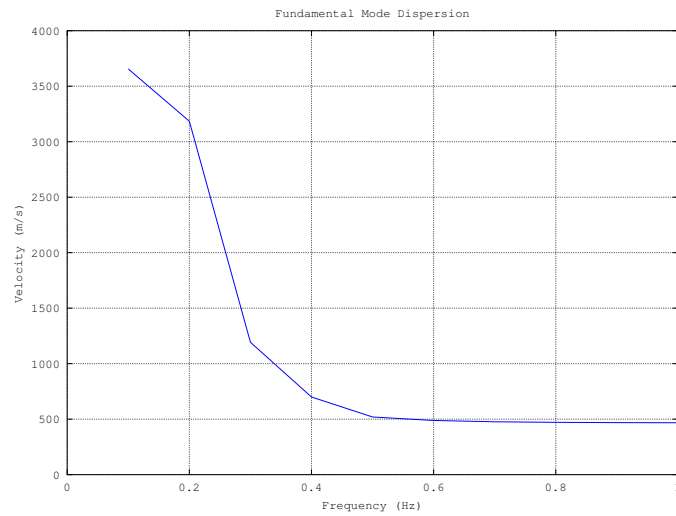
Figure 5.8: Forward model produced for the case of a layer over a half-space. Top layer shear velocity was set to 500 m/s (appropriate for unsaturated to saturated sandy material) with a thickness of 600 m. The half-space velocity was set to 4000 m/s to represent bedrock.
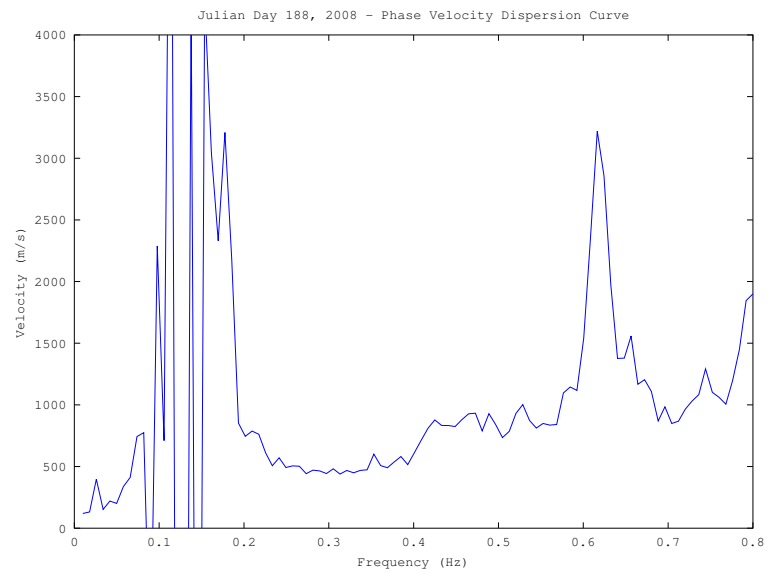
Figure 5.9: Phase velocity dispersion curve from shared good window for station pair AKRB-AKLV on Julian Day 188 (July 6), 2008 from 12:00 am to 12:20 am. Note possible arrival of a second mode above 0.6 Hz.

Chapter 6

DISCUSSION

6.1 Introduction

The current body of research on the analysis of ambient seismic surface waves shows little overlap between the works of scientists who are interested in characterizing the source of the noise (see, for example Tanimoto (2007), Webb (2008), or Yang and Ritzwoller (2008)) and those who use the noise for inversion of velocity dispersion curves for imaging the subsurface (references include Draganov *et al.* (2006), Nakata *et al.* (2011), and works cited in previous chapters). In the field of source parameterization, there generally are not attempts to use the findings about the source characteristics to inform inversion models. In the field of subsurface modeling, variations in the behavior of the source are generally not incorporated into analysis. Including information on the direction of arrival to produce dispersive velocity curves as in Chapter 5 ensures that the correct travel time between stations is used. Since the validity of the subsurface model produced from inversion of dispersion curves is reliant on the validity of the travel time measurements between stations, times with strongly directional energy should be isolated and analyzed to determine the correction factor for an angled approach. This section is dedicated to exploring the meaning

of the results in Chapter 5 and hypothesizing about the possible effects of incorporating the information learned in this study into future inversion routines.

6.2   Frequency-Dependent Source Direction

An unexpected result shown in Chapter 5 is the difference in the direction of approach for lower- versus higher-frequency passbands. The filters used to define the passbands for each of the plots in Figures 5.3 and 5.4 are both within the range of the bandwidth assumed to be appropriate for studying ambient seismic noise, although the range used for producing the second plot begins to move out of the bandwidth with the highest amplitude. This implies that signals at different frequencies may originate from different dispersive sources. That conclusion was further enhanced by performing the PCA analysis on signals that had been passed through a very narrow bandpass Butterworth filter.

In most cases, the ability to arrange for arrays that are dense enough and have a large enough geographic footprint allows researchers to analyze data using the beamforming method to determine the vector velocity of energy. However, in most of the literature reviewed to prepare this manuscript, the direction determined from beamforming was for the entire microseism bandwidth. Outside of the research on the 26 s microseism, there was generally no analysis of narrowband signals. Based on the findings in this dissertation, it is suggested that researchers who apply beamforming techniques to signals collected at arrays with better raypath coverage also examine the ways that the direction of the velocity vector changes with frequency.

6.3 Quality Control

It is particularly interesting that, regardless of the bandwidth examined, the quality control portion of PCA selected only certain azimuths out of the range of azimuths reported for all windows (Chapter 5, Figures 5.3 and 5.4). Combined with the observation of erratic particle motion found in so-called "bad" windows, as exemplified by the plot in Chapter 4, Figure 4.1, this is interpreted as an indication that local noise often contaminates the records. The near-surface geology of the island is diverse, due to the number of volcanic events affecting the stratigraphy at different points on the island (see Appendix C for details on the geology of Akutan). Local noise may be due to these geology differences, and one would especially expect to see differences in the case of station pairs AKLV-AKRB or AKGG-AKUT, where the waves likely pass through the walls of both the currently-active and ancestral calderas. The use of principal component analysis offers the possiblity of reducing the amount of contamination in the phase spectrum from off-path noise sources and allows one to interpret the results of the geology directly below the raypath.

In sparse arrays, it is otherwise impossible to determine when energy from off-path sources are included in the records. This is, of course, less of a problem when the networks used are particularly dense. For arrays with good station coverage, the additional information from off-path effects may be acceptable as resolution of phase ambiguities is much easier when there are more station pairs. In areas with sparse networks, the PCA algorithm allows the user a way to sort out local noise from directional noise. This is the only robust way to validate that the signals received at all stations are stationary stochastic processes over the space dimension.

6.4   Memory Management

Each of the time series analyzed in Chapter 5 were 20 minute long recordings with sampling intervals of 0.02 seconds. Implementing the window size determined by the function bpfilt3.m and the default quality control parameter in the function ratioanalysis.m, the number of windows for a signal filtered over a passband of 0.3-0.6 Hz was reduced from 473 windows down to less than 50 windows for each station. Out of those good windows, a much smaller portion of simultaneously recorded windows (on the order of two or three) were considered good for every station. Once the signals were passed through the very narrow bandpass filters, there were even fewer shared good windows. The user may, of course, relax the cutoff ratio, but this example serves to illustrate the point that PCA offers the opportunity for a major reduction in the size of data files to be analyzed.

As seen in Table 5.1 in Chapter 5, the amount of windows of good data is around 7% of any one time series, regardless of the decade of bandwidth over which the calculations are taken. For shorter time series, such a reduction is not desirable. However, with six months of data from four three-channel stations coming in at just over 30 GB, this analysis suggests that a maximum of only 2.1 GB of data will be selected as "good." The use of PCA to develop a quality control parameter thus offers a way to reduce the size of a data set while improving its quality.
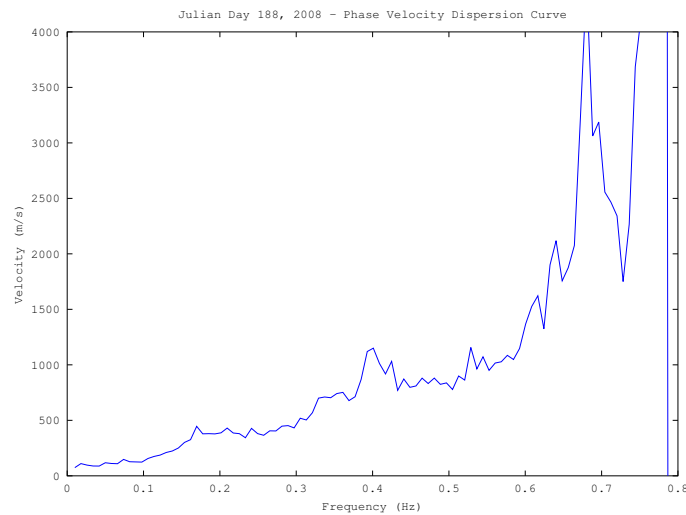
Figure 6.1: Phase velocity curve produced from measure of ϕ in filtered signals with center frequency at 0.7576 Hz in one of the windows where each station reports an eigenvalue ratio that is greater than the cutoff value. Note that the velocity units are in m/s, so the magnitudes at higher frequencies reach values that are unrealistic for earth materials, and at lower frequencies there does not appear to be dispersion, although the dispersion trend is evident in the frequency range from approximately 0.4-0.6 Hz. This suggests the need for a multistage selection criteria that incorporates both the original comparison between a window's maximum-to-minimum eigenvalue ratio and the maximum of that ratio at any window as well as a comparison between the azimuths reported by PCA at each station.

6.5   Areas for Future Research

6.5.1   Improving Quality Control

After narrow bandpass filtering the signals over 0.01-1 Hz, it was noted that many of the shared good windows produced unusual dispersion curves that could not represent the true geology between the station pair (compare Figure 6.1 to Figure 5.7 in Chapter 5 for an example).

Upon closer inspection, for the times that produced reasonable dispersion curves, the

ΔA and α that was calculated in Chapter 4 in Equation 4.12 were similar for each station because the angles, ϕ, used to determine ΔA and α for each window as calculated by PCA were approximately the same. Where unreasonable curves were produced, the angles for the windows were very different. The rose plots in Figure 5.3 show that, although the PCA quality control criterion does reduce the spread in azimuths of arrival, the signals recorded in different good windows still may have very different directions of arrival when compared to each other. In the function stationgeom3.m, the choice was made to average the azimuths of the approaching energy for shared good windows in the time series recorded at the pair of stations because it was assumed that the angles would be approximately equal if the source of the noise was distant. The differences between some of the angles for good windows suggests that this assumption is invalid. It would be best to add an additional quality control parameter that would allow one to eliminate shared windows with very different azimuths automatically. More detailed work on the causes of strong, simultaneous, local noise could also prove interesting for understanding the geology of the subsurface that produces these effects.

The choice to use a very long window length (equal to the number of samples in one wavelength of the lowest frequency in the microseism bandwidth) offered the advantage of allowing one to maintain a wide aperture and ensured that analysis was occurring for all frequencies over the same window of data. However, the price of keeping a wide aperture was the reduction of the maximum-to-minimum eigenvalue ratio in the good windows at higher frequencies. It is of interest to consider a new definition of the window length that is in fact dependent on frequency, but a full analysis of the implications of this choice is beyond the scope of this dissertation.

6.5.2 Applications

One of the original goals of the dissertation presented here was to work towards a monitoring system for imaging the magma chamber below Akutan. It was thought that the up-to-the-minute recordings of data from the permanent monitoring stations on Akutan could be rapidly inverted to produce real-time images of the subsurface that would allow scientists at AVO to detect any change in the size, shape, or depth of the magma chamber. As it was quickly realized that other, preliminary challenges had to be addressed before attempting to build such a system, the original goal was not accomplished. The work presented here would, however, serve to inform and improve any attempts to build the system in the future as it offers a way to instantly reject any data that includes too much noise from local sources. The improvement in data quality should lead to more reliable phase velocity dispersion curves and thus more accurate inversion images. In general, the author recommends that tomographic imaging studies in the future include an attempt to address the source parameters since it appears that a lot of the data currently assumed to be suitable for analysis may in fact be violating the principal of stationarity in space. Local noise does not travel between stations and, just because it is in the bandwidth appropriate for ambient seismic noise studies, it should not be included.

Another area for further research concerns the interesting findings in the differences in angle of approach for different bandwidths within the range of that expected of microtremors. The change in direction of approaching noise happens fairly gradually over the range from lower to higher frequency that was examined in the PCA analysis. As there is no mention of this elsewhere in the literature, it appears to be a new discovery and warrants investigation. The variation seems to suggest the source is complicated.

Finally, as noted in the literature on source parameters, one should expect to see differ-

ences in frequency content and signal amplitude at different times during the year because the primary natural source of ambient seismic noise is thought to be the interaction of waves with shallow sea floor. The 26 second microseism spectral peak, for instance, appears to have a larger amplitude during the winter months in the southern hemisphere when seasonal storms create higher seas conditions at the source, off the coast of Africa (Holcomb, 1980). Since records of ambient seismic noise are available through IRIS for time periods spanning many years, it is suggested that it may be possible to correlate the amplitude of seismic noise with large storm events. In areas that are particularly difficult to access, it may be of value to remotely monitor the weather for boating safety. More significantly, the correlation of long-term seismic records with weather records could allow for assessing climate variability over the past 25 years.

Chapter 7

CONCLUSIONS

For areas that are not well-suited to active seismic experiments due to geographic isolation or environmental sensitivity, the modeling the subsurface by ambient seismic signals is an attractive option. Over the past several decades, incredible advances in memory storage and data transfer speeds has allowed scientists at all levels to access data remotely and has thus quickly redefined the idea of a "field site" to now include places that the researcher has never seen in person. From remotely monitoring volcanoes to evaluating the earthquake hazards for buildings, the number of applications for analysis of ambient seismic noise has increased dramatically each year. Despite the advances, there still remain many places on earth where long-term installations of broadband instruments, each with a price tag of well over $10,000, is neither feasible nor practical. In these locations are included many of the most quickly-evolving sites on earth: active volcanoes. Especially along the Aleutians, where access to the islands is limited both by distance and by weather, direct observation of volcanism is difficult.

Ambient noise data from sparse arrays of broadband seismometers on Akutan and other volcanoes is available through IRIS, but the standard techniques of analysis include statistical methods that are impossible to implement with few stations and no control over station

geometry. The need for a new method to assess the directional approach of ambient seismic noise was addressed in this dissertation with the development of a principal component analysis algorithm. The requirement for the numbers of stations needed for analysis was reduced from four stations (for the most minimal SPAC array) to just two stations.

An additional benefit of implementing the principal component analysis algorithm was the ease with which a quality control parameter could be established. By setting some definite cutoff parameter based on the linearity of the signal received at a station in the array, one is able to discern between times in the recordings where directional noise is received through the ground from the ambient seismic source and times when the recordings featured local signals that defied the necessary assumptions of stationarity for traditional analysis. A comparison of the signal received at two stations by cross correlation resolves the 180-degree ambiguity in the source direction by observing whether the cross correlation peaked at a positive or negative lag. The amount of time that it took for a wave front to pass from one station to another was then determined and the azimuth of approaching energy was used to correct the apparent velocity. The phase velocity dispersion curves could then be calculated either by using the information about the azimuthal approach of the energy and station geometry directly or by sampling along the unwrapped phase spectrum of the cross correlation to determine the change in delay time as a function of frequency and then converting this to velocity using the calculated distance between stations. To ensure the correct calculation of the dispersion curve, an additional selection criterion to ensure that the energy in a shared good window is coming from the same direction for each signal.

Throughout the literature published by researchers interested in observing and determining the origins of different features of the ambient seismic spectrum, it is postulated that there is, in fact, a source of the noise. Furthermore, studies suggest that the source

may vary for different portions of the frequency spectrum (Tanimoto, 2007), and this was confirmed here. It is interesting to note that the idea of a source at different locations for different frequencies is not shared or, possibly, just not mentioned by researchers who use ambient seismic recordings to produce velocity maps of the subsurface. Given the evidence for the effects of directive noise from a specific source on travel times, as seen in the study by Shapiro *et al.* (2006), the study herein first identifies times when there is highly directional noise and then offers a way to correct the apparent traveltimes to produce accurate velocities for each frequency observed. The methods developed in this study will allow the user to apply a correction to the velocity profiles of the subsurface by incorporating the direction of the incoming energy into the calculations.

REFERENCES

2012 (Aug). *Alaska Community Database: Community Information Summaries (CIS).*

Aki, K. 1957. Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bulletin of the Earthquake Research Institute, University of Tokyo*, **35**, 415–457.

Aki, K., & Richards, P.G. 1980. *Quantitative Seismology - Theory and Methods*. 1 edn. W.H. Freeman & Co.

Alexander, S.S. 1963. *Surface Wave Propagation in the Western United States*. Ph.D. thesis, California Institute of Technology.

Arciniega-Ceballos, A., Chouet, B., & Dawson, P. 2003. Long-period events and tremor at Popocatepetl volcano (1994-2000) and their broadband characteristics. *Bulletin of Volcanology*, **65**(2-3), 124–135.

Aster, R.C., Borchers, B., & Thurber, C.H. 2005. *Parameter Estimation and Inverse Problems*. International Geophysics Series, vol. 90. Burlington, MA: Elsevier Academic Press.

AVO. 2012a (Aug). *About Alaska's Volcanoes*.

AVO. 2012b (Aug). *Akutan reported activity*.

Bendat, J.S., & Piersol, A.G. 1971. *Random Data: Analysis and Measurement Procedures*. New York: John Wiley & Sons, Inc.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M., & Yang, Y. 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophys. J. Int.*, **169**(Jan), 1239–1259.

Byers, Jr. F.M., & Barth, T.F.W. 1953. Volcanic activity on Akun and Akutan Island. *Pages 382–397 of: 7th Pacific Science Conference Proceedings*, vol. 2.

Byrne, T. 1979. Late Paleocene demise of the Kula-Pacific spreading center. *Geology*, **7**(Jul), 341–344.

Capon, J. 1969. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, **57**(8), 1408–1418.

Capps, S.R. 1934. *Geology of Alaska Peninsula and Aleutian Islands*. Tech. rept. 857-D. United States Department of the Interior - Geological Survey, Washington, D.C.

Cara, M. 1973. Filtering dispersed wavetrains. *Geophysical Journal of the Royal Astronomical Society*, **33**, 65–80.

Cessaro, R.K. 1994. Sources of primary and secondary microseisms. *Bulletin of the Seismological Society of America*, **84**(1), 142–148.

Chavez-Garcia, F.J., Rodriguez, M., & Stephenson, W.R. 2005. An alternative approach to the SPAC analysis of microtremors: Exploiting stationarity of noise. *Bulletin of the Seismological Society of America*, **95**(1), 277–293.

Chen, J.C., Yao, K., & Hudson, R.E. 2002. Source localization and beamforming. *IEEE Signal Processing Magazine*, Mar, 30–39.

Coats, R.R. 1950. *Volcanic Activity in the Aleutian Arc*. Tech. rept. 974-B. United States Department of the Interior - Geological Survey, Washington, D.C.

Detterman, R.L., Case, J.E., Miller, J.W., Wilson, F.H., & Yount, M.E. *Stratigraphic Framework of the Alaska Peninsula*. Tech. rept. Bulletin 1969-A. U.S. Geological Survey.

Dixon, J.P., Stihler, S.D., Power, J.A., Tytgat, G., Estes, S., Moran, S.C., Paskievitch, J., & McNutt, S.R. 2002. *Catalog of Earthquake Hypocenters at Alaskan Volcanoes: January 1, 2000 through December 31, 2001*. Tech. rept. Open-File Report 02-342. U.S. Geological Survey.

Draganov, D., Wapenaar, K., & Thorbecke, J. 2006. Seismic interferometry: Reconstructing the earth's reflection response. *Geophysics*, **71**(4), S161–S170.

Dziewonski, A., Bloch, S., & Landisman, M. 1969. A technique for the analysis of transient seismic signals. *Bulletin of the Seismological Society of America*, **59**(1), 129–139.

Ekstrom, G. 2001. Time domain analysis of Earth's long-period background seismic radiation. *Journal of Geophysical Research*, **106**(B11), 26483–26493.

Engebretson, D.C., Cox, A., and Gordon, R.G. 1984. Relative motions between oceanic plates of the Pacific Basin. *Journal of Geophysical Research*, **89**(B12), 10291–10310.

Ereditato, D., & Luongo, G. 1994. Volcanic tremor wave field during quiescent and eruptive activity at Mt. Etna (Sicily). *Journal of Volcanology and Geothermal Research*, **61**, 239–251.

Estrella, H.F., and Gonzalez, J.A. 2003. SPAC: An alternative method to estimate earthquake site effects in Mexico City. *Geofisica Internacional*, **42**(3), 227–236.

Ewing, M., Mueller, S., Landisman, M., and Sato, Y. 1959. Transient analysis of earthquake and explosion arrivals. *Geofis. Pura Appl.*, **44**, 83–118.

Feng, C.-C., & Teng, T.-L. 1983. An error analysis of frequency-time analysis. *Bulletin of the Seismological Society of America*, **73**(1), 143–144.

Finch, R.H. 1935. Akutan Volcano. *Zeitschrift fur Vulkanologie*, **XVI**, 155–160.

Fournelle, J.H., Marsh, B.D., and Myers, J. 1994. *Age, character, and significance of Aleutian arc volcanism*. Vol. G-1. Geological Society of America. Chap. 23, pages 723–757.

Gates, G.O., Powers, H., and Wilcox, R.E. 1971. *Geology of the Near Islands, Alaska*. Tech. rept. Bulletin 1028-U. U.S. Geological Survey.

Haney, M.M., Mikesell, T.D., van Wijk, K., and Nakahara, H. 2012. Extension of the spatial autocorrelation (SPAC) method to mixed-component correlations of surface waves. *Geophys. J. Int.*, **191**, 189–206.

Harmon, N., Gerstoft, P., Rychert, C.A., Abers, G.A., de la Cruz, M.S., and Fischer, K.M. 2008. Phase velocities from seismic noise using beamforming and cross correlation in Costa Rica and Nicaragua. *Geophysical Research Letters*, **35**, L19303.

Haubrich, R.A., Munk, W.H., and Snodgrass, F.E. 1963. Comparative spectra of microseisms and swell. *Bulletin of the Seismological Society of America*, **53**(1), 27–37.

Hayes, D.E., & Heirtzler, J.R. 1968. Magnetic anomalies and the relation to the Aleutian island arc. *Journal of Geophysical Research*, **73**, 4637–4646.

Hein, J.R., & McLean, H. 1980. *Paleogene sedimentary and volcanogenic rocks from Adak Island, central Aleutian Islands, Alaska*. Tech. rept. Professional Paper 1126-E. U.S. Geological Survey.

Holcomb, L.G. 1980. Microseisms: a twenty-six-second spectral line in long-period Earth motion. *Bulletin of the Seismological Society of America*, **53**, 27–37.

Holcomb, L.G. 1998. Spectral structure in the Earth's microseismic background between 20 and 40 seconds. *Bulletin of the Seismological Society of America*, **88**(3), 744–757.

Incorporated Research Institutions for Seismology. *IRIS*.

Incorporated Research Institutions for Seismology. 2010 (May). *SEED Reference Manual*. Electronic edition edn. Incorporated Research Institutions for Seismology.

Incorporated Research Institutions for Seismology. 2012a (Sep). *Data Available By Networks for Each Year*.

Incorporated Research Institutions for Seismology. 2012b (Aug). *IRIS Earthquake Browser*.

Incorporated Research Institutions for Seismology. 2012c (Feb). *rdseed Manual*.

Incorporated Research Institutions for Seismology. 2012d (Oct). *Waveform Data By Day*.

Inston, H.H., Marshall, P.D., and Blame, C. 1971. Optimization of the filter bandwidth in spectral analysis of wavetrains. *Geophysical Journal of the Royal Astronomical Society*, **23**, 243–250.

IRIS. 2012 (Sep). *BREQ_FAST Manual*.

Jolliffe, I.T. 2002. *Principal Component Analysis*. 2 edn. New York, NY: Springer-Verlag New York, Inc.

Kanai, K., Tanaka, T., and Osada, K. 1954. Measurement of the microtremor. *Bulletin of the Earthquake Research Institute, University of Tokyo*, **32**, 199–209.

Kobayashi, N., & Nishida, K. 1998. Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature*, 357–360.

Landisman, M., Dziewonski, A., and Sato, Y. 1969. Recent improvements in the analysis of surface wave observations. *Geophysical Journal of the Royal Astronomical Society*, **17**(4), 369–403.

Lauro, E. De, Martino, S. De, Falanga, M., Palo, M., and Scarpa, R. 2005. Evidence of VLP volcanic tremor in the band [0.2-0.5] Hz at Stromboli volcano, Italy. *Geophysical Research Letters*, **32**, L17303.

Lee, W.S., Sheen, D.-H., Yun, S., and Seo, K.-W. 2011. The origin of double-frequency microseism and its seasonal variability at King Sejong Station, Antarctica. *Bulletin of the Seismological Society of America*, **101**(3), 1446–1451.

Levshin, A.L., Pisarenko, V.F., and Pogrebinsky, G.A. 1972. On a frequency-time analysis of oscillations. *Ann. Geophys.*, **28**, 211–218.

Lin, F.-C., Ritzwoller, M.H., Townend, J., Savage, M., and Bannister, S. 2007. Ambient noise Rayleigh wave tomography of New Zealand. *Geophysical Journal International*, 18p.

Longuet-Higgins, M.S. 1950. A theory of the origin of microseisms. *Philosophical Transactions of the Royal Society of London, Series A*, **243**, 1–35.

Lonsdale, P. 1988. Paleogene history of the Kula Plate: Offshore evidence and onshore implications. *Geological Society of America Bulletin*, **100**, 733–754.

Lu, Z., C. Wicks, Jr., Power, J.A., and Dzurisin, D. 2000. Ground deformation associated with the March 1996 earthquake swarm at Akutan volcano, Alaska, revealed by satellite radar interferometry. *Journal of Geophysical Research*, **105**(B9), 21483–21495.

Michaels, P. 2001. Use of principal component analysis to determine down-hole tool orientation and enhance SH-waves. *Journal of Environmental and Engineering Geophysics*, **6**(4), 175–183.

Miller, T.P., and Smith, R.L. 1987. Late Quaternary caldera-forming eruptions in the eastern Aleutian arc, Alaska. *Geology*, **15**, 434–438.

Morton, R. 2010 (Sep). Perceiving tomorrow's marine shipping spill risk. *Pages 1–9 of: OCEANS 2010*.

Motyka, R.J., and Nye, C.J. 1988. *A geological, geochemical, and geophysical survey of the geothermal resources at Hot Springs Bay valley, Akutan Island, Alaska*. Tech. rept. Report of Investagation 88-3. Alaska Division of Geological & Geophysical Surveys.

Nakata, N., Snieder, R., Tsuji, T., Larner, K., and Matsuoka, T. 2011. Shear wave imaging from traffic noise using seismic interferometry by cross-coherence. *Geophysics*, **76**(6), SA97–SA106.

National Weather Forecast Service Office. 2012. *Puff Model*.

Nyman, D.C., and Landisman, M. 1977. The dispaly equalized folter for frequency-time analysis. *Bulletin of the Seismological Society of America*, **67**, 393–404.

Okada, H. 2003. *The Microtremor Survey Method*. Geophysical Monograph Series, vol. 12. Society of Exploration Geophysicists.

Okada, H., Matsushima, T., Moriya, T., and Sasatani, T. 1990. An exploration technique using long-period microtremors for determination of deep geological structures under urbanized areas. *Geophys. Explor.*, **43**, 402–417.

Plafker, G., and Berg, H.C. 1994. *The Geology of Alaska*. The Geology of North America, vol. G-1. The Geological Society of America, Inc. Chap. 1, pages 1–16.

Press, F. 1956. Determination of crustal structure from phase velocity of Rayleigh waves, I: Southern California. *Bulletin of the Geological Society of America*, **67**, 1647–1658.

Press, F. 1957. Determination of crustal structure from phase velocity of Rayleigh waves, II: San Francisco Bay region. *Bulletin of the Geological Society of America*, **47**, 87–88.

Press, F., Ewing, M., and Oliver, J. 1956. Crustal structure and surface-wave dispersion in Africa. *Bulletin of the Seismological Society of America*, **46**, 97–103.

Rhie, J., and Romanowicz, B. 2006. A study of the relation between ocean storms and the Earth's hum. *Geochemistry, Geophysics, Geosystems*, **7**(10), 1–23.

Richter, D.H., Waythomas, C.F., McGimsey, R.G., and Stelling, P.L. 1998. *Geologic map of Akutan Island Alaska*. Tech. rept. Open-File Report 98-135. U.S. Geological Survey.

Sabra, K.G., Gerstoft, P., Roux, P., and Kuperman, W.A. 2005. Surface wave tomography from microseisms in Southern California. *Geophysical Research Letters*, **32**, L14311.

Schuster, G. 2009. *Seismic Interferometry*. New York: Cambridge University Press.

Shapiro, N.M., and Campillo, M. 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, **31**, L07614.

Shapiro, N.M., Campillo, M., Stehly, L., and Ritzwoller, M.H. 2005. High-resolution surface-wave tomography from ambient seismic noise. *Science*, **307**(5715), 1615–1618.

Shapiro, N.M., Ritzwoller, M.H., and Bensen, G.D. 2006. Source location of the 26 sec microseism from cross-correlations of ambient seismic noise. *Geophysical Research Letters*, **33**, L18310.

Snieder, R. 2004. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Phys. Rev. E*, **69**, 046610.

Snieder, R., & Wapenaar, K. 2010. Imaging with ambient noise. *Phys. Today*, **63**(9), 44–49.

Tanimoto, T. 2007. Excitation of normal modes by non-linear interaction of ocean waves. *Geophys. J. Int.*, **168**, 571–582.

Tolstoy, M., Vernon, F.L., Orcutt, J.A., and Wyatt, F.K. 2002. Breathing of the seafloor: Tidal correlations of seismicity at Axial volcano. *Geology*, **30**(6), 503–506.

Turner, II C.G. 1972. Preliminary report of archaeological survey and test excavations in the Eastern Aleutian Islands, Alaska. *Arctic Anthropology*, **9**(2), 32–35.

Urick, R.J. 1984. *Ambient Noise in the Sea*. Washington, D.C.: Undersea Warfare Technology Office Naval Sea Systems Command Department of the Navy.

Vallier, T.L., Scholl, D.W., Fisher, M.A., Bruns, T.R., Wilson, F.H., von Huene, R., and Stevenson, A.J. 1994. *Geologic framework of the Aleutian arc, Alaska*. Vol. G-1. Geological Society of America, Inc. Chap. 11, pages 367–388.

Vidale, J.E. 1986. Complex polarization analysis of particle motion. *Bulletin of the Seismological Society of America*, **76**(5), 1393–1405.

Waythomas, C.F., Power, J.A., Richter, D.H., and McGimsey, R.G. 1998. *Preliminary Volcano-Hazard Assessment for Akutan Volcano, East-Central Aleutian Islands, Alaska*. Tech. rept. Open-File Report 98-360. U.S. Geological Survey.

Webb, S.C. 2008. The Earth's hum: the excitation of Earth normal modes by ocean waves. *Geophys. J. Int.*, **174**(2), 1–25.

Yang, Y., & Ritzwoller, M.H. 2008. Characteristics of ambient seismic noise as a source for surface wave tomography. *Geochemistry, Geophysics, Geosystems*, **9**(2), 18p.

Yang, Y., Ritzwoller, M.H., Levshin, A.L., and Shapiro, N.M. 2007. Ambient noise Rayleigh wave tomography across Europe. *Geophysical Journal International*, **168**, 259–274.

Yao, H., van der Hilst, R.D., and de Hoop, M.V. 2006. Surface-wave tomography in SE Tibet from ambient seismic noise and two-station analysis: I.–Phase velocity maps. *Geophysical Journal International*, **166**, 732–744.

Appendix A

DATA ACCESS

A.0.3 The IRIS Interface

To determine which seismic stations are available in a specific region, it is easiest to perform a map search through IRIS by using the Timeseries Query Form request utility, found at:

$$http://www.iris.edu/SeismiQuery/timeseries.htm \tag{A.1}$$

and clicking the map icon or entering the latitude and longitude of a bounding box around the area of interest, as in Figures A.1 and A.2.

The search results can be viewed in a list and chart as in Figure 2.7 or as a map with stations placed by their latitude and longitude coordinates.

$$http://www.iris.edu/data\_available/[network]?timewindow = [start] - [end] \tag{A.2}$$

where [network] is the two-character network code of interest (e.g., AV for the network named Alaska Volcano Observatory), [start] is the start time of the window of interest, and

Figure A.1: IRIS interface for performing a map search to find station coverage for an area of interest. Click the map icon to select a bounding box, which will automatically enter the latitudes and longitudes of its location, or manually enter values for latitude and longitude of a bounding box in the region highlighted by the yellow square.
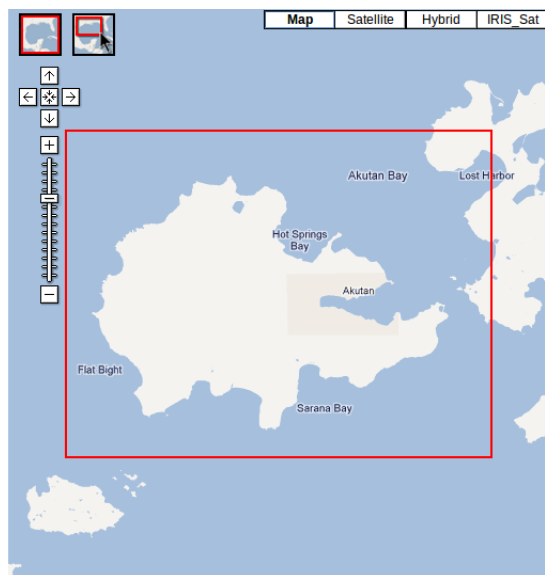
Figure A.2: A search for stations on Akutan was conducted over the bounding box selected using the map interface outlined in red.

[end] is the end time of that window. The format for the start and end times is YYYY/MM/DD, so that

$$http://www.iris.edu/data\_available/AV?timewindow = 2010/08/01 - 2010/09/07$$

(A.3)

searches from the first day in August of 2010 to the seventh day in September of 2010.

After searching the database to see if there are recordings over the period of interest, one must perform a breq_fast request to obtain the data by ftp transfer or some other, physical medium. The request is of the format:

$$STA\,NN\,YYYY\,MM\,DD\,HH\,MM\,SS.TTTT\,YYYY\,MM\,DD\,HH\,MM\,SS.TTTT\,\#\_CH1\,CH2\,CHn\,LI$$

(A.4)

where STA is the three- or four-digit station code and NN is the two-digit network code. The first YYYY MM DD HH MM SS.TTTT block is the start time of the request and the second is the end time. YYYY is the four-digit year, MM is the two-digit month, DD is the two-digit day, HH is the two-digit hour, MM is the two-digit minute, SS is the two-digit second, TTTT is the four-digit expression of the fraction of the second. The _ indicates the number of channel designators in the request and the CH1...CHn are the channel designators of interest (i.e., N, E, or Z for north channel, east channel, or vertical channel). LI is the location identifier, which is an optional parameter. The wildcard ? can be used for the channels to fetch any available data. When using the wildcard for all three channels, use the number 1 for the number of channel designators so that an example may be:

$$\textit{AKLV AV } 2010\,06\,01\,00\,00\,00.0000\,2010\,08\,30\,00\,00\,00.0000\,1\,??? \qquad (A.5)$$

to retrieve data from station AKLV in the AV network from June 1, 2010 at midnight to August 30, 2010 at midnight, where the indicator for the number of channels is set to 1 because the wildcards ??? are used. Note that this example fails to retrieve data because AKLV was not recording in this time period. This emphasizes the need to perform an inquiry on data holdings before submitting a request. Requests can either be emailed directly to IRIS by following the form requirements as in the following example:

```
.NAME Kathryn Decker
.INST Boise State University
.MAIL 1910 University Drive, Mail Stop #1536, Boise, ID 83702
.EMAIL katiedecker@u.boisestate.edu
.PHONE 315 794-3455
```

```
.FAX 208 426-3888

.MEDIA Electronic (FTP)

.ALTERNATE MEDIA DVD

.ALTERNATE MEDIA EXABYTE - 5 gigabyte

.LABEL AkutanSummer

.QUALITY B

.END


AKLV AV 2010 06 01 00 00 00.0000 2010 08 30 00 00 00.0000 1 ???
```

where as many lines as necessary may be used for requesting data from various stations. The choice of "B" for QUALITY selects the highest quality of data available for the time period of interest. Details for how to format other, more complicated requests are available on the BREQ_FAST Manual website IRIS (2012).

A.0.4   Extracting Data from SEED Format

Extraction of data from the SEED volume into one of several possible formats is done using the rdseed software published by IRIS. rdseed can be downloaded for free from IRIS at Incorporated Research Institutions for Seismology (2012c). The manual, also found at Incorporated Research Institutions for Seismology (2012c), offers detailed information on the options available for the format of extraction of information from the SEED volume. Since the download for a period of six months as recorded at four three-channel stations comes as a SEED volume containing about 30GB of data, recordings over shorter periods of time for any station within the volume were extracted by using the start- and endtime options for the rdseed utility. Although the start and end time for a select portion of data

can be arbitrarily chosen by the user for any time span contained in the SEED volume, the default buffer setting for the rdseed program breaks the extracted file down into smaller, more manageable files of approximately 20 minutes in duration, plus or minus a few seconds' worth of samples. For a starting point, this default limit on the length time series was accepted and analysis was conducted on series of approximately 20 minutes.

Only the SAC file type option was used for extracting data with rdseed for this study. This choice was based on the anticipation that, at a later stage of the research, analysis would be extended to longer time series. The SEG-Y header size has a sample limit of 32,767, which worked only for time series shorter than 10 minutes for series with sampling intervals of 0.02 s. Since it was assumed that this was not long enough for producing robust results under the planned processing flow, the SAC file format was chosen as an alternative to downsampling to accommodate the SEG-Y header format. Thus, for a starting point, the readsac2.c utility written by Dr. Paul Michaels was used to convert the file from SAC to regular ASCII text files. The program readsac2.c is included in Appendix B. After conversion to ASCII, an awk script can easily be written to trim the time series from each channel of each station down to reflect the latest start time and the earliest end time of all stations. This script is also included in Appendix B. The ASCII files were then imported into Matlab (version 7.10, R2010a) for analysis.

Appendix B

CODE

B.1 C Codes

The utility readsac2.c and its associated files is used to convert files from SAC to ASCII format.

B.1.1 readsac2.c

```
// |$Id: readsac2.c,v 1.7 2012-06-03 05:10:49 pm Exp $
          |
// Paul Michaels <pm@cgiss.boisestate.edu>
    #include <stdio.h>
    #include <stdlib.h>
    #include <float.h>
    #include <math.h>
    #include <string.h>
    #include <time.h>
```

```c
#include <../include/bsac.h>
// compile command: gcc -Wall -lm readsac.c -o readsac
// reads a evenly spaced SAC file and outputs to stdout
// Two Columns: sample_time data_value


// prototypes
    int     GetHeader(char infile[], int NZDTTM[], struct
        header *hd);


    void    DisplayHeader(struct header hd, int NZDTTM
        [],
            double *tzero, int *npts, float *delt );


    int     GetSacData(int npts, char infile[], float *
        s1);


  int   main(int argc, char *argv[])
  {


//... declare variables
    FILE            *h1;
    int             j;
    float           *s1;
    double          tm;
```

```
        float              delt;
        double             tzero;   //time zero GMT Hr Min Sec
            msec==>seconds
        double             stime;   //start time to extract
        double             etime;   //end time to extract
        int                npts;
        int                NZDTTM[6]={0,0,0,0,0,0};
        char               infile[80];
        char               help[3]="-h";
        struct header      hd;


//...test for command line help ?
        if((argc==2 && !strcmp(argv[1],help)) || argc<4)
        {

//...help display
        puts(" ");
        puts(" ");
        puts("USAGE:  readsac2 infile start(s) end(s)  >
            outfile   ");
        puts(" if both start and end = -1, just dump header
            info");
        puts(" ");
        return (1);
```

```
}
else {
strcpy(infile ,argv[1]);
stime=atof(argv[2]);
etime=atof(argv[3]);
} // endif


/* Get SAC Header for Input File */
if (GetHeader(infile ,NZDTTM,&hd) == 1 ) goto exit_1;


/* Echo command */
fprintf(stderr ," \n");
for (j=0; j<argc; j++) fprintf(stderr ," %s ",argv[j
   ]);
fprintf(stderr ," \n");


/* Display Header, extract some values */
DisplayHeader(hd,NZDTTM,&tzero ,&npts ,&delt );


/* Check for evenly spaced data */
if (hd.LEVEN != 1 ) {
fprintf(stderr ,"ABORT!!!! Data not evenly spaced \n
   ");
return(1);} // endif
```

```
// bail out if −1 for both start and end times
if ( stime == −1.0 && etime == −1.0 ) goto skipit;


//... dynamic allocation of memory for data
s1=calloc(npts, sizeof(float));
  if(s1==NULL)
 {
  fprintf(stderr,"ABORT:s1 can't allocate enough
     trace memory");
  fprintf(stderr," npts  = %d",npts);
  goto exit_s1;
 }


/* Read Data from Input File */
int     rcode;
if ((rcode=GetSacData(npts, infile, s1)) != 0 ) {
 fprintf(stderr," ERROR: reading data rcode=%d\n",
    rcode);
goto exit_1; } // endif


//... Compute start and end sample
int     jstart, jend;
```

```
        stime=stime−tzero ;
        etime=etime−tzero ;
        jstart=floor ((stime/delt + 0.5));
        jend=floor ((etime/delt + 0.5));
//...check for dumb choices , fixup
        if (jstart < 0  || jstart >jend) {
        jstart =0;
        fprintf (stderr ,"WARNING!! reset jstart to zero \n");
        fprintf (stderr ," either end time <start time , or
            start < tzero=%f \n",
                tzero );}
        if (jend>npts || jend < jstart) {
        jend=npts ;
        fprintf (stderr ,"WARNING!! reset jend to npts=%d \n",
            npts );
        fprintf (stderr ," either end time < start time , or
            endtime too large \n");}

        for (j=jstart ; j<jend ; j++) {
//        tm=(float) j ∗ delt ; // for first sample = time zero
        tm=(double) j ∗ (double) delt + tzero ;   // for
            seconds into day GMT
        fprintf (stdout ,"%.3 f %f \n",tm, s1[ j ]) ;
        }
```

```
exit_1:
//          fclose(h1);
exit_s1:
        free(s1);
skipit:
return(rcode);
}        //endmain


//==FUNCTIONS=========GetHeader FUNCTION
  =====================================
        int       GetHeader(char infile[], int NZDTTM[],
          struct header *hd)
        {
        FILE    *h1;
//...open input file
        if( (h1=fopen(infile ,"rb"))==NULL )
        {
          fprintf(stderr ,"\n ABORT !! Error opening file %s
            ",infile);
          return(1);
        }


//...read header
        if(fread(hd,sizeof(char),158*4,h1)!=158*4)
```

```
                {
                 puts("ABORT—Error reading input file ");
                 return(1);
                }
        NZDTTM[0]=(*hd).NZYEAR;
        NZDTTM[1]=(*hd).NZJDAY;
        NZDTTM[2]=(*hd).NZHOUR;
        NZDTTM[3]=(*hd).NZMIN;
        NZDTTM[4]=(*hd).NZSEC;
        NZDTTM[5]=(*hd).NZMSEC;
        fclose(h1);
        return(0);
        } //end GetHeader


//=============DisplayHeader FUNCTION
   ===============================
        void    DisplayHeader(struct header hd, int NZDTTM
            [],
         double *tzero, int *npts, float *delt )
        {
        /* Station */
        char    KSTNM[8]="          \0";
        strncpy(KSTNM, hd.KSTNM,7);
        fprintf(stderr,"\n Station Name: %s \n",KSTNM);
```

```c
fprintf(stderr," Station Latitude = %f degrees (+
    North)\n",hd.STLA);
fprintf(stderr," Station Longitude= %f degrees (+
    East)\n",hd.STLO);
fprintf(stderr," Station Elevation= %f meters\n",hd.
    STEL);


char    KCMPNM[8]="           \0";
strncpy(KCMPNM,hd.KCMPNM,7);
fprintf(stderr," Component: %s \n",KCMPNM);
/* copy time date to array */
fprintf(stderr,"\n Year=%d ",NZDTTM[0]);
fprintf(stderr," Julian Day=%d ",NZDTTM[1]);
fprintf(stderr,"\n Hour=%d ",NZDTTM[2]);
fprintf(stderr," Minute=%d ",NZDTTM[3]);
fprintf(stderr," Second=%d ",NZDTTM[4]);
fprintf(stderr," mSecond=%d ",NZDTTM[5]);


//... Calculate tzero
    *tzero= (double)(NZDTTM[1]*86400 + NZDTTM[2]*3600
     + NZDTTM[3]*60 + NZDTTM[4] + ((double) NZDTTM[5])
        /1000.0 );
    fprintf(stderr,"\n
        =====================================\n");
```

```
fprintf(stderr,"\n absolute tzero=  %.3f seconds  \n
    ",*tzero);


//... Calculate absolute end time
    double   etime;
    etime=*tzero+hd.E;
    fprintf(stderr," absolute endtime %.3f seconds \n",
        etime);
    fprintf(stderr
        ,"=======================================\n");



    /* Echo Check of key header values */
    *npts=hd.NPTS;
    fprintf(stderr,"\n npts=%d \n",*npts);
    *delt=hd.DELTA;
    fprintf(stderr," delt=%f \n",*delt);
    fprintf(stderr," Begin Time: %.3f seconds\n",hd.B);
    fprintf(stderr," End   Time: %.3f seconds\n",hd.E);

    /* Echo type of data */
    switch (hd.IFTYPE) {
        case 1:
```

```
                  fprintf(stderr ," Type: time series data \n")
                      ;
                  break ;
                  case 2:
                  fprintf(stderr ," Type: real and imag
                      spectrum \n");
                  break ;
                  case 3:
                  fprintf(stderr ," Type: spectral file ,
                      amplitude and phase \n");
                  break ;
                  case 4:
                  fprintf(stderr ," Type: general x vs y data \
                      n");
                  break ;
                  case 5:
                  fprintf(stderr ," Type: general x,y,z (3−D)
                      file \n");
                  break ;
          } // endswitch

       fprintf(stderr ," SCALE=%f   \n",hd.SCALE);

       switch (hd.IDEP) {
```

```
            case 5:
            fprintf(stderr," X-data Type: UNKNOWN \n");
            break;
            case 6:
            fprintf(stderr," X-data Type: displacement (
               nm) \n");
            break;
            case 7:
            fprintf(stderr," X-data Type: velocity (nm/
               sec) \n");
            break;
            case 50:
            fprintf(stderr," X-data Type: velocity
               signal in volts \n");
            break;
            case 8:
            fprintf(stderr," X-data Type: accel (nm/sec/
               sec) \n");
            break;
    } // endswitch

    if (hd.DEPMEN != -12345.0) {
            fprintf(stderr," Data Maximum = %f \n",hd.
               DEPMAX);
```

```
        fprintf ( stderr ," Data Mean     = %f \n",hd.
           DEPMEN) ;
        fprintf ( stderr ," Data Minimum = %f \n",hd.
           DEPMIN) ;
} // endif

if  (hd.IZTYPE !=  −12345.0) {
switch  (hd.IZTYPE) {
        case  5:
        fprintf ( stderr ," Reference Time : UNKNOWN \n
           ") ;
        break ;
        case  9:
        fprintf ( stderr ," Reference Time : Zero time
           is  start  of  file  \n") ;
        break ;
        case  10:
        fprintf ( stderr ," Reference Time : Zero time
           is  0000  of GMT  day  \n") ;
        break ;
        case  11:
        fprintf ( stderr ," Reference Time : Zero time
           is  event  origin  \n") ;
        break ;
```

```
                    case 12:
                    fprintf(stderr ," Reference Time: Zero time
                        is 1st arrival \n");
                    break;
                    default:
                    fprintf(stderr ," Reference Time:  IZTYPE=%d
                        \n",
                            hd.IZTYPE);
        } //endswitch
        } //endif
        } // end DisplayHeader


//========GetSacData FUNCTION
   =========================================
        int      GetSacData(int npts, char infile[], float *
            s1)
        {
        FILE     *h1;


//...open input file
        if( (h1=fopen(infile ,"rb"))==NULL )
        {
            fprintf(stderr ,"\n ABORT !! Error opening file %s
                ",infile);
```

```
            return(1);
        }


//move to end of header
        fseek(h1,158*4,SEEK_SET);


//read data
        if (fread(s1,sizeof(float),npts,h1)!=npts)
        { puts("error reading data");
                return(1);
        }
//close input data file
        fclose(h1);
        return(0);
        } //end GetSacData
```

B.1.2   sub4.h

```
/* |——————————————————————————————————————|
   | sub4.h                                 |
   | author: P. Michaels                    |
   | dated: 21 May 1997                     |
   |                                        |
   | purpose: include file for subroutine library |
   |          sub4.lib                      |
```

```
        |                                                          |
        |——————————————————————————————————————————————————————————|
*/


//... function prototypes


                // BASIC BSEGY FUNCTIONS


// Print Copyright label, time, date to screen
        void cr_labl(char pid[]);


// Print Help Panel to screen
        void hlp_labl(char msg1[],  char  msg2[]);


// Print Copyright label, time, date to file h3
        void fcr_labl(char pid[], FILE *h3);


// Generate output and listing file names
        void names(char infile[], char pid[], char ofile[],
          char suffix[]);


// Count traces, return number samples, interval
        int in_chk(int *ntraces, int *npts,
          float *fsamin, struct header hd, FILE *h1);
```

```
// Read a bsegy header+trace from file h1
    int c_bsegin(int jrec, float s1[], int npts,
        struct header *hd, FILE *h1);


// Write a bsegy header+trace to file h2
    int c_bsegout(int jrec, float s1[], int npts,
        struct header *hd, FILE *h2);


// Compute exbar parameters and draw progress bar
    void bargrid(int ntraces, float *bar, int *ibar, int *
        invbar);


// Write a progress bar step for jrec into ntraces
    void exbar(int ntraces, float bar, int ibar, int
        invbar, int jrec);


/*  Compute x,y,z coordinates and range of shot and
    receivers
                        *sx       =pointer shot x-coord.
                        *sy       =pointer shot y-coord.
                        *sz       =pointer shot z-coord.
                        *sd       =pointer shot depth
                        *rx       =pointer receiver x-coord.
```

```
                    *ry         =pointer receiver y-coord.

                    *rz         =pointer receiver z-coord.

                    *range      =pointer range from shot to
                        receiver

                    hd          =segy header structure

*/


    void findxyz(float *sx, float *sy, float *sz, float *
        sd,
        float *rx, float *ry, float *rz, float *range,
        struct header hd);
```

// SIGNAL PROCESSING FUNCTIONS

```
/* Boxcar smoothing
                        npts        =number of samples
                        iwide       =width of box in samples
                        ntimes      =number of times to apply
                        isign   -1=box car smoother (sinc^ntimes
                            )
                            0=high-pass 1-sinc^ntimes
                            +1=high-pass (1-sinc)^ntimes
```

```
                         s1         =input trace array

                         s2         =output trace array

    returns: 0=success    1=abort, can't allocate memory

                            (or invalid parameters)

*/

      int c_boxit(int npts,int iwide,int isign,int ntimes,

          float s1[], float s2[]);



/* Find max and min values

                         npts        =number of samples

                         s1          =input trace

                         isw1      0=max and min of s1

                                   1=max and min of absolute_value

                                      (s1)

                         *fmax      =pointer to max value

                         *fmin      =pointer to min value

                         *kmx       =pointer to max value index

                         *kmn       =pointer to min value index

*/

      void fmax_min(int npts, float s1[],

          float *fmax, float *fmin, int isw1,  int *kmx,  int

             *kmn);



/* Single Pole smoother ARMA operator
```

```
                              npts      =number of samples
                              s1        =input trace array
                              fsamin    =sample interval (s)
                              tcond     =decay constant (amp=>1/e) (s)
                              s2        =output trace array
      returns: 0=success    1=invalid parameters
*/
      int onepole(int npts, float fsamin, float tcond,
          float s1[], float s2[]);


/* Normalization Function
                              npts      =number of samples
                              s1        =input trace array
                            *fscal    0=normalize by L2 norm
                                      =scale factor 1/(*fscal)
                              s2        =output trace array
      returns: 0=success    1=invalid parameters
*/
      int  fnorm(int npts, float *fscal, float s1[], float
          s2[]);



/*  Compute Mean Absolute Value
                              npts      =number of samples
```

```
                   s1          =input  trace  array
   returns MAV as  float  value
*/
     float mav(int  npts , float  s1 [ ]) ;


/*  Least  Squares  Linear  or  Exponential  Fit  to  x,y data
    isw1    0=Linear  1=Log_e  for  y−variable
*/
     void  lsqufit(float  x[] ,  float  y[] ,  int  npts ,  int  isw1 ,
         float  *slope ,  float  *intrcpt ,  float  *lsqe ,  float  *
           r2 ) ;
```

B.1.3 c_bsegy.h

```
//... segy  header
//    $Id: c_bsegy.h,v  1.4  2012−03−07  00:07:57  pm Exp $
//NOTE:  long  int  (8  bytes  64 bit ,  4  bytes  32 bit )  int  (4  bytes
    both  64 & 32 bit )
     struct  header
    {
          int     tsnl ;    // Trace  sequence  number  within
             line  01−−04
          int     tsnt ;    // Trace  sequence  number  within
             reel  05−−08
```

```
        int     ofrn;      // Original Field Record Number
              09--12
        int     tnofr;     // Trace number original fld
          record   13--16
        int     espn;      // Energy Source point number
                17--20
        int     cdp;       // CDP ensemble number
                 21--24
        int     tncdp;     // Trace number within CDP
          ensemble   25--28
short int     tic;       // Trace ID code (1 = seismic
    data)   29--30
short int     nvs;       // Vertical stack fold
                31--32
short int     nhs;       // Horizontal stack fold
              33--34
short int     duse;      // Data use (1=production; 2=
    test)    35--36
        int     idist;     // Shotpoint-Receiver distance (
          m)     37--40
        int     irel;      // Station elevation (m)
                 41--44
        int     ishe;      // Surface elevation of source (
          m)     45--48
```

```
        int      ishd;      // Source  depth  (m)
                                49--52
        int      delr;      // Datum  elevation  of  receiver
            53--56
        int      dels;      // Datum  elevation  of  source
                57--60
        int      wds;       // Water  depth  of  source
                    61--64
        int      wdr;       // Water  depth  of  receiver
                65--68
short int     smul1;     // Scalar  multr/divr , bytes  41-68
        69--70
short int     smul2;     // Scalar  multr/divr , bytes  73-88
        71--72
        int      ishlo;    // Shotpoint  coordinate - x (
            East +)   73--76
        int      ishla;    // Shotpoint  coordinate - y (
            North +)  77--80
        int      irlo;     // Receiver  coordinate - x  (
            East +)   81--84
        int      irla;     // Receiver  coordinate - y  (
            North +)  85--88
short int     cunits;    // Coord.  Units  (1=m  2=sec.  of
      arc )   89--90
```

```
short  int      wvel;      // Weathering  velocity  (m/s)
              91--92
short  int      swvel;     // Subweathering  velocity  (m/s)
          93--94
short  int      utimes;    // Uphole  time  at  source
                  95--96
short  int      utimeg;    // Uphole  time  at  group
                    97--98
short  int      sstati;    // Source  static  correction  (ms)
            99-100
short  int      gstati;    // Group  static
                      101-102
short  int      tstati;    // Total  static
                      103-104
short  int      istime;    // Lag  time  A
                      105-106
short  int      ibtime;    // Lag  time  B
                      107-108
short  int      ictime;    // Delay  recording  time  (msec)
              109-110
short  int      mtimes;    // Mute  start  time
                  111-112
short  int      mtimee;    // Mute  end  time
                  113-114
```

```
short int      length;    // Number of samples in this
    trace   115−116
short int      isi;       // Sample interval (microsec.)
         117−118
short int      gaint;     // Gain (1=fixed 2=binary 3=
    float)   119−120
short int      gc;        // Gain constant
                          121−122
short int      gidb;      // Instrument or initial gain in
    DB  123−124
short int      tcorr;     // Correlated (1=no; 2=yes)
              125−126
short int      tsswee;    // Start sweep frequency (hz)
           127−128
short int      teswee;    // End sweep frequency (hz)
              129−130
short int      tsleng;    // Sweep length in milliseconds
           131−132
short int      tstype;    // Sweep type
                          133−134
short int      tstts;     // Sweep trace taper(msec) at
    start  135−136
short int      tstte;     // Sweep trace taper(msec) at
    end    137−138
```

```
short  int      tttype ;   // Taper  type
                     139−140
short  int     aif ;     // Antialias  filter  frequency
          141−142
short  int     ais ;     // Antialias  filter  slope
               143−144
short  int     nif ;     // Notch  filter  frequency
               145−146
short  int     nis ;     // Notch  filter  slope
                147−148
short  int     flc ;     // Low  cut  frequency
                  149−150
short  int     fhc ;     // High  cut  frequency
                151−152
short  int     slc ;     // Low  cut  slope
                    153−154
short  int     shc ;     // High  cut  slope
                  155−156
short  int     tyear ;   // Shot  Time − year
                  157−158
short  int     tday ;    // Shot  Time − day
                 159−160
short  int     thour ;   // Shot  Time − hour (24  hour
     clock )  161−162
```

```
short int      tmin;      // Shot Time − minute of hour
          163−164
short int      tsec;      // Shot Time − second of minute
          165−166
short int      tbcode;    // Time basis code (2 = GMT)
            167−168
short int      twf;       // Trace weigthing factor
                169−170
short int      ggrp1;     // Geoph group roll switch first
      pos171−172
short int      ggtp;      // Geoph group trc pos 1 on fld
      rec 173−174
short int      errlt;     // Time code translator error
      light 175−176
short int      daca;      // Distance−azimuth calc
      algorithm   177−178
short int      edc;       // Earth dimension code
                179−180
    int      mst;         // (BSEGY First Break Picks)
              181−184
short int      shtazi;    // Source pol azi (deg) East=90
      deg   185−186
short int      charge;    // Charge size in grams
                187−188
```

```
        short int      shtver;   // Source pol angle (deg.)Down
            =180    189−190
        short int      sday;     // Shot time − Day
                           191−192
        short int      shour;    // Shot time − Hour
                         193−194
        short int      smin;     // Shot time − Minute
                       195−196
        short int      sseco;    // Shot time − Sec
                         197−198
//               int      ssmic;   // Shot time − Microsecond
            199−202
        short int      ssmic1;   //long integer splits across
            alignment
        short int      ssmic2;   //use 2 short integers
        short int      azimut;   // Azimuth of rec from shot (deg
            )    203−204
        short int      geoazi;   // Geoph pol Azm of axis (deg) E
            =90  205−206
        short int      geover;   // Geoph pol Vert. deg. Down=180
               207−208
            int      ttrace;   // add to get trace start time
                   209−212
```

```c
    char        scrs[4];     // instrument name
                    213-216
    char        deploy[4];   // Deployment name
                    217-220
    char        spname[4];   // Shotpont name
                    221-224
    char        rstnam[4];   // Receiver site name
                    225-228
    char        shotid[4];   // Shot name
                    229-232
    char        lineid[4];   // Line name
                    233-236
    char        geoor[4];    // Geophone orientation
                    237-240
    };
/*

|--------------------------------------------------------


| Note:   bytes 181-240 are unassigned or for optional use
   as defined|
|       by Barry et al. (1975)-"Digital Tape Standards"-
   SEG           |

|--------------------------------------------------------
```

```
*/
```

B.1.4   bsac.h

```
//... bsac header (SAC Compatible Header Structure for BSAC)
//    $Id: bsac.h,v 1.5 2012-03-07 00:07:57 pm Exp $
//   P. Michaels <cgiss.boisestate.edu>
//   Based on the Seismic Analysis Code Data Format published
   by IRIS
//       usually independent variable is time, dependent a
   seismic signal
//       REF: http://www.iris.edu/software/sac
         struct   header
         {
         float    DELTA;   //sample interval (required)
         float    DEPMIN;  //min value dependent variable
         float    DEPMAX;  //max value dependent variable
         float    SCALE;   //UNUSED scale factor to multiply
            dep. var.
         float    ODELTA;  //observed sample interval if
            different from delta
         float    B;       //begining value of indep. variable
            (required)
         float    E;       //ending value of indep. variable (
            required)
```

```
float    O;        // event origin time seconds (sec.)
    wrt reference time
float    A;        // first arrival time seconds (sec.)
    wrt reference time
float    I009;     // internal sac word 09
float    T0;       // user defined time pick (sec.) wrt
    reference time
float    T1;       // user defined time pick (sec.) wrt
    reference time
float    T2;       // user defined time pick (sec.) wrt
    reference time
float    T3;       // user defined time pick (sec.) wrt
    reference time
float    T4;       // user defined time pick (sec.) wrt
    reference time
float    T5;       // user defined time pick (sec.) wrt
    reference time
float    T6;       // user defined time pick (sec.) wrt
    reference time
float    T7;       // user defined time pick (sec.) wrt
    reference time
float    T8;       // user defined time pick (sec.) wrt
    reference time
```

```
float     T9;        // user defined time pick (sec.) wrt
    reference time
float     F;          // fini (end of event time (sec.) wrt
    reference time
float     RESP0;   // instrument response parameter (not
    currently used)
float     RESP1;   // instrument response parameter (not
    currently used)
float     RESP2;   // instrument response parameter (not
    currently used)
float     RESP3;   // instrument response parameter (not
    currently used)
float     RESP4;   // instrument response parameter (not
    currently used)
float     RESP5;   // instrument response parameter (not
    currently used)
float     RESP6;   // instrument response parameter (not
    currently used)
float     RESP7;   // instrument response parameter (not
    currently used)
float     RESP8;   // instrument response parameter (not
    currently used)
float     RESP9;   // instrument response parameter (not
    currently used)
```

```
float    STLA;    // station latitude (degrees,
    positive north)
float    STLO;    // station longitude (degrees east
    positive)
float    STEL;    // station elevation (meters)
float    STDP;    // station depth below surface (
    meters)
float    EVLA;    // event latitude (degrees, positive
    north)
float    EVLO;    // event longitude (degrees, positive
    east)
float    EVEL;    // event elevation (meters)
float    EVDP;    // event depth (meters)
float    MAG;     // event magnitude
float    USER0;   // user defined storage area
float    USER1;   // user defined storage area
float    USER2;   // user defined storage area
float    USER3;   // user defined storage area
float    USER4;   // user defined storage area
float    USER5;   // user defined storage area
float    USER6;   // user defined storage area
float    USER7;   // user defined storage area
float    USER8;   // user defined storage area
float    USER9;   // user defined storage area
```

```
float     DIST;     // station to event distance KM

float     AZ;       // event to station azimuth (degrees)

float     BAZ;      // station to event azimuth (degrees)

float     GCARC;    // station to event great circle arc
     length (degrees)

float     I054;     // internal, word 54

float     I055;     // internal, word 55

float     DEPMEN;   // mean value of dependent variable

float     CMPAZ;    // component azimuth (degrees
     clockwise from north)

float     CMPINC;   // component inclination (degrees
     from vertical)

float     XMINIMUM;         // minimum x-value spectral
     files only

float     XMAXIMUM;         // maximum x-value spectral
     files only

float     YMINIMUM;         // minimum y-value spectral
     files only

float     YMAXIMUM;         // maximum y-value spectral
     files only

float     U063;     // unused word 63

float     U064;     // unused word 64

float     U065;     // unused word 65

float     U066;     // unused word 66
```

```
float      U067;    //unused word 67
float      U068;    //unused word 68
float      U069;    //unused word 69
int        NZYEAR; //GMT year corresponding to
   reference (zero time) in file
int        NZJDAY; //GMT julian day
int        NZHOUR; //GMT julian hour
int        NZMIN;   //GMT minute
int        NZSEC;   //GMT second
int        NZMSEC; //GMT millisecond
int        NVHDR;
int        NORID;
int        NEVID;
int        NPTS;     //number of samples in signal (
   required)
int        I080;     //internal, word 80
int        NWFID;
int        NXSIZE;
int        NYSIZE;
int        U084;     //unused word 84
int        IFTYPE; //Type of file (required) (see inc/
   SacHeader.h or extfunc.h)
           /* 1 = ITIME = time series data
              2 = IRLIM = real & imag spectrum
```

```
                    3 = IAMPH = spectral file , amplitude and
                        phase
                    4 = IXY   = general x vs y data
                    5 = IXYZ  = general x , y , z (3−D) file
                */
int       IDEP;    // Type of independent variable
          /* 5 = IUNKN = x data type unknown
                    6 = IDISP = x data displacement data (nm)
                    7 = IVEL  = x data velocity in nm/second
                   50 = IVOLTS =x data velocity signal in
                        volts
                    8 = IACC  = x data accel (nm/sec /sec )
                */
int       IZTYPE; // reference time equivalence :
          /* 5 = IUNKN = unknown
                    9 = IB = zero time : start of file
                   10 = IDAY = zero time : 0000 of GMT day
                   11 = IO  = zero time : event origin
                   12 = IA  = zero time : 1st arrival
                13−22 = ITn = user defined pick time n , n
                      = 0..9
                        IT0 =13 , IT1 =14 , ... . IT9 =22
                */
int       U088 ;
```

```
int        IINST;   // type of recording instrument (not
      currently used)
int        ISTREG; // station geographic region (not
      used)
int        IEVREG; // event geographic region (not used)
int        IEVTYP; // event type
           /*  5 = IUNKN  = unknown
             37 = INUCL  = nuclear event
             38 = IPREN  = nuke pre-shot
             39 = IPOSTN = nuclear post-shot event
             40 = IQUAKE = earthquake
             41 = IPREQ  = foreshock
             42 = IPOSTQ = aftershock
             43 = ICHEM  = chemical explosion
             72 = IQB    = quarry or mine blast confirmed
                by quarry
             73 = IQB1   = quarry or mine blast with
                designed
                         shot info=ripple fired
             74 = IQB2   = quarry or mine blast with
                         observed shot info-ripple
                              fired
             75 = IQBX   = quarry or mine blast , single
                shot
```

```
        80 = IME  = marine explosion
        77 = IEQ  = earthquake
        78 = IEQ1 = earthquakes in a swarm or
             aftershock sequence
        79 = IEQ2 = felt earthquake
        44 = IOTHER = other
      */
int     IQUAL;
int     ISYNTH;
int     IMAGTYP;
int     IMAGSRC;
int     U097;    // unused word 97
int     U098;    // unused word 98
int     U099;    // unused word 99
int     U100;    // unused word 100
int     U101;    // unused word 101
int     U102;    // unused word 102
int     U103;    // unused word 103
int     U104;    // unused word 104
int     LEVEN;   // REQUIRED even spaced data =
   logical true = 1
int     LPSPOL; // logical 1=true 0=false
int     LOVROK; // logical 1=true 0=false
int     LCALDA; // logical 1=true 0=false
```

```
int      U109;      //unused word 109

char     KSTNM[8];           // station name (8char=2
   words)

char     KEVNM[16];          // event name

char     KHOLE[8];

char     KO[8];

char     KA[8];

char     KT0[8];

char     KT1[8];

char     KT2[8];

char     KT3[8];

char     KT4[8];

char     KT5[8];

char     KT6[8];

char     KT7[8];

char     KT8[8];

char     KT9[8];

char     KF[8];

char     KUSER0[8];

char     KUSER1[8];

char     KUSER2[8];

char     KCMPNM[8];

char     KNETWK[8];

char     KDATRD[8];
```

```
char        KINST[8];
};
```

B.2   Matlab Codes

All Matlab codes used in the analysis of data and generation of figures for this dissertation
were written in Matlab 7.10 R2010a. The computer used was a Dell Optiplex 960 with the
Ubuntu Lucid Lynx version 10.04 LTS operating system.

The file datasetup4.m includes the commands to call all other functions from the direc-
tory AllCode3. It includes some plotting commands, as well.

```
%datasetup2.m differs from datasetup.m in that it passes
    narrow
%bandpassed signals into the PCA algorithm.  Each signal is
     filtered using
%the same coefficients as what was used to determine the
    time shift and
%true velocity.


%This script prepares data for entry into the PCA function.


clear all




%load the data of interest:  E,N, and Z channels for station
```

1, station 2

```
load GGE;

load GGN;

load GGZ;

load LVE;

load LVN;

load LVZ;

load RBE;

load RBN;

load RBZ;

load UTE;

load UTN;

load UTZ;

load GGLVRBUT;
```

```
%store each station as a variable:  data1, data2
% data1E=GGE;

% data1N=GGN;

% data1Z=GGZ;

% data2E=LVE;

% data2N=LVN;

% data2Z=LVZ;
```

```
staUTM=GGLVRBUT;
GGEt=GGE(:,1);
dt=GGEt(2)−GGEt(1);
npts=length(GGEt);
df=1/(npts*dt);




%enter the position in UTM of station 1, station 2
UTMGG=[staUTM(1,2),staUTM(1,3)];  %GG(Easting, Northing)
UTMLV=[staUTM(2,2),staUTM(2,3)];  %LV(Easting, Northing)
UTMRB=[staUTM(3,2),staUTM(3,3)];  %RB(Easting, Northing)
UTMUT=[staUTM(4,2),staUTM(4,3)];  %UT(Easting, Northing)

%User choose one station as UTM1, second station as UTM2
%UTM1=input('choose UTMGG, UTMLV, UTMRB, or UTMUT for
    station 1')
%UTM2=input('choose UTMGG, UTMLV, UTMRB, or UTMUT for
    station 2')
UTM1=UTMLV;
UTM2=UTMUT;

if (UTM1==UTMGG)
    data1z=GGZ(:,2);
```

```
        data1e=GGE(: ,2) ;

        data1n=GGN(: ,2) ;

elseif (UTM1==UTMLV)

        data1z=LVZ(: ,2) ;

        data1e=LVE(: ,2) ;

        data1n=LVN(: ,2) ;

elseif (UTM1==UTMRB)

        data1z=RBZ(: ,2) ;

        data1e=RBE(: ,2) ;

        data1n=RBN(: ,2) ;

elseif (UTM1==UTMUT)

        data1z=UTZ(: ,2) ;

        data1e=UTE(: ,2) ;

        data1n=UTN(: ,2) ;

end

if (UTM2==UTMGG)

        data2z=GGZ(: ,2) ;

        data2e=GGE(: ,2) ;

        data2n=GGN(: ,2) ;

elseif (UTM2==UTMLV)

        data2z=LVZ(: ,2) ;

        data2e=LVE(: ,2) ;

        data2n=LVN(: ,2) ;

elseif (UTM2==UTMRB)
```

```
        data2z=RBZ(: ,2);

        data2e=RBE(: ,2);

        data2n=RBN(: ,2);

elseif (UTM2==UTMUT)

        data2z=UTZ(: ,2);

        data2e=UTE(: ,2);

        data2n=UTN(: ,2);

end


t=GGEt;

zdata1=data1z ';

edata1=data1e ';

ndata1=data1n ';

zdata2=data2z ';

edata2=data2e ';

ndata2=data2n ';




fr1 =0: npts /2 −1;

fr1=fr1 /( npts ∗ dt );

Z1=fft ( zdata1 −mean( zdata1 ));

% figure

% plot ( fr1 , abs ( Z1 ( 1 : npts /2 ) ) )
```

```
%calculate the distance between the two stations
[delr,aa,bb]=delr(UTM1,UTM2);


%set up cascade of narrowband filters
%fbwcf=linspace(0.01,1,50);  %looks the best so far
fbwcf=linspace(0.01,1,50);
%flowest=fbwcf(1);
fdiff=0.001;
for j=1:length(fbwcf);
flo(j)=fbwcf(j)-fdiff;
fhi(j)=fbwcf(j)+fdiff;
[adv,win,b,a] = bpfilt3(dt,flo(j),fhi(j));
% zdata1f=filter(b,a,zdata1);
% Z1f=fft(zdata1f-mean(zdata1f));
% figure
% plot(fr1,abs(Z1f(1:npts/2)))



%set initial conditions
j1=1;  %start at position 1
```

```
j2=j1+(win−1);   %last point = end of window
%  t1 =1;
%  t2=t1+(win−1);
 counter =0;


%call PCA function
[dmfXE1 ,dmfXN1 , evecs1 , evecs2 , evals1 , evals2 , ratio1 , ratio2 , xx1
    , xx2 , yy1 , . . .
     yy2 , azim1 , azim2 , npts ] = pca2staGen (win , edata1 , edata2 ,
         ndata1 , ndata2 , . . .
     j1 , j2 , counter , adv , b , a ) ;



%analyze the resulting ratios to find good and bad time
    windows
[ goodtradeg1 , goodtradeg2 , badtradeg1 , badtradeg2 , goodaz1deg ,
    goodaz2deg , . . .
     goodtime1 , goodtime2 , maxratio1 , maxratio2 ]= r a t i o a n a l y s i s (
         azim1 , azim2 , npts , adv , win , ratio1 , . . .
     ratio2 ,GGEt) ;



% commented to reduce # of figures generated
%  [ fEd1 , fNd1 , timestamp1 , gwst1 , gr1 , fi ]= p a r t i c l e m o t i o n (
```

```
        goodtradeg1 ,...
%        GGEt, edata1 , ndata1 , zdata1 , flo ,b,a, win ) ;
%


%calculate  cross  correlations  for  good  windows  at  both
    stations


[numcomm12(j) , goodindx12 , st12 ,...
      gstarttime12]=plotxcorr_gwin12_2(goodtime1 , goodtime2 ,...
    GGEt, win , edata1 , edata2 , zdata1 , zdata2 ) ;
st12c{j}=st12 ;
goodindxc{j}=goodindx12 ;


% %find  the  value  of  deltax  and  alpha  for  each  good  azimuth :
      NEEDS  FIXING ,
%  basically  want  to  do  this  loop  for  all  elements  of
    goodindxc{j}
[ deltax(j) , alpha(j) , goodaz1s , goodaz2s]=stationgeom3(delr , aa ,
    bb , goodaz1deg ,...
      goodaz2deg , numcomm12(j) , goodtradeg1 , goodtradeg2 ,
        gstarttime12 , goodindx12 ) ;
```

```
goodaz1sc{j}=goodaz1s;
goodaz2sc{j}=goodaz2s;




% %find the true velocities as a function of frequency
% [vtruegeom(j),deltat(j)]=realdatatimeshift5(goodindx12,...
%       zdata1,zdata2,delr,deltax(j),...
%       dt,alpha(j),fbwcf(j),b,a,win);


end




%note that goodindxc{1,16} is a vector containing the start
    indices of good
%windows for the frequency 0.3131 Hz.  This is a
    particularly well
%populated entry in goodindxc, with 274 good windows.
%good3131=goodindxc{1,16};
good3131=goodindxc{1,47};



[ff]=method2(good3131,win,alpha,delr,npts,df,dt,zdata1,
    zdata2);
```

The following functions were written over the period from May 2011-September 2012.

B.2.1   delr.m

```
function [ delr , aa , bb]= delr (UTM1,UTM2)
```

```
%This function performs a velocity
%correction on the apparent velocity ,
%based on the station geometry and velocity
%model from the USGS for Akutan .
%The output is velocity , corrected for
%the angle of approach , and the
%cross−correlation of the two stations .
```

```
%dt=time sample interval
%t = timeseries from stations
%delx = distance between stations
%azim = angle of approaching energy
%zdata1 = z component of the first data recording
%zdata2 = z component of the second data recording
%vtrue = true velocity , corrected for azimuthal arrival
%XC12 = cross correlation of station 1,2
```

```
%calculate the distance between the two stations of choice
%UTM1, UTM2 are of the format (Easting, Northing)


aa=UTM1(1,1)−UTM2(1,1);  %difference between Easting
bb=UTM1(1,2)−UTM2(1,2);  %difference between Northing
cc2=(aa^2)+(bb^2);       %Pythagorean Thm to
%find distance between 1 and 2
 delr=sqrt(cc2);         %dist between station 1 and 2 in meters
```

B.2.2 bpfilt3.m

```
 function [adv,win,b,a] = bpfilt3(dt,flo,fhi)
%{input the time portion of the (data,time)
%format file to calculate the
%time interval, dt, and choose a lower cutoff frequency
%for the Butterworth
%bandpass filter. NOTE: since the time vectors are
%trimmed to match the
%shortest vector in the body of the main code, GGEt
%is the same as
%GGNt,GGZt,LVEt,LVNt, and LVZt. The output from this
%function consists of
%b and a, respectively the numerator and denominator
%coefficients of the
%filter. The filter may then be applied to each window
```

```
%in the PCA algorithm.  win, the third output, is the
%window length determined from
%the period of the lowcut portion of the filter and the dt.
%}


%bpfilt3.m differs from bpfilt2.m in that the lowest
%frequency in fbwcf is no longer used for defining
%the window length.  A default window length is
%used instead.
 flowest=0.01;
T=1/flowest;  %period (s) of the lowest f in fbwcf
nsamp=ceil(T/dt);  %# of samples in 1 wavelength;
%ceil to get integer value
win=nsamp*1;  %set window length to # of sample points x2
%win=nsamp*5;  %set window length to # of sample points x5
 adv=200;  %advance the window by 500 samples each time
 fs=1/dt;  %sampling frequency
fnyq=(fs/2)*2*pi;  %nyquist frequency in rad/s
%fhi=flo*2;  %look at a passband spanning one octave
wlo=(flo*2*pi)/fnyq;  %angular freq normalized by nyquist
whi=(fhi*2*pi)/fnyq;  %angular freq normalized by nyquist
Wn=[wlo,whi];  %define passband
[b,a]=butter(4,Wn);
%[b,a]=butter(2,Wn);
```

B.2.3 pca2staGen.m

```
function [dmfXE1,dmfXN1,evecs1,evecs2,evals1,evals2,ratio1
   ,...
    ratio2,xx1,xx2,yy1,yy2,azim1,azim2,npts] = pca2staGen(
        win,...
    edata1, edata2,ndata1,ndata2,j1,j2,counter,adv,b,a)


npts=length(edata1);
dmXE1=edata1(j1:j2)-mean(edata1(j1:j2));  %demean each
   window
dmXE2=edata2(j1:j2)-mean(edata2(j1:j2));
dmXN1=ndata1(j1:j2)-mean(ndata1(j1:j2));
dmXN2=ndata2(j1:j2)-mean(ndata2(j1:j2));
dmX1=[dmXE1;dmXN1];
dmX2=[dmXE2;dmXN2];
Xs1=(1/win)*(dmX1*dmX1');
Xs2=(1/win)*(dmX2*dmX2');
[evecs1, evals1]=eig(Xs1);
[evecs2, evals2]=eig(Xs2);
while (j2<npts-adv)
    counter=counter+1;
   %winnumber(counter)=counter+1;
   j1=j1+adv;
   j2=j1+(win-1);
```

```
%        t1=t1+adv;
%        t2=t2+(win-1);
        dmXE1=edata1(j1:j2)-mean(edata1(j1:j2));
        dmXE2=edata2(j1:j2)-mean(edata2(j1:j2));
        dmXN1=ndata1(j1:j2)-mean(ndata1(j1:j2));
        dmXN2=ndata2(j1:j2)-mean(ndata2(j1:j2));
        dmfXE1=filter(b,a,dmXE1);
        dmfXE2=filter(b,a,dmXE2);
        dmfXN1=filter(b,a,dmXN1);
        dmfXN2=filter(b,a,dmXN2);
        dmfX1=[dmfXE1;dmfXN1];
        dmfX2=[dmfXE2;dmfXN2];
        Xs1=(1/win)*(dmfX1*dmfX1');
        Xs2=(1/win)*(dmfX2*dmfX2');


        %find the eigenvectors and eigenvalues of Xs
        [evecs1,evals1]=eig(Xs1);
        [evecs2,evals2]=eig(Xs2);
        [maxeval1,index1]=max(max(evals1));
        [maxeval2,index2]=max(max(evals2));
        [mineval1,index21]=min(diag(evals1));
        [mineval2,index22]=min(diag(evals2));


        %determine the ratio of each window's max:min eigenvals
```

```matlab
        ratio1(counter)=maxeval1/mineval1;
        ratio2(counter)=maxeval2/mineval2;
        mj1=evecs1(:,index1)*15000;
        mj2=evecs2(:,index2)*15000;
        xx1(counter)=mj1(1);
        xx2(counter)=mj2(1);
        yy1(counter)=mj1(2);
        yy2(counter)=mj2(2);
        %8/2/12 update:  azim is now azim(counter)
       %find azim of each window
       azim1(counter)=(pi/2)-atan2(mj1(1),mj1(2));
       azim2(counter)=(pi/2)-atan2(mj2(1),mj2(2));
       if (azim1(counter)<0)
           azim1(counter)=360+azim1(counter);
       end
       if(azim2(counter)<0)
           azim2(counter)=360+azim2(counter);
       end
end
```

B.2.4   ratioanalysis.m

```matlab
function [goodtradeg1,goodtradeg2,badtradeg1,...
   badtradeg2,goodaz1deg,goodaz2deg,goodtime1,...
   goodtime2,maxratio1,maxratio2]=ratioanalysis(azim1,...
```

```
   azim2 , npts , adv , win , ratio1 , ratio2 ,GGEt)


%this  function  identifies  the  ratios  above  and  below  the
%cutoff  ratio  and  creates  files  with  matrices  of  the
%form  [col1=start  time  of  a  window ,  col2
%=  ratio  for  that  window ,  col3  =  azimuth  of  that  window]


 maxratio1=max( ratio1 );
 maxratio2=max( ratio2 );
%minratio1=min( ratio1 );
%minratio2=min( ratio2 );


%user  chooses  cutoff  parameter
%cutopt=input ('Choose  cutoff  paramter :   1  =  60%  of  max ,
%2  =  1.5*1SD  above  mean ')
 cutopt =1;
 if  ( cutopt ==1)
     cutoff1 =0.2* maxratio1 ;
     cutoff2 =0.2* maxratio2 ;
 elseif ( cutopt ==2)  %originally  1.5 ,  not  1.2
     cutoff1 =1.2* std ( ratio1 )+mean( ratio1 );
     cutoff2 =1.2* std ( ratio2 )+mean( ratio2 );
 end
```

```
b1=find(ratio1<cutoff1);  %indices where ratio1 < cutoff
b2=find(ratio2<cutoff2);
g1=find(ratio1>cutoff1);  %indices where ratio1 > cutoff
g2=find(ratio2>cutoff2);
ratiob1=ratio1;  %establish new vector to record bad ratios
ratiob2=ratio2;
ratio1(b1)=0;  %this sets the value of any ratio<cutoff
%(bad ratios) to zero in the new vector
ratio2(b2)=0;
ratiob1(g1)=0;  %this sets the value of any good ratios to
    zero
ratiob2(g2)=0;




%match the timestamps from the nonzero columns of ratio1,
    ratio2
T=GGEt;
i1=1;
i2=i1+(win-1);
counter2=0;
timewin=T(i1:i2);  %first time window (200 points)
time1=timewin(1);  %first time stamp on first 200-pt time
    window
while (i2<npts-adv)
```

```
        counter2=counter2 +1;
        i1=i1+adv ;
        i2=i1 +(win −1);
        timewin=T( i1 : i2 ) ;
        time1 ( counter2 )=timewin (1) ;
 end




%make matrix of good ratios and times
 trg1 =[ ratio1 ; time1 ];
 trg2 =[ ratio2 ; time1 ];


%make matrix of bad ratios and times
 trb1 =[ ratiob1 , time1 ];
 trb2 =[ ratiob2 , time1 ];


%find the indices of nonzero values of ratio vector (good
     ratios )
 indices1=find ( ratio1 );
 indices2=find ( ratio2 );


%same for bad ratios
 indicesb1=find ( ratiob1 );
```

```
indicesb2=find(ratiob2);


%goodratio is all of nonzero samples of ratio
goodratio1=nonzeros(ratio1);
goodratio2=nonzeros(ratio2);


%similarly, badratio
badratio1=nonzeros(ratiob1);
badratio2=nonzeros(ratiob2);


%select start times for the good windows
goodtime1=[time1(indices1)]';
goodtime2=[time1(indices2)]';


%and now bad windows
badtime1=[time1(indicesb1)]';
badtime2=[time1(indicesb2)]';


%find the good azimuths
goodaz1=azim1(indices1)';
goodaz2=azim2(indices2)';
badaz1=azim1(indicesb1)';
badaz2=azim2(indicesb2)';
```

```
%convert azims to degrees
 goodaz1deg=goodaz1 *(180/ pi ) ;
 goodaz2deg=goodaz2 *(180/ pi ) ;
 badaz1deg=badaz1 *(180/ pi ) ;
 badaz2deg=badaz2 *(180/ pi ) ;


%make matrices of good ratio , good time , good azimuth
 goodtradeg1 =[ goodratio1 , goodtime1 , goodaz1deg ];
 goodtradeg2 =[ goodratio2 , goodtime2 , goodaz2deg ];


%make matrices of bad ratio , bad time , bad azimuth
 badtradeg1 =[ badratio1 , badtime1 , badaz1deg ];
 badtradeg2 =[ badratio2 , badtime2 , badaz2deg ];
```

B.2.5 particlemotion.m

```
 function [ fEd1 , fNd1 , timestamp1 , gwst1 , gr1 , fi ]= particlemotion (
    goodtradeg1 ,...
      GGEt, edata1 , ndata1 , zdata1 , flo , b,a, win)
%this function produces fEd1 and fEd1 , which can be plotted
    by the command:
%  plot ( fEd1 , fNd2 , ' gs ' )
%to produce a graph of the particle motion for the first six
     good windows
```

```
%of the PCA for edata1 , ndata1

%lines for plotting are commented out , 9/23/12

zdata1=zdata1 ';
for i = 1:6
    timestamp1(i)=goodtradeg1(i,2); %select time stamp of
        each start time
    gwst1(i)=find(GGEt==timestamp1(i)); %find the index for
         each timestamp
    gr1(i)=goodtradeg1(i,1); %select the ratio of the good
        window
    fi(i)=i+1; %create a figure index
    Ed1=edata1(gwst1(i):gwst1(i)+win)-mean(edata1(gwst1(i):
        gwst1(i)+win));
    Nd1=ndata1(gwst1(i):gwst1(i)+win)-mean(ndata1(gwst1(i):
        gwst1(i)+win));
    Zd1=zdata1(gwst1(i):gwst1(i)+win)-mean(zdata1(gwst1(i):
        gwst1(i)+win));
    fEd1=filter(b,a,Ed1);
    fNd1=filter(b,a,Nd1);
    fZd1=filter(b,a,Zd1);
%     figure(fi(i)) %plot the good and the bad with this
    option OR
```

```
%        %figure  %plot the good and the z with this option
%        plot(fEd1,fNd1,'gs')
%        axis square
%        title({['edata1 v ndata1 for good time window starting
        with ...
%              time ',num2str(timestamp1(i))],['ratio ',
    num2str(gr1(i))],..
%              ['low cut frequency in Hz = ',num2str(flo)]})
%        figure
%        plot(fZd1,fEd1,'cs')
%        axis square
%        title('Particle motion for Z v. East')
%        figure
%        plot(fZd1,fNd1,'cs')
%        axis square
%        title('particle motion for Z v. North')
end
```

B.2.6 plotxcorr_gwin12_2.m

```
function [numcomm12,goodindx12,st12,...
    gstarttime12]=plotxcorr_gwin12_2(goodtime1,goodtime2,...
    GGEt,win,edata1,edata2,zdata1,zdata2)
```

```
%this function identifies good start windows in both the
    first and second
%data series.

%find where there's a good window in both stations
st12=intersect(goodtime1,goodtime2);

%count the number of these good windows over the whole
    series
numcomm12=length(st12);

%the following plot options are commented out to reduce
    runtime
%find index of good time
%
% if (numcomm12>0)
%      for ii=1:numcomm12
%          goodindx12(ii)=find(GGEt==st12(ii));
%          gstarttime12(ii)=GGEt(goodindx12(ii));
%          gstartdata12(ii)=edata1(goodindx12(ii));
%          goodstarts12=[gstarttime12(ii),gstartdata12(ii)];
%store values of the good starting time/data point in a
    matrix for later use
%          gooddata1E=edata1(goodindx12(ii):goodindx12(ii)+
```

```
         win ) ;
%        gooddata2E=edata2 ( goodindx12 ( i i ) : goodindx12 ( i i )+
    win ) ;
%        goodtime12=GGEt ( goodindx12 ( i i ) : goodindx12 ( i i )+win )
    ;
%        gooddata1Z=zdata1 ( goodindx12 ( i i ) : goodindx12 ( i i )+
    win ) ;
%        gooddata2Z=zdata2 ( goodindx12 ( i i ) : goodindx12 ( i i )+
    win ) ;
%        dmgood1E=gooddata2E−mean ( gooddata2E ) ;
%        dmgood2E=gooddata2E−mean ( gooddata2E ) ;
%        [ c12E , lags12E ]= xcorr (dmgood1E , dmgood2E ) ;
%        figure
%        stem ( lags12E , c12E ) ;
%        title ({[ ' xcorr  for  first  good  window ,  ...
%             E  channels  of  1  and  2 ' ] ,[ ' start  time  ' ,
    num2str ( goodtime12 ( i i ) ) ] })
%        xlabel ( ' Lags  in  Number  of  Samples ' )
%        ylabel ( ' Amplitude  ( Not  Normalized ) ' )
%        dmgood1Z=gooddata1Z−mean ( gooddata1Z ) ;
%        dmgood2Z=gooddata2Z−mean ( gooddata2Z ) ;
%        [ c12Z , lags12Z ]= xcorr (dmgood1Z , dmgood2Z ) ;
%        figure
%        stem ( lags12Z , c12Z ) ;
```

```
%           title ({[ ' xcorr  for  first  good  window ,  Z  channels
   of  1  and  2 ' ] ,...
%                  [ ' start  time  ' , num2str ( goodtime12 ( ii ) ) ] } )
%           xlabel ( ' Lags  in  Number  of  Samples ' )
%           ylabel ( ' Amplitude ( Not  Normalized ) ' )
%      end
% else
%      goodindx12 =0;
% end

% %just  do  it  for  the  first  good  window
% if  ( numcomm12 >0)
%      goodindx12 ( 1 )= find ( GGEt== st12 ( 1 ) ) ;
%      gstarttime12 ( 1 )=GGEt( goodindx12 ( 1 ) ) ;
%      gstartdata12 ( 1 )= edata1 ( goodindx12 ( 1 ) ) ;
%      goodstarts12 =[ gstarttime12 ( 1 ) , gstartdata12 ( 1 ) ] ;
%      gooddata1E= edata1 ( goodindx12 ( 1 ) : goodindx12 ( 1 )+win ) ;
%      gooddata2E= edata2 ( goodindx12 ( 1 ) : goodindx12 ( 1 )+win ) ;
%      goodtime12=GGEt( goodindx12 ( 1 ) : goodindx12 ( 1 )+win ) ;
%      gooddata1Z= zdata1 ( goodindx12 ( 1 ) : goodindx12 ( 1 )+win ) ;
%      gooddata2Z= zdata2 ( goodindx12 ( 1 ) : goodindx12 ( 1 )+win ) ;
%      dmgood1E=gooddata2E −mean ( gooddata2E ) ;
%      dmgood2E=gooddata2E −mean ( gooddata2E ) ;
%      [ c12E , lags12E ]= xcorr ( dmgood1E , dmgood2E ) ;
```

```
% %        figure
% %        stem(lags12E,c12E);
% %        title({['xcorr for first good window, E channels of
    1 and 2'],...
%          ['start time ',num2str(goodtime12(1))]})
% %        xlabel('Lags in Number of Samples')
% %        ylabel('Amplitude (Not Normalized)')
%      dmgood1Z=gooddata1Z-mean(gooddata1Z);
%      dmgood2Z=gooddata2Z-mean(gooddata2Z);
%      [c12Z,lags12Z]=xcorr(dmgood1Z,dmgood2Z);
% %        figure
% %        stem(lags12Z,c12Z);
% %        title({['xcorr for first good window, Z channels of
    1 and 2'],...
%          ['start time ',num2str(goodtime12(1))]})
% %        xlabel('Lags in Number of Samples')
% %        ylabel('Amplitude(Not Normalized)')
% else
%      goodindx12=0;
%      gstarttime12=0;
% end

%do it for all good windows
 if (numcomm12>0)
```

```
for i0 =1:numcomm12
goodindx12(i0)=find(GGEt==st12(i0));
gstarttime12(i0)=GGEt(goodindx12(i0));
gstartdata12(i0)=edata1(goodindx12(i0));
goodstarts12=[gstarttime12(i0),gstartdata12(i0)];
gooddata1E=edata1(goodindx12(i0):goodindx12(i0)+win);
gooddata2E=edata2(goodindx12(i0):goodindx12(i0)+win);
goodtime12=GGEt(goodindx12(i0):goodindx12(i0)+win);
gooddata1Z=zdata1(goodindx12(i0):goodindx12(i0)+win);
gooddata2Z=zdata2(goodindx12(i0):goodindx12(i0)+win);
dmgood1E=gooddata2E−mean(gooddata2E);
dmgood2E=gooddata2E−mean(gooddata2E);
[c12E,lags12E]=xcorr(dmgood1E,dmgood2E);
% figure
% stem(lags12E,c12E);
% title({['xcorr for first good window, E channels of 1
and 2'],...
% ['start time ',num2str(goodtime12(1))]})
% xlabel('Lags in Number of Samples')
% ylabel('Amplitude (Not Normalized)')
dmgood1Z=gooddata1Z−mean(gooddata1Z);
dmgood2Z=gooddata2Z−mean(gooddata2Z);
[c12Z,lags12Z]=xcorr(dmgood1Z,dmgood2Z);
% figure
```

```
%        stem(lags12Z,c12Z);
%        title({['xcorr for first good window, Z channels of 1
    and 2'],...
%          ['start time ',num2str(goodtime12(1))]})
%        xlabel('Lags in Number of Samples')
%        ylabel('Amplitude(Not Normalized)')
    end
else
    goodindx12=0;
    gstarttime12=0;
end
```

B.2.7   stationgeom3.m

```
function[deltax,alpha,goodaz1s,goodaz2s]=stationgeom3(delr,
    aa,bb,goodaz1deg,...
    goodaz2deg,numcomm12,goodtradeg1,goodtradeg2,
        gstarttime12,goodindx12)


%
% %calculate the angle between the raypath from station A to
    station B and
% %North
% psi=atan(aa/bb);  %aa is easting, bb is northing
```

```
% %set phi = the average of the azimuth at the first station
    and the azim of
% %the second station for each good window:   this is the
    number of degrees
% %off of north
% if numcomm12>0
% for ii=1:numcomm12
%      wheresg1(ii)=find(goodtradeg1(:,2)==gstarttime12(ii));
%      wheresg2(ii)=find(goodtradeg2(:,2)==gstarttime12(ii));
%      goodaz1s(ii)=goodaz1deg(wheresg1(ii));
%      goodaz2s(ii)=goodaz2deg(wheresg2(ii));
%      %%%FIX THIS:   need to subtract the angles from 180
%      if goodaz1s(ii)>180
%          goodaz1s(ii)=goodaz1s(ii)-180;
%      else
%          goodaz1s(ii)=goodaz1s(ii);
%      end
%      if goodaz2s(ii)>180
%          goodaz2s(ii)=goodaz2s(ii)-180;
%      else
%          goodaz2s(ii)=goodaz2s(ii);
%      end
```

```
%
%
%      %take  average  angle  of  approaching  energy  for  each
   good  window
%      phi(ii)=((goodaz1s(ii)+goodaz2s(ii))/2)*(pi/180);   %
   convert  to  rad
%      %identify  the  unit  vector  pointing  towards  the  source
%      xx1(ii)=sin(phi(ii));
%      yy1(ii)=cos(phi(ii));
%      zz1=[xx1(ii);yy1(ii)];
%      matzz1(:,ii)=zz1;   %first  col=xx1(1),yy1(1);   2nd  col=
   xx1(2),yy1(2);
%      %identify  the  unit  vector  pointing  from  one  station  to
    another
%      xx2=abs(aa);
%      yy2=abs(bb);
%      %xx2=aa;
%      %yy2=bb;
%      zz2=[xx2;yy2];
%      %find  inner  product  between  the  two  vectors
%      inprodzz12(:,ii)=matzz1(:,ii)'*zz2;
%      normzz1(:,ii)=norm(matzz1(:,ii));   %it's  a  unit  vector
    ->  should  =  1
%      normzz2=norm(zz2);
```

```
%        cosbetaprime ( ii )=inprodzz12 (: , ii ) ./( normzz1 (: , ii )∗
    normzz2 ) ;
%        betaprime ( ii )=acos ( cosbetaprime ( ii ) ) ;
%      %be  sure  that  beta  is  acute
%        if  betaprime ( ii )>( pi /2) ;
%           %beta ( ii )=(2∗ pi)−betaprime ( ii ) ;  this  is
    subtracting  from  360..
%            beta ( ii )=pi−betaprime ( ii ) ;
%        else
%            beta ( ii )=betaprime ( ii ) ;
%        end
%        alpha ( ii )=( pi /2)−beta ( ii ) ;
%        deltax ( ii )=delr ∗( sin ( alpha ( ii ) ) ) ;  %hypotenuse ∗sin (
    alpha )=deltax
% end
% else
%        alpha =0;
%        deltax =0;
% end
% if  numcomm12>0
% for  ii =1: length (numcomm12)
%        wheresg1 ( ii )=find ( goodtradeg1 (: ,2)==gstarttime12 ( ii ) ) ;
%        wheresg2 ( ii )=find ( goodtradeg2 (: ,2)==gstarttime12 ( ii ) ) ;
%        goodaz1s ( ii )=goodaz1deg ( wheresg1 ( ii ) ) ;
```

```
%          goodaz2s ( ii )=goodaz2deg ( wheresg2 ( ii ) ) ;
%       %%%FIX THIS :   need  to  subtract  the  angles  from  180
%        if  goodaz1s ( ii )>90
%            goodaz1s ( ii )=goodaz1s ( ii ) −180;
%       else
%            goodaz1s ( ii )=goodaz1s ( ii ) ;
%       end
%        if  goodaz2s ( ii )>90
%            goodaz2s ( ii )=goodaz2s ( ii ) −180;
%       else
%            goodaz2s ( ii )=goodaz2s ( ii ) ;
%       end
%
%
%      %take  average  angle  of  approaching  energy  for  each
   good  window
%       phi ( ii )=(( goodaz1s ( ii )+goodaz2s ( ii ) )/2) ∗( pi /180) ;    %
   convert  to  rad
%      %identify  the  unit  vector  pointing  towards  the  source
%       xx1 ( ii )=sin ( phi ( ii ) ) ;
%       yy1 ( ii )=cos ( phi ( ii ) ) ;
%       zz1 =[xx1 ( ii ) ; yy1 ( ii ) ] ;
%       matzz1 ( : , ii )=zz1 ;   %first  col=xx1 (1) , yy1 (1) ;   2nd  col=
   xx1 ( 2 ) , yy1 ( 2 ) ;
```

```
%      %identify the unit vector pointing from one station to
    another
%      xx2=abs(aa);
%      yy2=abs(bb);
%      %xx2=aa;
%      %yy2=bb;
%      zz2=[xx2;yy2];
%      %find inner product between the two vectors
%      inprodzz12(:,ii)=matzz1(:,ii)'*zz2;
%      normzz1(:,ii)=norm(matzz1(:,ii)); %it's a unit vector
    -> should = 1
%      normzz2=norm(zz2);
%      cosbetaprime(ii)=inprodzz12(:,ii)./(normzz1(:,ii)*
    normzz2);
%      betaprime(ii)=acos(cosbetaprime(ii));
%    %be sure that beta is acute
%      if betaprime(ii)>(pi/2);
%          %beta(ii)=(2*pi)-betaprime(ii); this is
    subtracting from 360..
%          beta(ii)=pi-betaprime(ii);
%    else
%          beta(ii)=betaprime(ii);
%    end
%      alpha(ii)=(pi/2)-beta(ii);
```

```
%         deltax(ii)=delr*(sin(alpha(ii)));   %hypotenuse*sin(
    alpha)=deltax
% end
% else
%       alpha=0;
%       deltax=0;
%       goodaz1s=0;
%       goodaz2s=0;
% end


%try it for first good window
 if numcomm12>0
 for ii=1
     wheresg1(ii)=find(goodtradeg1(:,2)==gstarttime12(ii));
     wheresg2(ii)=find(goodtradeg2(:,2)==gstarttime12(ii));
     goodaz1s(ii)=goodaz1deg(wheresg1(ii));
     goodaz2s(ii)=goodaz2deg(wheresg2(ii));
     %%%FIX THIS:   need to subtract the angles from 180
      if goodaz1s(ii)>90
          goodaz1s(ii)=goodaz1s(ii)-180;
      else
          goodaz1s(ii)=goodaz1s(ii);
      end
      if goodaz2s(ii)>90
```

```
    goodaz2s(ii)=goodaz2s(ii)−180;
else
    goodaz2s(ii)=goodaz2s(ii);
end


%take average angle of approaching energy for each good
    window
phi(ii)=((goodaz1s(ii)+goodaz2s(ii))/2)*(pi/180);   %
    convert to rad
%identify the unit vector pointing towards the source
xx1(ii)=sin(phi(ii));
yy1(ii)=cos(phi(ii));
zz1=[xx1(ii);yy1(ii)];
matzz1(:,ii)=zz1;   %first col=xx1(1),yy1(1);   2nd col=
    xx1(2),yy1(2);
%identify the unit vector pointing from one station to
    another
xx2=abs(aa);
yy2=abs(bb);
%xx2=aa;
%yy2=bb;
zz2=[xx2;yy2];
%find inner product between the two vectors
```

```
inprodzz12 (: , ii )=matzz1 (: , ii ) '* zz2 ;
normzz1 (: , ii )=norm ( matzz1 (: , ii ) ) ;   %it 's  a  unit  vector
   -> should  =  1
normzz2=norm ( zz2 ) ;
cosbetaprime ( ii )=inprodzz12 (: , ii ) ./( normzz1 (: , ii )*
   normzz2 ) ;
betaprime ( ii )=acos ( cosbetaprime ( ii ) ) ;
%be  sure  that  beta  is  acute
if  betaprime ( ii )>( pi /2) ;
   %beta ( ii )=(2* pi )-betaprime ( ii ) ;  this  is  subtracting
      from  360..
   beta ( ii )=pi -betaprime ( ii ) ;
else
   beta ( ii )=betaprime ( ii ) ;
end
alpha ( ii )=( pi /2)-beta ( ii ) ;
deltax ( ii )=delr *( sin ( alpha ( ii ) ) ) ;   %hypotenuse * sin ( alpha
   )=deltax
end
else
   alpha =0;
   deltax =0;
   goodaz1s =0;
   goodaz2s =0;
```

end

B.2.8   method2.m

```
function [ff]=method2(good3131,win,alpha,delr,npts,df,dt,
    zdata1,zdata2)
%this is the function version of method 2 for determining
    the phase
%velocity dispersion curve, which uses cross correlation of
    two signals
%during a good window.


f=0:npts-1;
f=(f*df);
fs=1/dt;
NFFT = 2^nextpow2(win);


ff=fs/2*linspace(0,1,NFFT/2+1);




for i2=1:length(good3131);
    matz1=zdata1(good3131(i2):good3131(i2)+win);
    matz2=zdata2(good3131(i2):good3131(i2)+win);
    z1m=matz1-mean(matz1);
```

```matlab
z2m=matz2-mean(matz2);

Z1m=fft(z1m,NFFT)/win;

Z2m=fft(z2m,NFFT)/win;

Z1m=conj(Z1m);

xcorr12=Z2m.*Z1m;

angw=angle(xcorr12);

UWw=unwrap(angw);

xx=linspace(.1,1,15);

yy=spline(ff,UWw(1:NFFT/2+1),xx);

maxgoodxx=find(xx==1);

for jj=1:maxgoodxx;

    phaseangsw(jj)=yy(xx==xx(jj));

    deltsw(jj)=-phaseangsw(jj)/(2*pi*xx(jj));

    vtruesw(jj)=(sin(alpha(16))*delr)/deltsw(jj);

    goodxxw=xx(1:jj);

end

plot(goodxxw,vtruesw)


title('real data phase velocity dispersion curve')

xlabel('frequency (Hz)')

ylabel('velocity (m/s)')

drawnow; pause(2)

end
```

Appendix C

GEOLOGY OF THE ALEUTIANS AND ALASKA

C.1   Introduction

The geology of southern Alaska is complex and, in its most extreme geographic locations, is still poorly understood. To familiarize the reader with the geologic setting of Akutan, where the methods developed in Chapter 4 are applied, and to inform the reading of results and discussion Chapters 5 and 6, an overview of the geology both for Alaska as a whole and Akutan specifically is provided here. As this project was originally conceived with the intent to build a fully automated early warning system based on the methods developed here, an understanding of the historic and prehistoric volcanic behavior of Akutan is also desirable. This chapter begins with a description of the geological and concludes with a brief review of the volcanic activity along the arc and at Akutan.

C.2   General Geological Setting

Alaska geology may be broken down into four major physiographic subdivisions (Detterman *et al.*, n.d.). The northernmost physiographic province consists of the Interior

Plains division of North America. The Arctic Coastal Plain, an area of current interest for petroleum exploration, is entirely contained within this province. South of the Arctic Coastal Plain are the units comprising the Cordillera orogenic belt of North America. From north to south, they are the Rocky Mountains System, corresponding to the northern Cordillera, the Intermontane Plateaux System, consisting of the region now referred to as Interior Alaska, and the Pacific Mountains System, which is the Alaskan extent of the southern Cordillera. Unlike Interior Alaska, which remained mostly unglaciated in the Pleistocene and allowed for overland travel from Eurasia, the northern and southern Cordillera were mostly covered in ice during the glaciations. Glacial topography is evident throughout (Plafker and Berg, 1994).

The Aleutian Arc is the western extent of the Alaska Range, which is a part of the large Pacific Mountains System that follows the west coast of North America. Underlying the southern Cordillera is an intraoceanic arc terrane, oceanic plateaux with extensive intrusions of mid-Cretaceous to Paleogene-aged plutons, accretionary prisms that formed due to the unusually high rate of sediment deposition into the intraoceanic arc, and flysch basins, all ranging in age from Proterozoic to Cenozoic. The mountainous topography on the mainland formed as a result of the ongoing accretion along the northern Gulf of Alaska. The Aleutian Islands, however, are composed of Cenozoic volcanic rock from stratovolcanoes formed north of the Aleutian Megathrust boundary. There are at least 36 volcanoes that have been active since 1760 along the island chain (Coats, 1950) and they are spaced, with few exceptions, at about 60-70 km apart (Fournelle *et al.*, 1994). The western Aleutian magmatic arc is built on relatively young oceanic crust while the eastern portion overlays the older accreted crust of the southern Cordillera (Plafker and Berg, 1994). The volcanism along the arc occurs along a linear, segmented front with a decrease in activity from east to

west. Historic volcanic activity is at its maximum at the area of greatest convergence between the Pacific and North American plate and at its least in the west, where the boundary tends towards transform movement (Fournelle *et al.*, 1994).

Petroleum exploration and the construction of the infrastructure needed to transport gas and oil from Prudhoe Bay, a supergiant field on the North Slope, was the primary driver for geological and geophysical exploration of Alaska from the time of the discovery of Prudhoe Bay in 1968 through the 1980s. Exploration techniques included exploratory drilling, aeromagnetic mapping, ship-borne magnetic, gravity, and seismic surveys, and mapping of permafrost near the proposed track of the Trans-Alaska pipeline. The efforts of the National Science Foundation-funded Deep Sea Drilling Program revealed the first data on stratigraphy of the boundary between the oceanic and continental crust. Along the subduction boundary, extensive research was and still is currently devoted to mapping the complex structure of the megathrust and upper crust. In response to the eruption at Mount Spurr volcano in 1992, the USGS established the Alaska Volcano Observatory to study volcanology and geophysics in the region and to develop early-warning systems for the active volcanoes on both the mainland and along the Aleutians (Plafker and Berg, 1994).

Despite the increase in exploration, geological mapping, and geophysical and volcanological investigations over the latter half of the last century, many details of the geology and subsurface in Alaska remain vague. Where there is no interest in oil or mineral exploration, the geophysical research is limited to analyses of the recordings collected at permanent seismic networks or temporary field installments. The number of mapping and subsurface investigations has increased with time but, given the great size of Alaska, a comprehensive picture is still incomplete. Even for active volcanoes, the isolation and limited commercial interest in some locations has left us with only geological maps and

no understanding of the magmatic systems below the surface. For Akutan Volcano, the most detailed geological information specifically limited to the island is contained in the geological map by Richter et al. (1998).

At the time of publication of this document, the subsurface of Akutan Volcano has not been modeled. Many challenges are presented in the study of Akutan: it is located near the halfway point on the east-west stretch of the Aleutians and can be accessed only by boat or amphibious aircraft (Aku, 2012), and the permanent seismic network of broadband stations that allow deeper glimpses of the subsurface is particularly sparse with only four out of five total stations featuring overlapping time windows of more than a couple of months. The sparsity of stations, remote nature of the site, and need for analysis to develop an early-warning system to protect residents, air traffic, and commercial enterprises are all trademark challenges of geophysical investigations in Alaska.

C.3   Geology of Aleutian Islands - An Overview

Unimak Pass marks the division between the two geologically dissimilar portions of the Aleutian Islands. West of the division is the Aleutian Ridge segment. To the east is the Alaska Peninsula-Kodiak Island segment. Although the two sections have shared a similar evolutionary history since the Eocene, there is evidence of significantly different earlier histories. The primary focus of this portion of the text is on the geology of the western portion of the Aleutians, which includes Akutan Island. A broad outline of the geologic history of the Alaska Peninsula-Kodiak Island segment is also given in order to provide regional context and a point of comparison. Figure C.1 serves as a visual guide to the following overview. In this figure, blue, yellow, and pink highlighting indicates the timespan of the Lower Series, Middle Series, and Upper Series, respectively, and serves to demon-

strate the relationship between the named formations and the chronostratigraphic units. Straight lines between units are used for conformable contacts. Wavy lines show unconformities. Patterned sections represent missing geological record. When dashed versions of the straight and wavy lines are used, the precise date of the boundary is unknown. Question marks show uncertain contacts. The supplemental Table C.1 provides a description of each formation named in the stratigraphic columns.

C.3.1   Aleutian Ridge

The submarine basement structure of the Aleutian Ridge was formed from the volcaniclastic products of a long period of Tertiary volcanism. The Ridge has been volcanically active for at least 55 million years (Vallier *et al.*, 1994). Magmatism seems to have reached a peak every 2.5 million years for the past 10 million years, and every 5 million years over the period from 10-20 mya (Fournelle *et al.*, 1994). Along the arc, the rate of subduction, presence and absence of particular lithofacies, and structural features vary with location and in time. The ridge is segmented and rotation of the blocks has occurred due to stresses from the formation of strike-slip faulting in response to the change in spreading direction of the (now extinct) Kula plate, approximately 56-55 Ma (Lonsdale, 1988). The date of this directional change is marked elsewhere by the change in orientation of the Hawaiian Islands with respect to the Emperor Seamount chain. Later, westward translation and oblique stress affected the geographical movement of the blocks due to the oblique collsion of the Pacific plate with the North American plate (Vallier *et al.*, 1994).

Dating the different geological units is challenging because similar lithologies and stratigraphic sequences are found in units of all ages. Many of the rocks are volcanoclastic debris or igneous in nature without any fossils and, in the sequences where fossils are

Figure C.1: Google Earth image with superimposed stratigraphic columns showing differences in geology along Aleutian Arc from east to west. Akutan Island is indicated. Adapted from Vallier et al., 1998.

Geological Formations of Aleutian Island Arc and Alaskan Peninsula

Chronostrat. Unit	Formation Name	Description
Older Rocks	Kodiak Formation	Siliciclastic interbedded conglomerates, coarse sandstones, siltstones, and shales deposited in deep-sea fan environment ?.
	Ghost Rocks Fm	Folded and faulted unit with pillow basalts, greenstone, claystone, thin beds of limestone, and local deposits of tuffaceous sandstone, deposited in deep sea environment ?.
Lower Series	Komandorsky Fm	Volcaniclastic sedimentary rock with foraminifer fossils
	Basement Rocks	Stratified volcanic flows, volcaniclastic sediments, limestones, chert, sills, and dikes. Tholeiitic composition for igneous components.
	Krugoli Fm	Flysch beds with sandstone, argillite, and limestones; distinctly stratified.
	Nevidiskov Fm	Similar to Krugoli, likely correlative.
	Finger Bay Volcanics	Calc-alkaline composition basalt flows on Adak Island, tholeiitic, transitional and calc-alkaline compositions on Atka and Amalia Islands. Intermediate-composition hypabyssal rocks.
	Andrew Lake Frm	Sedimentary unit comprised of volcaniclastics from contemporaneous volcanoes and some sediment eroded from Finger Bay Volcanics, laminated cherts, porcelanite Hein and McLean (1980).
	Tolstoi Fm	Volcaniclastic and carbonaceous sedimentary rocks deposited in nonmarine environment.
	Copper Lake Fm	Similar to Tolstoi Fm, correlative.
	Stepovak Fm	Volcaniclastics and sandstones with interbedded volcanics, deposited in deep water turbidite environment ?.
	Meshik Fm	Andesite and dacite volcanic rocks and associated volcaniclastics.
	Sitkalidak Fm	Thick sandstone units, interbedded sandstones and siltstones, mudstone, and conglomerate, deposited in series of submarine fans ?.
Middle Series	Nikol'Skoye Suite	Volcaniclastic sedimentary rocks, pyroclastic flow deposits, conglomerates, and dolomites. Terrestrial plant fossils and marine animal fossils.
	Vodopad Suite	Lava flows, volcaniclastic rocks, mafic to intermediate composition rocks.
	Chuniksak Fm	Sandstones with felsic parent rocks.
	Hemlock Conglomerate	Fluvial sandstones and conglomerates with occasional finer textured sediments and coal interbedded. Broadleaf deciduous plant fossils and evergreen needle fossils Detterman et al. (n.d.).
	Belkofski Fm	Sandstones, siltstones, and conglomerates of tuffaceous volcaniclastic origin; interbedded volcanic tuffs and breccias. Marine transitioning up into nonmarine depositional environment Detterman et al. (n.d.).
	Unga Fm	Terrestrial coarse-grained sedimentary rocks, locally-derived volcaniclastic and volcanic rocks interbedded. Partially correlative to the Belkofski Fm ?, Detterman et al. (n.d.).
	Tachilni Fm	Sandstones, conglomerate, siltstone, shale, thin coal beds, and volcaniclastics deposited in a shallow marine to nonmarine environment.
	Bear Lake Fm	Correlative to Tachilni Formation.
	Sitkinak Fm	Coal-bearing sandstones, siltstones, and conglomerates deposited in a mostly nonmarine environment with occasional marine deposition of sandstone and siltstone.
	Narrow Cape Fm	Fossiliferous sandstone and siltstone, deposited in quiet shallow marine environment.
Upper Series	Faneto Fm	Thick fluvial sedimentary unit of uncertain age.
	Massacre Bay Fm	Lavas of andesitic and dacitic composition with associated volcaniclastics, sills and dikes.
	Volcanic Rocks	Quaternary deposits of basalt flows and associated volcaniclastics from local eruptions.
	Milky River Fm	Volcaniclastic rocks with occasional basaltic or andesitic flows, deposited in nonmarine environment.
	Morzhovoi Volcanics	Interbedded flows of basalt and pyroclastic and associated volcaniclastic rocks. Detterman et al. (n.d.)
	Tugidak Fm	Glacial sandstones and siltstones with occasional pebbles and cobbles, fossiliferous, deposited in a marine environment.

Table C.1: Names and descriptions for formations found in the Aleutian Arc.

found, there is a lack of index fossils. Due to the long period of extrusive volcanic activity and prevalence of plutonic intrusions, potassium-argon dating is generally inconclusive because large portions of the strata have been thermally altered. Although some specific studies were conducted to map the geology on particular islands in the Aleutians (see Gates *et al.* (1971) and Hein and McLean (1980), for example), with named formations from these studies shown on the stratigraphic columns of Figure C.1, large portions of the geologic record are absent and dates of formations are uncertain. In order to correlate the units found in the Aleutian Ridge with those found on the mainland and in the Alaska Peninsula-Kodiak Island sector of the Aleutians, the use of three chronostratigraphic units is generally preferred over lithological unit definitions. Divisions are based on the results of seismic reflection studies in the Adak-Amlia portion of the Aleutians (to the west of Akutan), where stratigraphic boundaries appear as reflectors that can be traced to shore and correlated with well logs and geologic observations of outcrops (Vallier *et al.*, 1994). Each chronostratigraphic unit corresponds to a stage in the ridge development.

Ridge building along the Aleutian Arc occurred due to the subduction of the now-extinct Kula plate below the North American plate. The Kula plate was first identified during analysis of magnetic mapping surveys over the Pacific Ocean by Hayes and Heirtzler (1968). The distinctive east-west striking magnetic anomalies, with older anomalies to the south, indicated the existence of a plate that had been mostly consumed over time. The Kula plate was the first extinct plate recognized by scientists and the direction and rate of motion were estimated from the array of magnetic anomalies. Its history included a split from the Farallon plate due to rifting during the Late Cretaceous, followed by a spreading rate from a Kula-Pacific plate ridge of 45-35 mm/yr until approximately 43 Ma when Kula plate likely fused with the Pacific plate (Lonsdale, 1988). The ancestral plate spread-

ing rates are significantly different than the current rate of convergence between the North American and Pacific plates, which is approximately 85-90 mm/yr (Vallier *et al.*, 1994). The accretion from the Kula-Farallon ridge was subducted beneath the North American plate and the rate of spreading at this margin cannot be determined. Although early publications claimed that the Kula plate had been completely subducted beneath the North American plate margin along the Aleutian Arc, one small remanent of the spreading center and a portion of the Kula plate itself can be found to the south of Attu Island (Lonsdale, 1988). Estimates for the date of cessation of spreading at the Kula-Pacific ridge range from 56 Ma (Byrne, 1979), when magnetic data shows a change in the configuration of the triple junction between the Kula, Pacific, and Farallon plates that realigned the spreading centers and required the extinction of the Kula-Pacific ridge, to 43 Ma (Engebretson *et al.*, 1984), when the plate changed direction and the Kula-Farallon spreading center rotated into alignment with the Pacific-Farallon spreading direction. The evolving nature of the spreading centers associated with the Kula, Farallon, and Pacific plates spurred and then ceased the volcanism and construction of the Aleutian ridge. The rotation of the Kula Plate is also a likely explanation for the apparent movement of the continental terranes from the Pacific Northwest in a westward and counterclockwise direction to their current position along the Gulf of Alaska (Lonsdale, 1988). The definitions of the Aleutian Arc chronostratigraphic series' start and end dates is based on the evolution of the tectonic stress regimes, subduction rates, and spreading rates of the Kula, Pacific, and Farallon Plates.

The lower series (LS) rocks found on the islands along the Aleutian Ridge are more heavily metamorphosed than the younger units but are otherwise difficult to date. The majority of the LS was probably emplaced beginning at the time of the rotational change in the direction of the motion and subsequent subduction of the Kula plate, when the major

period of ridge building began (Lonsdale, 1988). The underthrusting, which occurred in a northerly direction and was likely rapid, spurred a period of intense volcanism and growth of the Aleutian Arc (Vallier *et al.*, 1994). Assuming that spreading at the Kula-Pacific spreading center had ceased at around 43 Ma, by approximately 37 Ma, ridge building and igneous activity had slowed significantly in response. This is therefore taken to be the cut-off age for the LS. The LS rocks are comprised of tholeiitic, calc-alkaline, and transitional volcanics, plutonic rocks, and smaller-scale subsurface sills and dikes, as well as occasional deeper plutonic intrusions. Metamorphism to greenschist, prehnite-pumpellyite, and zeolite facies is common throughout. The LS also includes deeper mafic plutonic rock that may represent the basement rock of the ridge. Discontinuous thick layers are evident in the offshore seismic sections and the top of the LS is a strong reflector. In the back arc sections, the boundary between the Cretaceous igneous ocean crust is indistinct in the seismic records but it is thought that the LS merges with or overlies the crust. In the forearc section, the sedimentary rocks of the LS are mostly turbidites and debris flows.

Above the LS, the middle series (MS) consists of mostly sedimentary rocks formed after ridge building had ceased, during a time of erosion and sedimentation with minimal surificial volcanic activity. Products of erosion formed a 2-3 km thick sediment blanket over the fore- and back arc basins, suggesting a slow rate of subsidence and minimal structural relief (Vallier *et al.*, 1994). Plutonic activity occurred in two significant phases during the formation of the MS. The first, which includes the emplacement of the large Hidden Bay pluton, has been dated between 36 and 28 Ma and was mostly island-arc tholeiitic in composition. The later plutonic intrusions were calc-alkaline in nature and developed over the period from 20 - 9 Ma. The change in composition between the two episodes of plutonism may reflect a change in tectonic events, but more research is needed to confirm

the causes (Vallier *et al.*, 1994). Interestingly, it was the discovery of plutonic silica-rich rocks that spurred our current understanding of the relative ages of the volcanic and plutonic formations throughout the Aleutians. Felsic outcrops were identified by Capps (1934) on his brief excursion to the shore on Unalaska Island. The observations that the islands were mostly volcanic and that several other islands had smaller coarse grained, more felsic, intrusions led to the correction of the then-common assumption that the felsic rocks predated the volcanoes and the volcanoes had intruded through the granites (Capps, 1934). The cutoff age for the MS occurs at the start of renewed igneous activity along the ridge during the early Pliocene or latest Miocene.

The upper series (US) begins at 5.3 Ma. The period of renewed volcanism marking the emplacement of the US rock is not extensive enough to counteract the significant erosion of the topographic highs along the arc. The sediments in the US fill the fore-arc basins and the Aleutian Trench. The sediments also drape over the summit platform and form the main constituent of the accretionary wedge beneath the trench. Overall, the Ridge segment does not show evidence of westward travel since the Eocene.

C.3.2 Alaska Peninsula-Kodiak Island

The Alaska Peninsula-Kodiak Island (AP-KI) group is divided into discrete terranes with a contact between the shallow-marine and continental rocks of the northern portion and the deep-marine formations of the southern portion occurring along the Border Ranges Fault (Vallier *et al.*, 1994). The rocks forming the Alaska Peninsula and northern sections of its adjacent islands are mostly Permian to Holocene age, and most of the outcrops in the area are volcanic and plutonic igneous or shallow marine or coastal sedimentary rocks with predominantly volcanic and plutonic parent rocks, dating from the Paleozoic to Holocene. The

southern section, which includes the southern portion of Kodiak Island, the outer Shumagin Islands, and the Sanak Islands, is instead composed of flysch deposits from the Late Cretaceous and Paleogene with plutonic intrusions into the Cretaceous rocks of approximately 60 Ma.

The AP-KI Group has a significantly longer geological history than the Aleutian Ridge and has been more extensively studied. During the Cenozoic, this portion of the Aleutians had characteristics of a continental-oceanic plate subduction margin rather than an island arc. During the Mesozoic, the AP-KI was likely an island arc at tropical rather than sub-Arctic latitudes and has moved northward since the Triassic. The Paleozoic history of the AP-KI region is poorly constrained and mostly unknown. Although there is a lack of early Late Cretaceous rock in the AP-KI portion of the arc, rocks from later in the Late Cretaceous are present and allow for the observation of continued erosion of the paleo-arc and deposition of the sediments into the paleo-Aleutian trench. It does not appear that there was significant volcanic activity in this portion of the Aleutian Arc during the Late Cretaceous. The Meshik Formation, which consists of volcanic rocks and their associated volcaniclastics, is the first formation to suggest renewed volcanism in the AP-KI region. Volcanic activity continued and, in fact increased at the time of the shift in direction of the Kula plate. Significant plutons formed in this portion of the arc at or around the same time as the emplacement of the Hidden Bay pluton and the Finger Bay volcanics along the ridge segment. Another round of hypabyssal igneous activity occurring at the same time as the more recent plutonic activity along the ridge has also been preserved. Rapid uplift and erosion occurred during the late Miocene and, while the southern portion of this region has large deposits of volcaniclastic rocks from a second round of magmatism, the northern portion has instead a history of mostly sedimentary processes. The Quaternary

rock units are mostly sedimentary and show evidence of periods of glaciation throughout their emplacement (Vallier *et al.*, 1994).

Although the rocks of the AP-KI Group are easily separated into lithological groups rather than chronostratigraphic groups due to the existence of more well-defined ages and units, it is possible to loosely correlate the named formations of the AP-KI Group and the LS, MS, and US rocks of the Aleutian Ridge, as is indicated on Figure C.1.

C.3.3   Geology of Akutan Volcano

The current city of Akutan is located on the eastern shore of the island and is located in an area with abandoned beach deposits, undifferentiated volcanics, undifferentiated colluvium, tephra, and volcanic domes. Figure 2.6 in Chapter 2 provides a Google Earth-generated image of the island.

The majority of scientific literature published about this portion of the Aleutians focuses on anthropological history, effects on and of freight shipping through Unimak Pass, and biology. Geological studies are few in number and are motivated mostly by the potential for geothermal resources. On Akutan itself, geophysical studies are limited to analyses of the passive and earthquake-source data collected at the permanent seismic network stations and very near-surface seismic refraction and electromagnetic investigations of the geothermal resources at Hot Springs Bay (Motyka and Nye, 1988). Remote sensing studies of the island mostly focus on the large earthquake swarm event of 1996, which opened cracks in the ground trending at N $60°$ W (Lu *et al.*, 2000).

A map of Akutan Island, constructed by Richter et al. in 1998, is the most detailed reference available on the geology of Akutan. The map has been republished here on Plate I. The most recent caldera is on the summit of present-day Akutan Volcano, which

is located on the northern portion of the island, while an ancestral caldera wall built in the late Pleistocene is found farther inland. The younger caldera contains at least three lakes, first noted by Byers and Barth (1953), small glaciers, and areas of perennial snow cover, as well as a 200 m high and 1 km wide cinder cone where all historic volcanic activity has taken place (Waythomas *et al.*, 1998). The rim of the modern caldera was breached by a lava flow in 1978 that almost reached the Bering Sea. Older lava flows and lahar deposits have been mapped throughout the western portion of the island and are from the modern as well as the ancestral Akutan volcano (Richter *et al.*, 1998). To the east, older 1.5-3.3 Ma volcanic rocks were deposited by eruptions at the ancestral Akutan caldera and eruptions of other, offshore volcanoes. Analysis of the older deposits suggests a major drop in sea level during the Pliocene, as the earlier deposits were below water and the later deposits occurred in a terrestrial environment.

Outside of the modern cone, four eruptive centers have been found. The Lava Point flow on the northwest shore of the island represents the youngest of these eruptive centers and was likely emplaced some time after 1800 A.D., based on the unweathered appearance of the cinder cone and aa lavas that it produced (Waythomas *et al.*, 1998). Cascade Bight is the next youngest, with a smaller flow dated at 0.15 Ma. Flat Top has the most significant amount of activity out of all of the satellite eruptive centers and dates for these flows have an age of approximately 0.25 Ma. The oldest center consists of a series of domes arranged linearly along Long Valley, with a date of 0.58 Ma (Richter *et al.*, 1998). The rocks from the ancestral Akutan eruptions are more mafic than those erupted in modern times from the main cinder cone and the older rocks from the flows at the satellite centers. Dating of the volcaniclastic rocks and observations of lahar stratigraphy date the modern caldera forming eruption at approximately 1611 years B.P (Richter *et al.*, 1998). Motyka and Nye

report in their 1988 publication on the geothermal resources of Akutan that, in the earliest published geological study of Akutan (Finch, 1935), it was noted that Dr. T.A. Jaggar had visited the island twice and had correlated the exposed geology with the units of the Alaska Peninsula, but unfortunately these results were not published. This author thus assumes, based on the ages of the most recent volcanics on Akutan, that the rocks fall into the US and are likely members of either the Morzhovoi Volcanics or the group of Volcanic Rocks indicated on Figure C.1 in the Alaska Peninsula stratigraphic column. In addition to the aforementioned cracks opened by the earthquake swarm in 1996, structure on Akutan Island trends mostly WNW. Dikes tend to follow this trend in the east, but in the west a more chaotic arrangement may hint at a radial distribution around the main peak of Akutan or around the ancestral volcano (Richter *et al.*, 1998; Motyka and Nye, 1988) .

C.4 Quaternary Volcanism

For this study, we wish to understand the current and historical nature of volcanic eruptions on and near Akutan Island to understand the range of risks from an eruption of Akutan at present. This motivates the exploration of the subsurface at the site of Akutan Volcano. The general history of eruptions in the Aleutians is known, but the remoteness of the site may preclude knowledge of every single eruption and, in many cases, it is difficult to establish the exact source for volcanic rock deposits.

Akutan Volcano is a composite volcano, or stratovolcano, composed of layers of lava flow, volcaniclastic debris, and tephra (Waythomas *et al.*, 1998). The caldera-forming eruption of Akutan in the Late Quaternary was one of at least twelve such eruptions along the Aleutian Arc. It is thought that ten of the twelve eruptions during this time were significant enough to produce the climate changes during the late Quaternary (Miller and Smith,

1987). This in turn resulted in the exposure of the land bridge from Russia to North Amer-
ica. Radiometric carbon dating of organic material found beneath pyroclastic flow deposits
on Akutan Island yields an age of 5200 +/-200 years (Miller and Smith, 1987). The other
eleven caldera-forming eruptions occurred at Okmok, Makushin, Fisher, Emmons Lake,
Aniakchak, Black Peak, Veniaminof, Ugashik, Katmai, Kaguyak, and the well-studied No-
varupta volcanoes (Miller and Smith, 1987). Most of these volcanoes have had historic
eruptions.

Geological evidence has informed the current understanding of prehistoric eruptive pat-
terns. The earliest ash deposits in the populated area neaer the harbor are 9500 years old
and are likely to be from more distant eruptions to the west of the island. In the harbor area
itself, the earliest evidence of ash from Akutan is approximately 6100 years old. At Reef
Bight, Akutan ash appears in the stratigraphic record at about 8500 B.P. Lahar deposits
at Reef Bight predate the ash deposits and are assumed to be the oldest evidence of an
older Holocene eruption of the volcano, whereas other lahar and volcaniclastic deposits are
also present and were emplaced during Holocene eruptions dated later than 8500 B.P. It is
thought that these more recent deposits may be from the eruption that formed the ancestral
caldera (Waythomas *et al.*, 1998). In Long Valley and Broad Bight, a 1-2 m thick tephra
deposit, followed by pyroclastic flow and lahar deposits, were likely emplaced during the
eruption that formed the current caldera approximately 1600 years ago. It is not clear
whether the island was populated at the time of this eruption (Turner, 1972). Evidence of
minor eruptive activity from 470-600 years ago also is found in the form of thin ash layers
in Cascade Bight and, possibly, correlative lahars in Flat Bight (Waythomas *et al.*, 1998).

C.5 Eruptive History at Akutan

The formation and eruptive history of Akutan seems to match the evolutionary development of large volcanoes in the oceanic sector of the Aleutians as outlined by Fournelle et al. (1994): first, thin basalt lava flows with interbedding layers of tephra are deposited from eruptions at a main vent to form a shield-style volcano. Next, the eruptions continue to produce materials and the proportion of pyroclastic materials to basalt flows increases. The deposits of pyroclastic material maintain a higher angle of repose than the relatively flat lava flows and begin to build up the sides of the stratocone. As the stratocone is built, new vents begin to form away from the main vent on the sides of the shield. The cones formed at the new vents are built at the same time as the main stratocone. As buildup continues, composition of the magmas at the main cone become increasingly silica-rich while the magmas at the satellite vents remain basaltic. As the composition of magma at the main cone becomes dacitic, a dome is built and, eventually, the cycle culminates in a large silicic eruption. In the case of Akutan, this eruption was large enough to form a caldera. Once the main eruption has occurred, volcanism at the satellite cones subsides and the majority of new activity occurs in the caldera. After the volcanism in the caldera comes to an end, a new stratocone is built in the same fashion to the west of the first cone. This style of volcanism is significantly different from the pattern of continental volcanism, where there are few satellite vents and stratovolcanoes tend to form in clustered arrangements rather than in linear formations, as in the oceanic sector. On land, basaltic compositions are rare and magmas tend to be silicic (Fournelle *et al.*, 1994). Smaller volcanic centers are also found in the oceanic sector, but they do not show evidence of satellite vents and tend to produce summit and flank domes in their final stages of activity rather than extrusive eruptions. Others formed in ways similar to the large volcanoes but do not have calderas

and produce crystalline lavas. Cinder cones are found at other locations and, while rare, tend to occur in large clusters around stratovolcanoes. Their presence may be indicative of fissure-style eruptions (Fournelle *et al.*, 1994).

The current episode of active volcanism along the Aleutian Arc appears to have started approximately 1 Ma (Fournelle *et al.*, 1994). Today, Akutan has a maximal elevation of approximately 1,100 m above sea level as defined by the National Geodetic Vertical Datum of 1929 (Waythomas *et al.*, 1998). The hot springs, subject of an extensive investigation by the Alaska Division of Geological and Geophysical Surveys (Motyka and Nye, 1988), occur in the glacially-carved Hot Springs Valley. A fumarole field is found to the south of the hot springs on the flank of the volcano. The diameter of the current Akutan caldera is about 2 km and it ranges in depth from 60-365 m. Extrapolating from this size and the mappable extent of associated pyroclastic flows, as well making as comparisons to more thoroughly studied caldera-forming eruptions, has resulted in an estimated eruptive volume of under 10 km ^{3} (Miller and Smith, 1987). This suggests a Volcanic Explosivity Index (VEI) of $\geq$ 6. Although the eruptive volume necessary for climate-affecting eruptions is assumed to be between 10-50 km^3, an understanding of the eruptive history of Akutan is necessary for estimating current and understanding future volcanic hazards.

In the historical period between 1790 and 1980, Akutan Volcano has erupted 28 times, which placed it second on the list of most active Aleutian volcanic vents after Shishaldin Volcano (Fournelle *et al.*, 1994). Since then, it has erupted at least five more times and has thus gained first place on the list (AVO, 2012a), although the ranking is somewhat uncertain because not all eruptions are likely observed for the more remote volcanoes in the Aleutians. The majority of the recent eruptions are small-to-moderate explosions in VEI category 2 (Lu *et al.*, 2000) and only four eruptions since 1911 have resulted in volcanic

ash falling on the harbor (Waythomas *et al.*, 1998). Products of the recent eruptions tend to be tholeiitic basalts or andesites. The most recent reported eruption occurred in 1992. Several false alarms have been reported since then, but were eventually credited to steam from the fumarolic field on the flank of the volcano and geothermal activity in Hot Springs Valley (AVO, 2012b).

Perhaps the most notable event over the last 20 years was the earthquake swarm of 1996. At the time, the closest seismometers were located at Dutton Volcano, approximately 260 km to the East. More than 80 earthquakes of magnitudes >3.5 were recorded there on March 11 over an 11 hour period. Three days later, a 19-hour long swarm with over 120 earthquakes at magnitudes >3.5 were recorded. Thousands of other, smaller quakes were not recorded due to the distance between the monitoring station and the source. Although the largest earthquake out of the two active days had a magnitude of 5.3 (Waythomas *et al.*, 1998), the final, additive total of the seismic moment for the swarm is on par with that of a single magnitude 6.0 earthquake and the energy released during the swarm was greater than or equal to the energy released before the eruptions of Mount St. Helens, Mount Spurr, and Redoubt (Lu *et al.*, 2000). The lack of seismic stations on Akutan was reconciled with the emergency installation of four seismometers that were able to measure several thousand additional earthquakes from March 18-August (Waythomas *et al.*, 1998). After the quakes, cracks in the ground were found between Lava Point and the southeast shore of the island, all trending at N60°W to N70°W, which is approximately parallel to the motion of the Pacific plate relative to the North American plate. This trend direction is appropriate for the current tectonic stress regime (Lu *et al.*, 2000). The largest concentration of cracks was in a zone of approximately 500 m by 3 km, beginning at Lava Point and ending near the summit of Akutan volcano.

Synthetic aperture radar (SAR) interferometry studies measured an upward motion on the western flank of Akutan and an overall downward motion on the eastern side of the island as a result of the swarm event. Lu et al. (2000) concluded that the bulge on the western flank of the volcano was due to the intrusion of approximately 0.2 km^3 of magma into a reservoir located approximately 13 km below the surface. The resulting pressurization in the magma chamber forced magma to a point less than 1 km below the surface, which produced the bulge. The cause of the subsidence in the eastern portion of the island is not well understood. In 2007, an earthquake of magnitude 8.2 in Kurile Islands generated seismic activity along the ground cracks opened in 1996. Later that year, remote sensing methods again detected inflation on the western portion of the island, but there was no subsequent eruption (AVO, 2012b). Events like these are poorly understood and point to the potential for unexpected eruptions. This lends credence to the need for continuous monitoring to provide early warning.

C.6 Volcanic Hazards Associated with Eruptions at Akutan

Hazards associated with volcanic eruptions in general include volcanic ash clouds, ash fallout, pyroclastic flows and surges, floods or lahars, lava flows, avalanches, volcanic gases, earthquakes, tsunamis, and rock fall avalanches. A few comments on the types of likely hazards for eruptions of Akutan in particular are warranted.

Historical eruptions of Akutan tend to be Strombolian. Strombolian eruptions generally do not emit large amounts of ash, and ash from most recent eruptions has not traveled beyond the harbor (Waythomas *et al.*, 1998). The caldera-forming eruptions during pre-

historic times, however, would have produced much more significant quantities of ash and eruptions of this style may occur again in the future. The primary risk associated with ash is for air traffic. Over 5,000 aircraft fly through oceanic air space that could be affected by ash plumes from Aleutian volcanoes per month (William Comer, 2012, personal communication). To provide a general idea of the area affected by eruptions, the Puff Model, created by NOAA to model the spreading of ash clouds from eruptions along the Aleutians, was run with an assumption of the likely maximum height of volcanic ash clouds from Akutan's historical eruptions (National Weather Forecast Service Office, 2012). Particle amount and horizontal diffusivity were chosen randomly as 6,000 and 6,000 m^2/s, respectively, since the intent was simply to show the spreading direction for the ash cloud rather than simulating an actual eruption. The plume height was chosen as 5 km to reflect a likely maximum for a Strombolian eruption of the type that has occurred in recent history at Akutan. Plume base was arbitrarily selected to be 1,000 m. Figure C.2 shows the general movement of particles away from the source over a 12-hour period. Ash plumes from larger eruptions could disrupt air traffic over the Northern Pacific and mainland Alaska for days.

On the island, ash fallout from eruptions in 1911, 1948, and 1989 was reported at Akutan Harbor (Waythomas *et al.*, 1998). The fine particulate nature of ash decreases air quality and can lead to health problems for sensitive members of the population. Pyroclastic flow and surge deposits have been less of a hazard of historical eruptions as they seem to be limited to the flanks of the volcano and, in some eruptions, may not have breached the caldera on the summit of the volcano (Waythomas *et al.*, 1998; Richter *et al.*, 1998). Lahars and floods are primarily concerns during the winter and spring eruptions when there is snow on the volcano. In past eruptions, lahar deposits have traveled down topographic lows formed by runoff streams and have almost reached the waterfronts at Broad Bight and
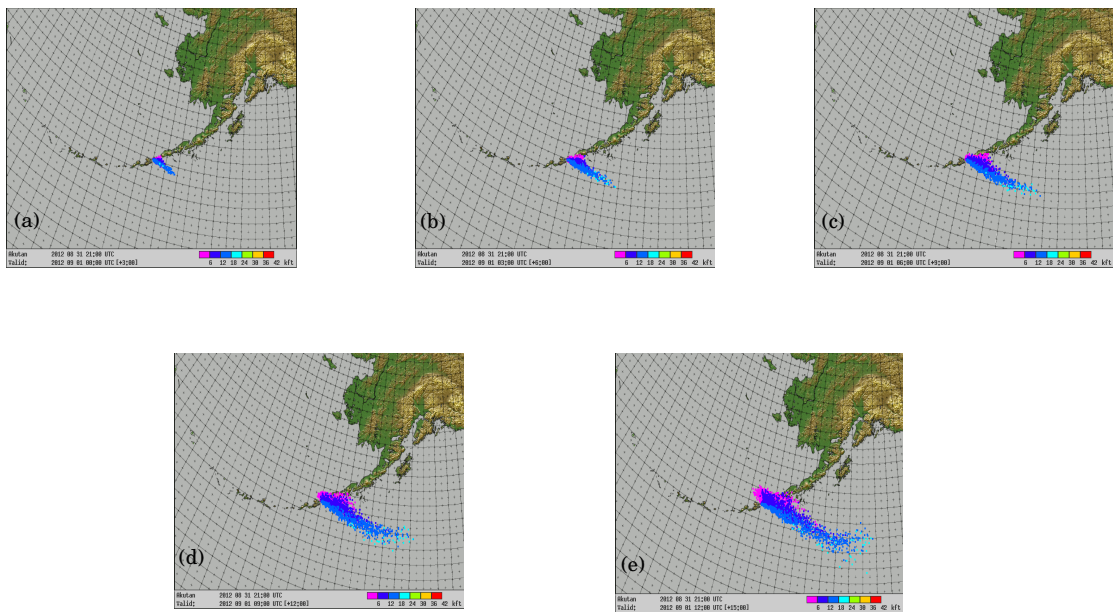
205

Figure C.2: Spread of ash at (a) 0:00 UTC, (b) 3:00 UTC, (c) 6:00 UTC, (d) 9:00 UTC, and (e) 12:00 UTC. Plots generated with the Puff Model to represent possible results of a large Strombolian eruption on Akutan.

Cascade Bight (Richter *et al.*, 1998), but since only the easternmost portion of the island is inhabited this does not pose a risk to settlements or structures. The lava flow towards Cascade Bight originated at the Cascade Bight peripheral volcanic center rather than at the main caldera. Lava flows on Akutan tend to move at rates of only tens of meters per hour and are not much of a hazard to people, but occasional blocky fronts to the lava flows may pose a risk of avalanches (Waythomas *et al.*, 1998). The remaining possible volcanic hazards are either unlikely to occur during the typical Strombolian eruption for Akutan or are uncommon (Waythomas *et al.*, 1998).

Appendix D

DERIVATIONS

D.1 Introduction

The SPAC method and the maximum likelihood method referenced in Chapter 3 are derived completely in a number of references, including the textbook by Aki and Richards (1980). The versions provided in this Appendix are informed by the derivations that have come before it and are printed in abbreviated form here to serve as a quick and convenient reference for the reader.

D.2 Derivation of the Spatial Autocorrelation Coefficient

Considering the case of two stations, for the case of a wave of frequency ω, the function of ground motion at the first station where the origin may be arbitrarily set is $u(0, 0, \omega, t)$ and the ground motion at the second station, located at point (r, θ), is $u(r, \theta, \omega, t)$. The spatial autocorrelation function $\phi(r, \theta, \omega)$ is then written as

$$\phi(r, \theta, \omega) = \overline{u(0, 0, \omega, t)u(r, \theta, \omega, t)}, \tag{D.1}$$

where the bar represents an averaged velocity in the time domain. Taking the spatial autocorrelation function for the station at the origin, $\phi(0,\omega)$, and using it to compute the directional average of the autocorrelation function over the circular array yields the spatial autocorrelation coefficient,

$$\rho(r,\omega) = \frac{1}{2\pi\phi(0,\omega)} \int_0^{2\pi} \phi(r,\theta,\omega)d\theta. \tag{D.2}$$

An integration of this equation easily shows that $\rho(r,\omega)$ is defined by a zeroth order Bessel function J_0 of the first kind,

$$\rho(\omega,r) = J_0(rk), \tag{D.3}$$

where r indicates the radius of the circular array and k is the wavenumber. A substitution of the relationship between wave number k and phase velocity, c,

$$k = \omega/c(\omega), \tag{D.4}$$

allows the SPAC to be written as

$$\rho(r,\omega) = J_0\left(\frac{\omega r}{c(\omega)}\right), \tag{D.5}$$

where $c(\omega)$ is the phase velocity for frequency ω. A quick substitution of the relationship $\omega = 2\pi f$ yields

$$\rho(f,r) = J_0(2\pi f r/c(f)). \tag{D.6}$$

This presentation of the derivation is based on the derivations by Aki (1957), Chavez-Garcia *et al.* (2005), Estrella and Gonzalez (2003), and Okada (2003).

D.3 Derivation of the Maximum Likelihood Method

To derive the MLM, first consider a normally distributed observed dataset $d_{t,i}$ for $t = (1,2,...,N)$ where N is the number of time samples in the data set, and $i = (1,2,...,M)$ where M is the number of stations where signals are observed. The actual signal portion of $d_{t,i}$ is $s_{t,i}$ and the noise portion is $n_{t,i}$, so that:

$$d_{t,i} = s_{t,i} + n_{t,i}. \tag{D.7}$$

To first consider the simple case of one moment in time as recorded at multiple stations ($N = 1$), rewrite Equation (D.7) using only the subscript i as:

$$d_i = s_i + n_i. \tag{D.8}$$

Assume that d_i is a Gaussian distribution with mean s_i. The variance between d_j and d_k (where j and k are elements in the set i) is represented as

$$\rho_{j,k} = <n_j, n_k> . \tag{D.9}$$

The $M \times M$ matrix with $\rho_{j,k}$ as the element at its jth row and k^{th} column is represented as ρ. Then, define Φ as the inverse of matrix ρ and represent the element at the jth row and k^{th} column as $\phi_{j,k}$. For $d = [d_1, d_2, ..., d_M]$, an M-dimensional vector, the joint probability density function (jdf) for the multivariate normal distribution of the observed normally-distributed datasets at each of the M sensors at only one point in time is then (Aster *et al.*,

2005; Aki and Richards, 1980; Okada, 2003):

$$f(d_1, d_2, ..., d_M) = \sqrt{\frac{det(\Phi)}{(2\pi)^M}} exp \left[-\frac{1}{2} \sum_{j,k=1}^{M} \phi_{j,k}(d_j - s_j)(d_k - s_k) \right], \qquad \text{(D.10)}$$

where *det* is the determinant of a matrix. Assuming the same signal is recorded at each of the sensors in the array, $s_1 = s_2 = ... = s_M = s$. To maximize the probability that a given d_i occurs, s is chosen to minimize the summation in Equation (D.10) (Capon, 1969), the weighted least squares criteria. It is found that the function is minimized when

$$s = \frac{\sum_{j,k=1}^{M} \phi_{j,k} d_k}{\sum_{j,k=1}^{M} \phi_{j,k}}. \qquad \text{(D.11)}$$

which is the weighted average of the data. For the case where the noise component of the observed signal at each station is not correlative between the stations, the matrix Φ will be diagonal, and

$$s = \frac{\sum_{k=1}^{M} d_k/\rho_k}{\sum_{k=1}^{M} 1/\rho_k}. \qquad \text{(D.12)}$$

As seen in Equation (D.11), the weighting acts to boost the contribution from the station with the least noise (Aki and Richards, 1980). The expansion of this brief analysis for the case of finite time series of sample length N rather than single time observations requires only the inclusion of the time shifting effect as a wave travels from station to station in the observed data and the inclusion of a second summation in the joint probability density function over the time dimension. To incorporate a finite number of sample times, the observed data equation as

$$d_{t,i} = s_{t,i} + n_{t+t_i,i}, \qquad \text{(D.13)}$$

where s is now a time series signal common to each station, and t_i is defined as in (3.1).

For the informative case of only one station, the jdf is now

$$f(t_1, t_2, ...t_N) = \sqrt{\frac{det(\Phi)}{(2\pi)^N}} exp\left[-\frac{1}{2}\sum_{l,m=1}^{N} \phi^{lm}(d_l - s_l)(d_m - s_m)\right], \qquad \text{(D.14)}$$

where l and m represent times in the range of t.

The unique trait of Capon's maximum likelihood method becomes apparent when one considers the term $s = [s_1, s_2, ..., s_N]$. It is assumed that the shape of the signal is known, but the amplitude is unknown so that the total expression for s may be written as

$$s = cf, \qquad \text{(D.15)}$$

where f is the shape factor and c is the amplitude factor. Substituting this definition into the exponent term in the expression (D.14), finding the derivative with respect to c, setting this equal to zero, and exploiting the symmetry of Φ, the maximum likelihood estimate for signal amplitude is found to be

$$c_{est} = (d^T \Phi f) \cdot (f^T \Phi d), \qquad \text{(D.16)}$$

where the superscript T indicates the vector transpose. It is easily proven that the estimate of amplitude does not distort the signal (Aki and Richards, 1980). The variance of the signal is then

$$c_{est,VAR} = (f^T \Phi f)^{-1}, \qquad \text{(D.17)}$$

which may be proven by definition. Finally, for signal shape $f_l = exp[i\omega(l-1)\Delta t]$ (a sinusoid with an amplitude of 1), the power spectrum P can be estimated at any angular

frequency ω in the receiver bandwidth by:

$$P(l,\omega) = \frac{1}{f_l^T \Phi f_l^*} = \frac{1}{\sum_{l=1}^N \sum_{m=1}^N \phi^{lm} exp\left[i\omega(l-m)\Delta t\right]} \tag{D.18}$$

where the asterix indicates a conjugate. To expand this analysis to the case of more than one station, begin with the jdf for the case varying in both space and time:

$$f(d_1,d_2,...,d_M,t_1,t_2,...t_N) = \sqrt{\frac{det(\phi)}{(2\pi)^{MN}}} exp\left[-\frac{1}{2}\sum_{j,k=1}^M \sum_{l,m=1}^N \phi_{jk}^{lm}(d_{l,i}-s)^T(d_{m,k}-s)\right], \tag{D.19}$$

and repeat the steps above.