P-Wave Study of the San Andreas Fault Near Parkfield, CA, From Ambient Noise Interferometry of Borehole Seismic Data

Stephen Mosher

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Master of Science degree in Earth and Environmental Sciences

Ottawa-Carleton Geoscience Centre
Department of Earth and Environmental Sciences
Faculty of Science
University of Ottawa

Summary

In this thesis, we investigate and develop the optimal data processing procedures necessary to recover Green's functions for body waves propagating among a network of borehole seismometers near Parkfield, CA. Applying these procedures, we detect P-waves propagating among these stations, which allows us to produce a first-order crustal velocity model for the San Andreas Fault in the Parkfield region. We also discuss under what conditions body wave phenomena such as reflections and mode conversions (P to S) may be observed, as further observing these would provide a dramatic improvement in our ability to characterize seismic velocity structures. Finally, we discuss the potential of seismic interferometry to produce time-lapse body wave characterizations of the San Andreas Fault, in which properties of the fault can be seen to change in time.

Deus Scientiarum Dominus Est

Acknowledgments

I am indebted to several people who have helped to make this thesis much better than I ever could have on my own.

I would first like to thank my supervisor, Pascal Audet, for your constant willingness and availability to provide guidance, support, and encouragement while working on this thesis. I'm often reminded of how fortunate I am to have such a great mentor when I hear about the experiences of other graduate students with their supervisors. I truly would not be where I am without your input.

I would also like to thank Andrew Schaeffer for always being available to listen, grab a coffee, and chat about research, which I truly appreciate. The comments you've provided to the many revisions of this thesis have been most helpful.

I'm especially grateful to Tyler Shendruk at Oxford, who generously lent me templates to help in preparing this document. I'm quite confident that you saved me countless hours of work formatting my thesis, which would not have looked nearly as good without your generosity.

I also wish to thank Taka'aki Taira at Berkeley for answering questions that I had regarding the HRSN instrumentation and Xavier Serra at Universitat Pompeu Fabra for giving me permission to use some of his data. Thank you both for your helpfulness in these matters.

Finally, I'm grateful for the encouragement and support I've received from my family during the time spent working on this thesis, especially from my lovely bride to be. Leah, I could not have asked for a better person to be at my side during my many endeavours, both presently and for those to come.

Contents

I cannot help feeling that seismology will stay in the place at the center of solid earth science for many, many years to come. The joy of being a seismologist comes to you, when you find something new about the earth's interior from the observation of seismic waves obtained on the surface, and realize that you did it without penetrating the earth or touching or examining it directly.

Keiiti Aki (1980)

1

Introduction

Contents

During the past two years while working on this thesis I often stole away to the department kitchen for respite, which, being on the fifteenth floor of one of the tallest buildings on campus, offered a fantastic view of the Rideau Canal. I would stand there, still, watching from above as glistening ripples flowed across the surface of the canal below. I watched as various undulations reflected and interfered, producing complicated patterns which ultimately faded from view. It was a comforting experience, because right there, in a beautiful simplicity, it seemed that the very thing I was trying to understand was being manifest.

1.1 Foreword

Seismic interferometry, which has developed rapidly within the past 15 years, is a large set of extremely powerful techniques for investigating subsurface Earth structures in a variety of settings. One particular technique within this greater scope is that of ambient noise interferometry. In a manner of speaking, one might describe ambient noise interferometry as a data processing methodology whereby we "see with noise", which to most, if not all, seems very counter-intuitive. After all, how could we ever "see with noise"? Moreover, to learn the inner workings of this technique is not a trivial task, it takes a determined effort to gain a deep appreciation and understanding of how and why "seeing with noise" actually works in practice. To understand the method one requires knowledge of elastodynamics, wave phenomena, linear-time-invariant systems, Fourier analysis, signal processing, etc. In a way, this thesis represents the undertaking of a journey to gain a deeper understanding of ambient noise interferometry, which ultimately culminated in our ability to study the properties of seismic waves "seen by noise" rather than those caused by ostensive sources.

1.2 Summary

1.2.1 Traditional Seismology and Ambient Noise Interferomerty

The fundamental aim of seismology is to understand the interior of the Earth and the distribution of its physical properties, especially in regions where direct access is impossible. Traditionally, the discipline accomplishes this by studying the properties of elastic vibrations, or seismic waves, that propagate within the Earth and which are caused by explicit sources such as earthquakes or large explosions. According to this framework then, the required elements for indirectly gleaning information about the Earth's interior are seismic waves, receivers (which record the ground motion produced by such waves), and sources of seismic waves (primarily earthquakes). However, given that the spatial distribution of earthquakes is generally limited, seismic waves necessary for studying features of interest within the Earth are also limited, and thus, the level of detail with which we may infer the physical properties of structures within the Earth is restricted. Put simply, traditional seismology doesn't always present as clear a picture as we would like. Unlike methods rooted in the traditional seismological framework, however, ambient noise interferometry approaches the same fundamental problem but from a radically different starting point, namely, one which obviates the reliance of seismic investigations on explicit sources, and thus, on direct seismic waves. Instead, ambient noise interferometry recovers information characterizing the Earth structure between two receivers simply by manipulating seismic noise data. As a consequence, given that seismic noise is not spatially constrained to the same degree as earthquakes, ambient noise interferometry has the potential to significantly overcome the challenges faced by traditional seismic methods. In particular, ambient noise interferometry has to the potential to improve the level of detail at which we can "see" inside the Earth, and also the accessibility of where we can "see".

1.2.2 A Brief Word on Seismic Noise

The Earth is in fact constantly shaking, and in many cases this ground shaking is directly attributable to definitive causes such as earthquakes or explosions. However, in the absence of discernible sources ground shaking persists, albeit at low levels, and it is precisely this ambient shaking that is referred to as seismic noise. Of course, this it not to say that seismic noise is without sources, on the contrary, it is known that sources of ambient seismic noise include tidal influences, ocean waves, human activity, and severe storms among other agents (Stein and Wysession, 2003). Nonetheless, the distinguishing factor between seismic noise and ground motion induced otherwise is the irreducibility of seismic noise to explicit causes.

1.2.3 A Simple Analogy

Conceptually, the possibility of gaining insight into the Earth's interior using ambient seismic noise is counter-intuitive. Furthermore, the mathematical and physical arguments underpinning ambient noise interferometry from a theoretical standpoint can seem cumbersome to work through. Therefore, we describe the basic mechanics of ambient noise interferometry by means of a physical analogy inspired by the work of Weaver and Lobkis (2002) in the field of ultrasonics. For a mathematically rigorous proof of the concept, with a focus on applications to seismic imaging, we refer the reader to a succinct derivation by Wapenaar (2004). Alternatively, for a more detailed yet simpler introduction we recommend the tutorial by Wapenaar et al. (2010).

We begin by envisioning an expanse of water sustaining random small amplitude fluctuations on its surface. Upon the surface of the expanse two receivers are anchored at some fixed distance from one another, each continuously measuring minute vertical displacements due to the surface fluctuations. Specifically, we imagine these fluctuations to be the result of a superposition of infinitely many waves, and we further impose that each wave contributing to the superposition state, while having equal amplitude, vibrates at a unique frequency and random phase with respect to all other waves. Thus, at any single instant of time, a snapshot of the surface of the water will yield a random configuration of wave fronts, and coherent wave propagation between the receivers will not be observed. However, over a long enough time interval, the individual parts of a coherent, impulsive-like source will effectively propagate between the receivers. Therefore, using this fact, the signal due to an impulsive-like source propagating between the two receivers can be built or recovered. To do so, one need only cross-correlate the individual signals recorded at each receiver, over many distinct time intervals, and perform an ensemble average. The reason that the cross-correlation operation is involved is because cross-correlation measures the similarity between two signals. Thus, each distinct cross-correlation measurement in the ensemble average can be thought of as contributing a small element of the recovered impulsive signal that is coherent between the receivers despite the overall random characteristic of the wavefield.

Apart from the medium, the situation depicted in this analogy corresponds to the exact experimental setup required to employ ambient noise interferometry in a seismic context. If one exchanges the water in the analogy for the Earth, all other aspects remain unchanged. However, we choose to use this analogy to convey the mechanics of ambient noise interferometry, simply because we find that randomly propagating waves on the surface of water are easier to visualize than randomly propagating seismic velocity fluctuations (ambient seismic noise) on the Earth's surface. Therefore, the result of performing ambient noise interferometry on seismic noise data recorded by two receivers is again, the recovery of the impulse response between those receivers. The impulse response, synonymous with the term Green's function, can be thought of as uniquely containing the information that characterizes the Earth structure between the two receivers. Said yet another way, the Green's function between the two receivers represents the signal that would be observed at one of the receivers, given an impulsive source had occurred at the other.

In developing this analogy we have glossed over three critical aspects which we here highlight. First, the analogy we developed involved a homogeneous medium, namely, an expanse of water. Second, we developed the analogy by considering a diffuse wavefield, that is to say, a wavefield in which the amplitudes of each vibrational mode are independent and have equal expected energies. Third, we implied a homogeneous distribution of surface fluctuations surrounding the receivers. In practice, however, the Earth is an inhomogeneous body, ambient seismic noise cannot always be treated as having a diffuse character, and ambient seismic noise is often inhomogeneously distributed around receivers. While each of these aspects indeed impact ambient noise interferometry from a theoretical standpoint, they have been treated in detail in the literature, notably in the derivation given by Wapenaar (2004), and subsequently do not render the method inapplicable to an Earth setting. We further address these issues in the body of this work.

1.2.4 Research Intent

In this thesis we explore the optimal conditions under which ambient noise interferometry may be applied to the High-Resolution Seismic Network (HRSN) of borehole seismometers distributed around the San Andreas Fault (SAF) in the Parkfield region of Southern California. We aim to examine several data processing techniques and propose a set of procedures that are specifically optimized for recovering body wave Green's functions from ambient noise data. Finally, we seek to apply such procedures to data obtained from the HRSN stations in order to demonstrate the potential of ambient noise interferometry to enhance future body wave tomography studies of the SAF.

We shall introduce the concepts of signal and noise. We define the signal as the desired part of the data and the noise as the unwanted part. Our definition of signal and noise is subjective in the sense that a given part of the data is "signal" for those who know how to analyze and interpret the data, but it is "noise" for those who do not.

Aki and Richards (1980)

2

Article: P-Wave Study of the San Andreas Fault Near Parkfield, CA, From Ambient Noise Interferometry of Borehole Seismic Data

Contents

2.1 Introduction

2.1.1 Ambient Noise Interferometry and Seismic Tomography

In the past decade and a half, seismic interferometry has been established as an immensely useful tool, with a wide variety of applications to seismological investigations (see e.g. Weaver and Lobkis, 2002; Wapenaar and Fokkema, 2006; Schuster, 2009; Wapenaar et al., 2010a). Of particular interest is the use of ambient noise interferometry, wherein the Earth's Green's function (impulse response) between two receivers is approximated by the ensemble average of cross-correlations of the ambient seismic noise field recorded by both receivers (Wapenaar, 2004). Mathematically this concept is described as

$$G(A, B, t) \approx \langle V(A, t) \otimes V(B, t) \rangle, \tag{2.1}$$

where $G(A, B, t)$ represents the Green's function between receivers A and B, $V(A, t)$ and $V(B, t)$ represent the ground velocities caused by ambient seismic noise and recorded at receivers A and B, respectively, $\otimes$ represents the cross-correlation operation, and the brackets $\langle \; \rangle$ denote an ensemble average. Such a result represents an approximation of the seismic waveform that would be recorded by one receiver given an impulsive source had occurred at the other. Using ambient noise interferometry, Green's functions for surface waves propagating between receiver pairs are readily obtainable, and have been used for surface wave tomography studies in a variety of settings (e.g., Shapiro et al., 2005; Brenguier et al., 2007; Lin et al., 2008). Because seismic noise is largely a surface phenomenon (Wapennar et al., 2010a) that originates mainly from ocean waves, atmospheric phenomena and human activities, surface waves are the most easily extracted type of seismic wave from ambient noise cross-correlations. However, it has been shown that the extraction of Green's functions for body waves is also possible, provided that receivers are spaced sufficiently close (less than 11 km) (Roux et al., 2005). Therefore, ambient noise interferometry further presents the possibility of performing not only surface wave tomography, but also body wave tomography on small scale networks. The first example of such an application was performed by Nataka et al. (2015), who used a small and dense network of approximately 2,500 receivers with 100 m spacing to perform a body wave tomography study at Long Beach, California.

Given that ambient noise interferometry effectually creates new sources at receiver locations, seismic tomography performed in this manner presents significant advantages compared to traditional tomographic techniques. In particular, model resolution no longer depends upon sporadic sources, both spatially and temporally, and station locations (i.e. virtual sources) are known exactly, removing the uncertainty in source location. Therefore, these advantages allow for the production of potentially more accurate and higher-resolved seismic velocity models compared to models based on traditional tomographic techniques.

2.1.2 Body Wave Recovery Among High-Resolution Seismic Network Stations

The San Andreas Fault (SAF) in the Parkfield region of Central California currently separates the dominantly granitic Salinian block, formerly part of the Sierra Nevada and Peninsular Ranges, on the southwest side, from sedimentary units of the Franciscan Complex on the northeast side (Audet, 2015). A related yet poorly constrained and understood aspect of the SAF is the presence of long-duration, low-magnitude seismic events with no clear P- or S-waves (Nadeau and Dolenc, 2005). Such events, called tremors, occur at depths below the seismogenic zone (between 20 and 40 km depths) and are thought to represent repeating slow earthquakes where an episode can last several tens of minutes, or hours (Peng and Gomberg, 2010). One of the major challenges remain to relate the occurrence of tremors with the specific geological environment of the SAF. In addition, due to the heterogeneity in geologic structure, the seismic velocity structure of the SAF in this region is highly complex. Notable velocity features include sharp contrasts across the fault (Thurber et al., 2004), as well as has been recently shown, low-velocity anomalies coincident with regions of tremor activity, and a high V_p/V_s zone in the mid-crustal region on the northwest side of the fault (Zeng et al., 2016).

While three-dimensional seismic velocity models for the SAF in this region have continuously improved (see e.g. Eberhart-Phillips and Michael, 1993; Thurber et al., 2003; Thurber et al., 2006; Zeng et al., 2016), a specific need to further improve the local S-wave velocity model has been noted by Zeng et al. (2016), who also propose the use of ambient noise based tomography to do so. Ultimately, improving body wave velocity models of the SAF in Central California may help in characterizing its structure and geology as well as that of the surrounding area, thus improving our understanding of the faulting process. Since the results of any ambient noise based body wave tomography study depend upon the waves recovered, we explore the feasibility of recovering body waves from seismic noise, propagating

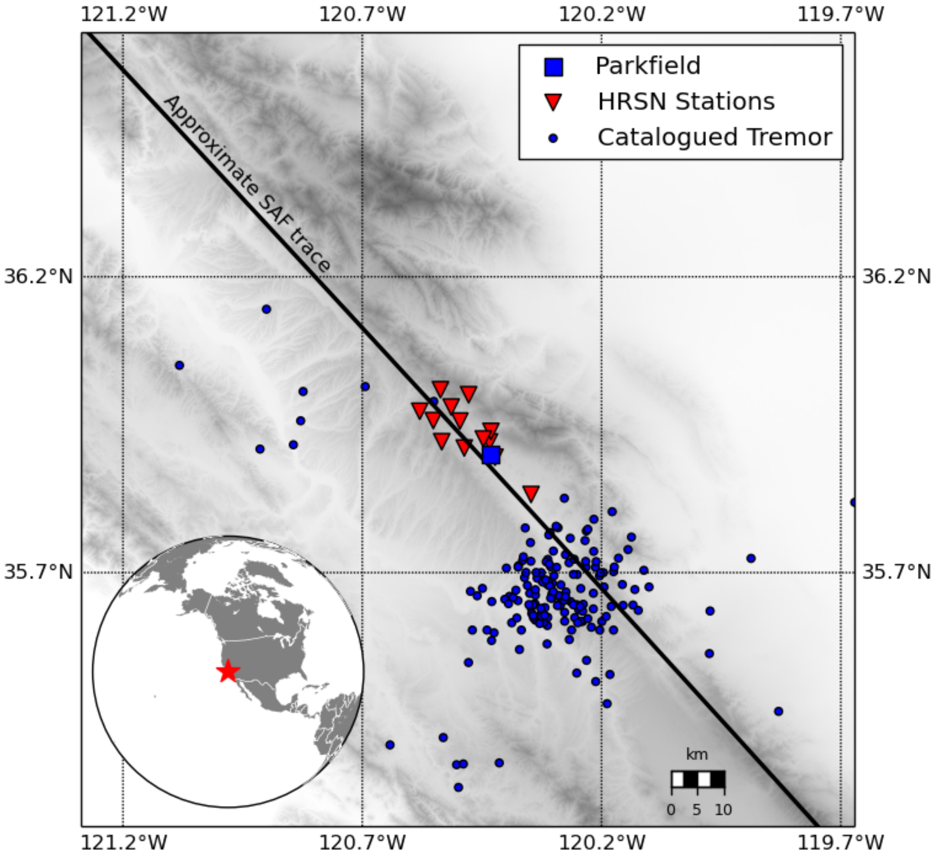

Figure 2.1.1: Region of study, featuring the approximate surface trace of the SAF, HRSN stations, the town of Parkfield, and locations of tremors during 2003 catalogued by Nadeau and Guilhem (2009).

between borehole stations of the High-Resolution Seismic Network (HRSN) in the Parkfield region (Figure 2.1.1, Table 2.1.1). More specifically, we demonstrate improvements to the data processing procedures that yield optimally recovered body wave travel-times, with an initial focus on P-waves.

Station	Lat. (degrees)	Lon. (degrees)	Elevation (m)	Depth (m)
CCRB	35.9572	-120.5516	595	350
EADB	35.8952	-120.4226	469	254
FROB	35.9109	-120.4869	515	258
GHIB	35.8323	-120.3473	393	370
JCNB	35.9390	-120.4311	533	335
JCSB	35.9212	-120.4340	454	332
LLCB	35.9801	-120.5142	385	385
MMNB	35.9565	-120.4960	701	510
RMNB	36.0009	-120.4777	1164	1125
SCYB	36.0094	-120.5366	947	695
SMNB	35.9730	-120.5799	648	450
VARB	35.9261	-120.4471	475	-61
VCAB	35.9216	-120.5339	755	590

Table 2.1.1: HRSN station coordinates, elevations, and depths. Sensor elevations are relative to mean sea level. Depth values refer to the depth to the sensor from the surface.

2.2 Methodology

2.2.1 Necessary Conditions for Ambient Noise Interferometry

Generally speaking, the applicability of approximating seismic Green's functions through ambient noise cross-correlations is contingent upon the spatio-temporal properties of the seismic noise field incident upon the network or station pairs being used. In theory, the seismic noise field must satisfy two primary and two secondary conditions if Green's functions are to be retrieved through use of this technique. The primary conditions are that contributors to the noise field must be: 1) uncorrelated, or, incoherent, and 2) share the same power spectrum apart from a normalization factor (Wapenaar, 2004). The secondary conditions to be met are that the noise field be: 1) homogeneously distributed and 2) spectrally white with respect to vibrational frequency. In practice, however, neither of these secondary conditions is satisfied, and for this reason, ambient noise cross-correlation studies require data pre-processing workarounds that address this.

Although the theory requires a homogeneously distributed noise field, this is in fact the weaker of the two secondary conditions to be met. The theory still holds as long as, for any given station pair, the noise field is sufficiently distributed within one of the Fresnel volumes surrounding the stations (Wapenaar et al., 2010a). In this case, the retrieved Green's function for the station pair will be asymmetric with respect to zero-lag time. For a station pair surrounded by a homogeneously distributed noise field, the contribution to the

Green's function by noise sources distributed outside the Fresnel volumes cancels out, and the Green's function estimate will be time symmetric. Therefore, the degree of homogeneity, with respect to the distribution of the noise field, controls the degree of symmetry of the retrieved Green's function, and furthermore, gives insight into which station pairs can yield usable Green's function estimates. In the case of significantly heterogeneous noise fields, a beamforming analysis can be performed to identify the origin of the incident noise field with respect to the network of stations. Once the directionality bias is known, ambient noise cross-correlation results are limited to station pairs whose inter-station orientations conform to the direction bias (see e.g. Roux et al., 2005).

When addressing a seismic noise field with a non-white frequency spectrum, a spectral whitening procedure is applied to seismograms in the data set. In all cases, implementing a spectral whitening procedure is an essential step for the successful use of ambient noise interferometry, as the spectrum of the observed noise field controls the frequency interval over which Green's function estimates are possible. A further pre-processing requirement, when approximating Green's functions from ambient noise cross-correlations, is to perform a time-domain normalization of individual station data. The purpose of time-domain normalization is to minimize the effects of seismic events or instrumental irregularities on cross-correlation results, in accordance with the first of the primary conditions noted above. Bensen et al. (2007) provide a detailed summary and comparison of common pre-processing procedures used in ambient noise cross-correlation studies for surface wave analyses as well as their merits. In this study we propose a departure from the typical work flow outlined by Bensen et al. (2007). This departure is necessitated by the distribution of HRSN stations around the SAF, the origin of seismic noise with respect to the fault, ongoing changes in HRSN station parameters, and our choice of whitening method. Our new work flow and its rationale are described in the following sections and are summarized schematically in Figure 2.2.1.

2.2.2 Data Pre-Processing Workflow

Data Preliminaries

According to Equation (2.1), the inter-station Green's function estimate between a pair of receivers must be produced by an ensemble average of noise recordings cross-correlated between the pair (Wapenaar, 2004). In other words, the Green's function estimate must be computed from many distinct cross-correlation measurements taken over an interval of time. However, for the purpose of ambient noise based body wave tomography, the previous

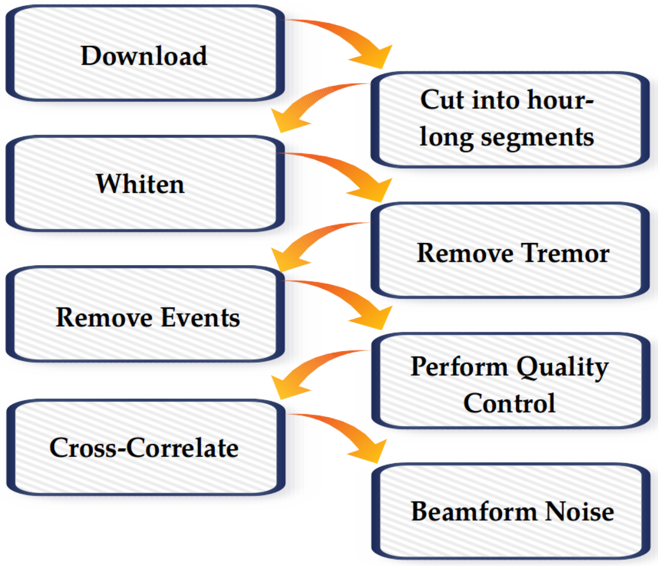

Figure 2.2.1: Data pre-processing workflow specifically developed for this study.

statement warrants caution, given that seismic noise (mainly ocean-related noise) varies seasonally. For this study, we therefore choose to use data recorded by borehole stations, since these effectively reduce the effects of seasonal variations on seismic data, and may be better shielded from surface noise sources (e.g, from human activity and other local sources of noise).

In exploring the feasibility of body wave retrieval using ambient noise interferometry, for the express purpose of investigating the SAF velocity structure, we seek data during a seismically quiescent period. For instance, between 2002 and 2007 two major earthquakes occurred in proximity to the HRSN, the M6.5 San Simeon earthquake (Dec. 22, 2003) and the M6.0 Parkfield earthquake (Oct. 11, 2004), both of which saw significant clusterings of tremor activity around the mainshock events as well as many related aftershock events (Brenguier et al., 2008). However, levels of tremor activity during 2003, prior to the San Simeon earthquake, were found to be comparatively lower. Therefore, our data processing begins by obtaining a year of 3-component short-period geophone data during 2003 from the HRSN in the form of day-long segments. This primary data set consists of 10,701 seismograms sampled at 20 Hz. Note that of 13 HRSN stations only 12 were operational during 2003 and that seismograms with less than a full day of data were excluded from the set.

Over the course of HRSN's latest stage of implementation (2001-present) many features of the stations have changed, notably component orientations, which also differ for each station, and instrument responses. Therefore, care must be exercised when working with data taken from this network over a long time span in order to ensure a consistent data set (Taira, personal communication, 2015). During 2003, all HRSN short-period geophone component orientations remained unchanged. However, the vertical components of stations GHIB and JCSB recorded data of opposite polarity with respect to all other stations; we therefore flip the polarity of vertical component data recorded by these stations during 2003. Finally, we cut day-long seismograms into hour-long segments to reduce the computational cost associated with manipulating long time series.

Whitening Procedure

Generally speaking, the goal of any data pre-processing is to make a set of raw data tractable for a specific analysis. Where many methods may suffice to achieve the same goal, the most desirable choice is often the simplest and easiest to implement. Therefore, in preparing a large set of seismic data for ambient noise interferometry, we seek the whitening method that is most effective while easy to implement.

The process of whitening a time series involves making operations in the frequency domain in order to equalize the amplitude information as much as possible, while minimally distorting the phase information, and without introducing biasing effects related to manipulating digitized time series. Any time series reconstructed exclusively from its phase spectrum is necessarily white since this operation discards spectral amplitude information. Therefore, we first attempt to whiten seismograms by reconstructing them exclusively from their raw phase spectra. In other words, we whiten seismograms by first computing their Fourier transform, obtaining their raw phase spectrum, and then back-transforming this phase spectrum. Although blindly performing this procedure without adaptation is indeed an effective whitening method, signals whitened in this manner are unsuitable for ambient noise interferometry. The reason being that inverse Fourier transforming time series that have been abruptly cut introduces large edge artifacts to reconstructed signals, which would remain coherent between cross-correlations, and thereby nullify the applicability of ambient noise interferometry given the primary theoretical conditions outlined in § 2.2.1. Hence, we pre-emptively taper signals in the time-domain before whitening. In this way we produce seismograms that are optimally whitened in as broad a frequency band as possible, and which are well-suited for ambient noise interferometry. Examples of the whitening operation without and with tapering are shown in panels five and six of Figure 2.2.2, respectively. To gain a deper insight into the appropriateness of this whitening method for ambient noise interferometry, we refer the reader to Appendix E.

We now examine the effects that other pre-processing steps have on the amplitude and phase of a resulting signal and show that in our context they are either unnecessary or undesirable. In many signal processing applications a decimation of signals is performed to increase the computational efficiency of further data manipulations. However, decimating signals limits their frequency bandwith while also introducing aliasing effects. Thus, in the case of ambient noise interferometry, this operation reduces the frequency range over which Green's function estimates can be obtained. Moreover, a reduction of frequency resolution in this context is especially undesirable, given that the HRSN components being used sample at 20 Hz. If we consider decimation by a factor of two for example, then the frequency resolution limit of these components becomes 5 Hz (i.e. the Nyquist frequency), yet, body waves generally propagate at frequencies ranging from 0.1 - 10 Hz, while reflected waves propagate at frequencies greater than 10 Hz (Stein and Wysession, 2003). Therefore, in order to optimally recover body waves from ambient noise data, we must to be able to resolve signal frequencies of at least 10 Hz. In addition to frequency limitations, in Figure 2.2.2 we note that of all the pre-processing steps examined, decimation also tends to alter raw signal phase the most. Upon these considerations, we therefore opt not to decimate raw HRSN data, which, sampled at 20 Hz, will not inhibit computational efficiency.

Other common pre-processing practices are the de-trending and demeaning of signals, in which linear trends are subtracted from time series and in which data become re-centred with a mean of zero, respectively. The purpose of these steps is again to avoid introducing edge effects if later filtering data, and to remove large DC components in the spectra, all of which serve to regularize signal amplitude (note for example, the difference in scale between the amplitude spectra of panels 1 and 3 in Figure 2.2.2). Additionally, in the case of raw HRSN data, a further step is often required, which is the application of a zero-phase high-pass filter with a corner frequency of 0.1 Hz. Applying a high-pass filter to raw HRSN data is necessary in order to remove unwanted low-frequency components that often persist in these signals. In the fourth panel of Figure 2.2.2, we see that the sample time series, which has been de-trended, demeaned, and high-pass filtered, has been sufficiently regularized for routine seismic analysis compared to the unfiltered, but otherwise identical time series in panel 3. Further note that the sequence of operations performed to produce the time series in the fourth panel of Figure 2.2.2 modifies the raw signal phase only slightly. While we could have de-trended, demeaned, and applied a high-pass filter to raw HRSN data before whitening, we find that by simply tapering signals pre-emptively and whitening them, we

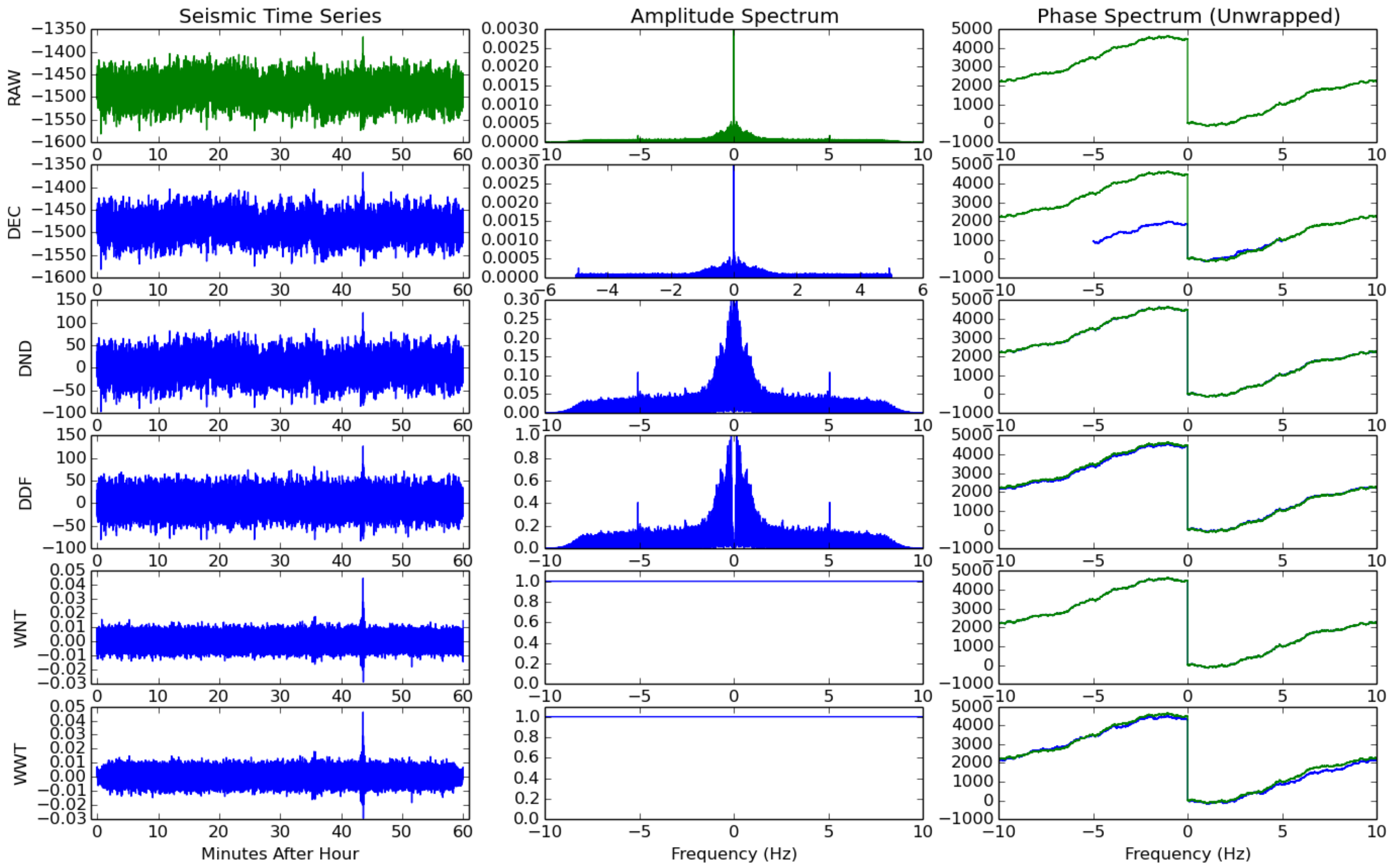

Figure 2.2.2: Time series, amplitude, and phase after applying various pre-processing steps to a raw signal. Top panel: sample of raw vertical-component HRSN seismic data recorded by station JCNB during January 29, 2003, between 10:00-11:00 UTC, and plotted in green (RAW). Second panel: raw signal decimated by a factor of 2 (DEC). Third Panel: raw signal de-trended and demeaned (DND). Fourth Panel: raw signal de-trended, demeaned, and high-pass filtered (DDF). Fifth panel: raw signal whitened without pre-emptive tapering (WNT). Sixth Panel: raw signal pre-emptively tapered and whitened (WWT). Uncorrected seismic time series showing V/m/s. Amplitude spectra normalized by their own maxima. Phase spectra in radians. The original phase spectrum is plotted on top of each resulting phase spectrum, in green, for ease of comparison.

achieve a similar result (Figure 2.2.2, panel 6). The reason for this is that by discarding raw amplitude information, the problem of regularizing raw signal amplitude disappears. Furthermore, upon comparing the fourth and sixth panels of Figure 2.2.2, we also note the presence of a low-magnitude seismic event in both time series, which is more clearly evident in the whitened time series with respect to background noise level. This event corresponds to a M1.2 earthquake located 21 km ESE of Anza, CA, which occurred at 10:43:21 UTC. The reason that low-magnitude events are enhanced via phase reconstruction, is again due to the fact that by discarding amplitude information, disparate signal amplitudes become effectively equalized. Therefore, not only does this whitening method achieve optimal whitening and satisfactory signal regularization with minimal data processing, it also enhances the presence of low-magnitude seismic events, all the while minimally altering raw signal phase.

Instrument Response

So far in our data pre-processing workflow we have neglected to discuss removing instrument responses from signals, this is because in our analysis we are able to ignore the effects of the instrument response for the following reasons. The purpose of removing the instrument response from a seismic record is to remove the effects of signal distortions caused by measuring the signal by a band-limited instrument, and also to obtain the signal in terms of intrinsic ground motion units (displacement, velocity, acceleration). However, when cross-correlating ambient seismic noise data, signals units are irrelevant and signal distortion can be overlooked. The crucial aspect is whether signals recorded by different stations have the same instrument response. If stations being used to estimate Green's functions each have the same instrument response, then signals, being identically representative of seismic motion from different stations, will converge to an accurate Green's function estimate when cross-correlated regardless what motion the signals constitute. In other words, the cross-correlation of the same instrument response (i.e., an auto-correlation) will essentially appear as a zero-phase signal convolved with the Green's function (recall Equation (2.1)). In the case of the HRSN, instrument response functions, specifically gain factors, and hence amplitude responses, did change throughout 2003. However, since we neglect signal amplitudes when whitening, and instead considered only raw signal phase, the problem of differing instrument responses vanishes, given that the phase responses of each HRSN remained unchanged during this period.

Considerations of Time-Domain Normalization Techniques and Data Rejection

Following the whitening of all our data, rather than performing a time-domain normalization procedure, such as a truncation scheme, event muting, or 1-bit normalization (Bensen et al., 2007), we opt to classify and segregate hour-long seismograms from our data set, based on whether they contain a seismic event, or whether we observe instrumental irregularities. The reason for this decision is fourfold.

First, we note the abundance of tremors in the Parkfield region, which can be thought of as quasi-continuously repeating, and therefore highly coherent low-amplitude seismic events (Nadeau and Guilhem, 2005). Thus, while clipping seismograms by either a multiple of the standard deviation or RMS diminishes the contamination to cross-correlation results from large-amplitude events (Bensen et al., 2007), such methods may not be appropriate when dealing with stations in the Parkfield region, given that tremors may survive truncation methods, and given that the impacts of tremor signals on ambient noise cross-correlation results have not been thoroughly assessed. We speculate that because tremors repeat at stable locations, cross correlating time series concurrent with tremor activity will likely result in coherent cross-correlation signals, which would rule out the use of ambient noise interferometry.

Second, although 1-bit normalization removes the amplitude information of all seismic events from prospective time series, this operation significantly alters raw signal phase, as seen in Figure 2.2.3, and subsequently is undesirable given our whitening method. Likewise, event muting drastically alters signal phase as well. Seats et al. (2012) also stress the importance of preserving phase, noting that the preservation of raw signal phase is quintessential to obtaining accurate and unbiased travel-times measurements from cross-correlation results.

Third, while we choose to use HRSN data from 2003 due to lower levels of tremor activity during this time period, as well as few mainshock and related aftershock events, that is, prior to the San Simeon earthquake, many of the HRSN stations exhibit instrumental irregularities coincident with this time frame. Therefore, rather than clip, mute, or 1-bit normalize data recorded during intervals where stations are not operating optimally, subsequently contaminating our cross-correlation results, we instead identify and reject such signals.

Fourth and finally, since whitened seismograms enhance the presence of low-magnitude seismic events, searching for events among whitened seismograms increases overall event

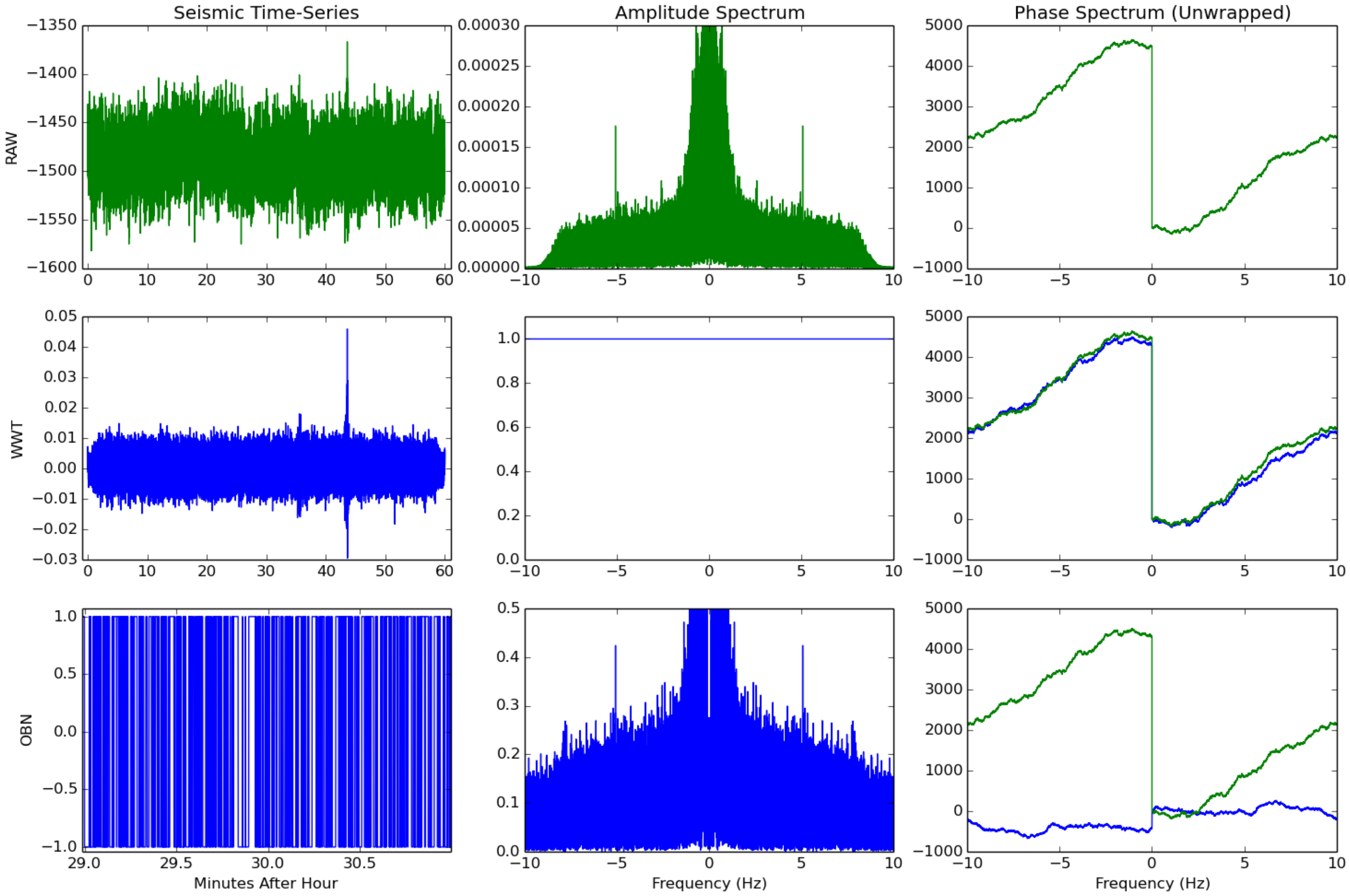

Figure 2.2.3: The resulting time series, amplitude, and phase after manipulations of a raw signal. Top panel: sample of raw vertical-component HRSN seismic data recorded by station JCNB during January 29, 2003, between 10:00-11:00 UTC, and plotted in green (RAW). Second panel: raw signal pre-emptively tapered and whitened (WWT). Third panel: raw signal 1-bit normalized and zoomed in for effect (OBN). Uncorrected seismic time series showing V/m/s. Amplitude spectra normalized by their own maxima. Phase spectra in radians. The original phase spectrum is plotted on top of each resulting phase spectrum, in green, for ease of comparison.

detection. Therefore, an improved detection rate with respect to low-magnitude events gives confidence that as many seismic events as necessary can be removed from the final data set to be cross-correlated.

In order to identify hour-long seismograms in the data set containing either events or irregularities, we use a combination of automated and manual identification schemes. We first operate automatically, segregating seismograms coincident with those tremor events tabulated by Nadeau and Guilhem (2009) occurring during 2003. In this way we ensure that as many known tremor events as possible are automatically removed from the data set. We then operate manually, performing the necessary component rotations to bring each station into the same ZNE reference frame. Once rotated, we plot 3-component, whitened, hour-long seismograms across the network. As stated, plotting whitened seismograms across the network aids manual event detection since low-magnitude seismic events are thus enhanced with respect to background level noise.

To be classified as an event, potential events need to be observed on at least three stations in the network. An example of a manually detected event, according to this scheme, is shown in Figure 2.2.4, in which an event occurring at approximately 52 minutes is visible on most of the components in the network. This event corresponds to an M3.5 earthquake which occurred 13 km southwest of Santa Barbara Island, CA, at 00:51:44 UTC. Likewise, a null detection is shown in Figure 2.2.5, in which the components of station VCAB seem to exhibit two distinct events at approximately 28 and 48 minutes respectively, and which are in fact not observed by any of the other HRSN stations. Hour-long time series containing manually detected events have been catalogued and are given in Appendix D. From the catalogue of manually detected events, we then segregate from the data set all components operating during those hours in which events were detected.

Lastly, being confident that as many seismic events as possible have been removed from the data set, series of seismic-like spikes and swells remaining in the set are therefore most likely anomalous or irregular. Thus, we perform an automated pass on each remaining seismogram in order to detect such patterns of behaviour. This process involves assessing station components separately since, for a given station, some components may be operating irregularly while the others are not. Such an example of non-uniform behaviour among the components of HRSN station GHIB is shown in Figure 2.2.6, in which the behaviour of the BP3 component is aberrant.

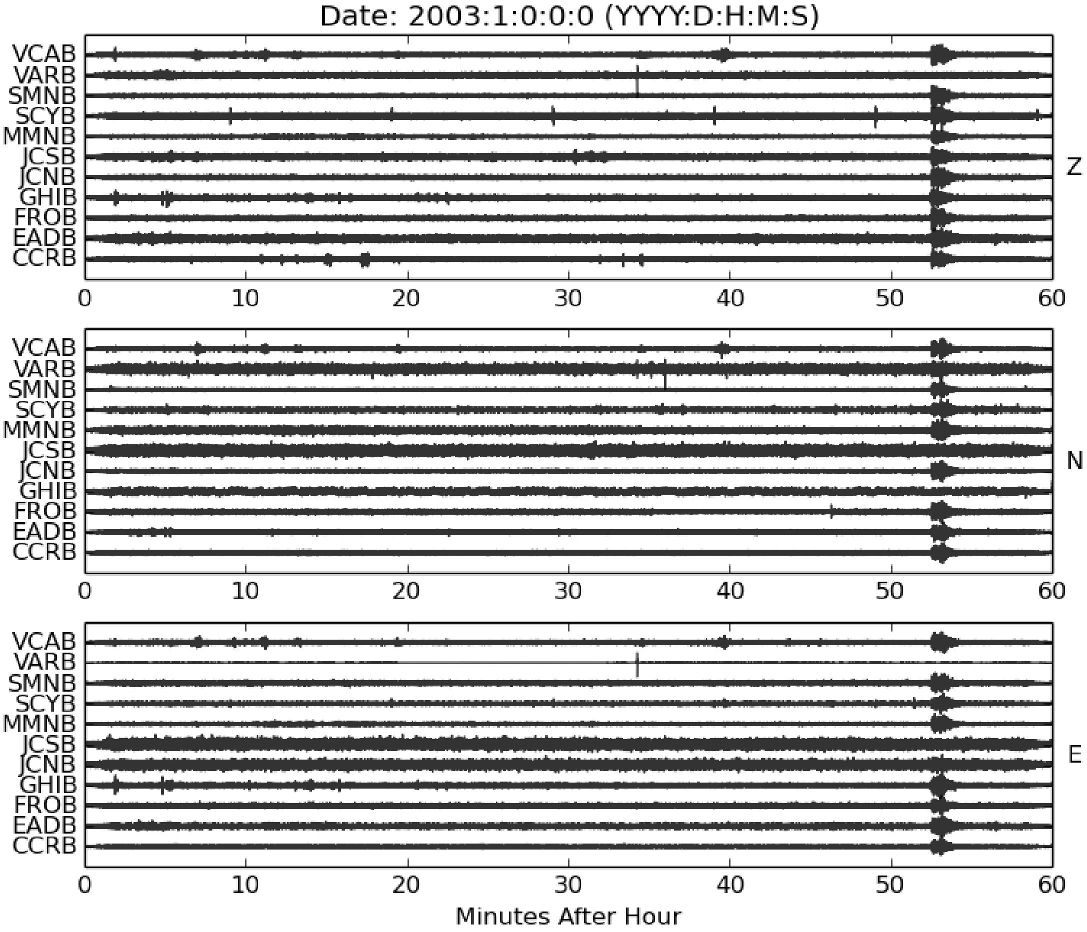

Figure 2.2.4: Sample event detection using whitened and rotated HRSN components. An event is detected at approximately 52 minutes, corresponding to an M3.5 event detected 13 km SW of Santa Barbara Island, CA.

In order to assess HRSN components for instrumental irregularities we perform two tests on each hour-long component remaining in the data set. The first test involves computing a kurtosis function for each prospective signal. Kurtosis, defined as the fourth standardized moment of a statistical distribution, is expressed mathematically as

$$Kurtosis[X] = \frac{\mu_4}{\sigma^4}, \tag{2.2}$$

where μ_4 represents the fourth moment about the mean and σ is the standard deviation. The kurtosis of a random variable X with an arbitrary distribution is a measure of the

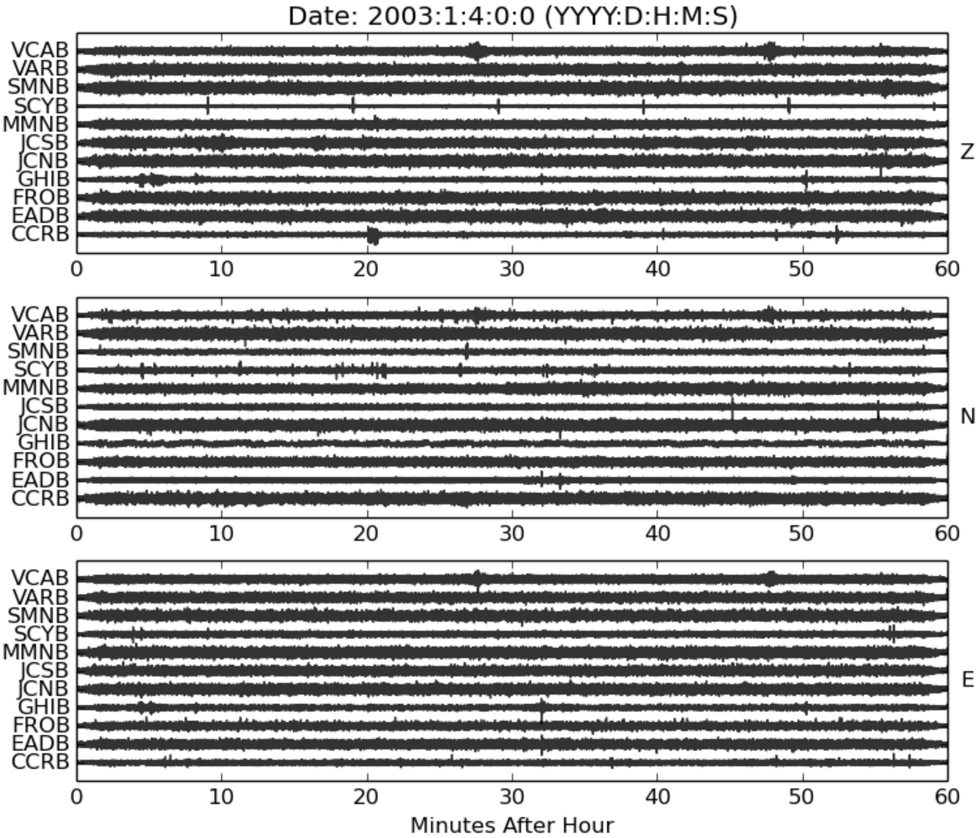

Figure 2.2.5: Sample null detection using whitened and rotated HRSN components. The components of station VCAB seem to show two event-like waveforms at approximately 28 and 48 minutes respectively, which are not observed by any of the other stations.

peakedness of that distribution with respect to the shape of a normal distribution. Thus, large positive values of kurtosis indicate a heavily peaked distribution, and therefore, this statistic is a suitable candidate for detecting large anomalous spikes present in the time series of HRSN components. We compute a kurtosis function by breaking the signal into several windows of 50 samples, each overlapping by 1 sample. We then compute a kurtosis value for each window and stitch these values together to form the kurtosis function. Overlaying time series with their kurtosis functions, one can easily associate anomalous spikes in time series with large values of the computed kurtosis function. In this way we test whether prospective time series pass or fail a kurtosis test, wherein failure is achieved when the maximum of the kurtosis function is greater than a certain threshold (specified momentarily).

Likewise, for each time series remaining in the data set, we also compute the unique short-

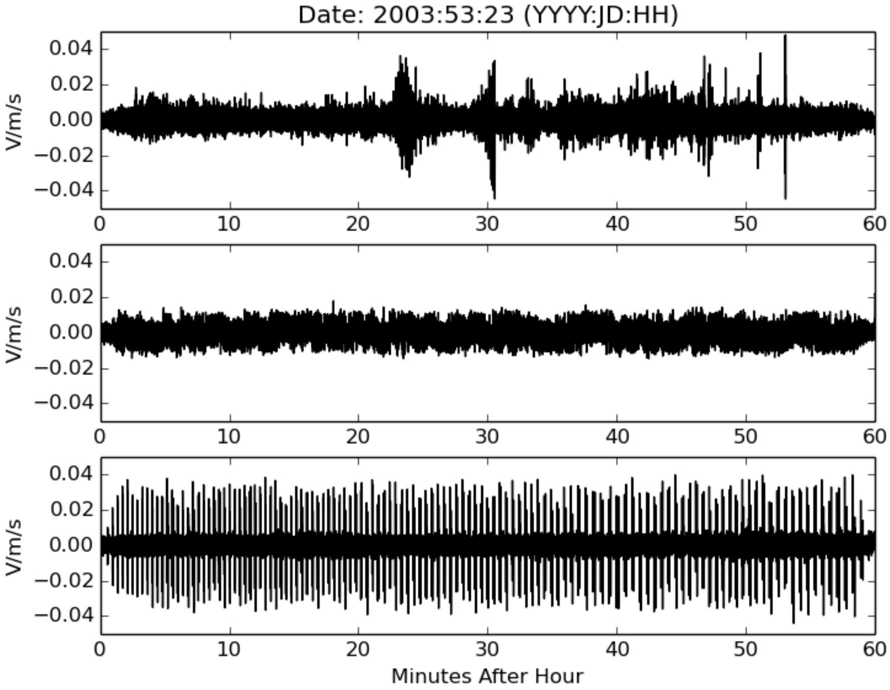

Figure 2.2.6: An example of non-uniform behaviour among components of HRSN station GHIB.

term-average/long-term average (STA/LTA) function using short and long window lengths of 200 and 35000 samples, respectively (10 seconds and 29.16 minutes). Ordinarily, computing the STA/LTA function with a long window length approximately half the length of the time series being tested means that the STA/LTA function can only be computed over the second half of time series. However, we can still compute a unique STA/LTA function over the entire time series, by overlaying the time series with two STA/LTA functions, one computed over the signal and the other computed over its time-reversed version, which overlap slightly. The STA/LTA function is sensitive to any event-like wave trains present in seismic data, due to the fact that abrupt short-term increases in the average energy of the time series will result in large values of the STA/LTA function. Therefore, we test whether the maximum of the computed STA/LTA function of each time series is also beyond a selected threshold, chosen independently of the kurtosis threshold (specified subsequently).

We test time series for irregularities using both methods as they are independent and complementary. A given signal may pass the kurtosis test while yet containing abnormal characteristics. Likewise, a given signal may pass the STA/LTA test while yet remaining unqualified for ambient noise cross-correlations. However, to catch the bulk of misbehaving seismograms remaining in the data set, it is sufficient to perform both tests, such that a time-

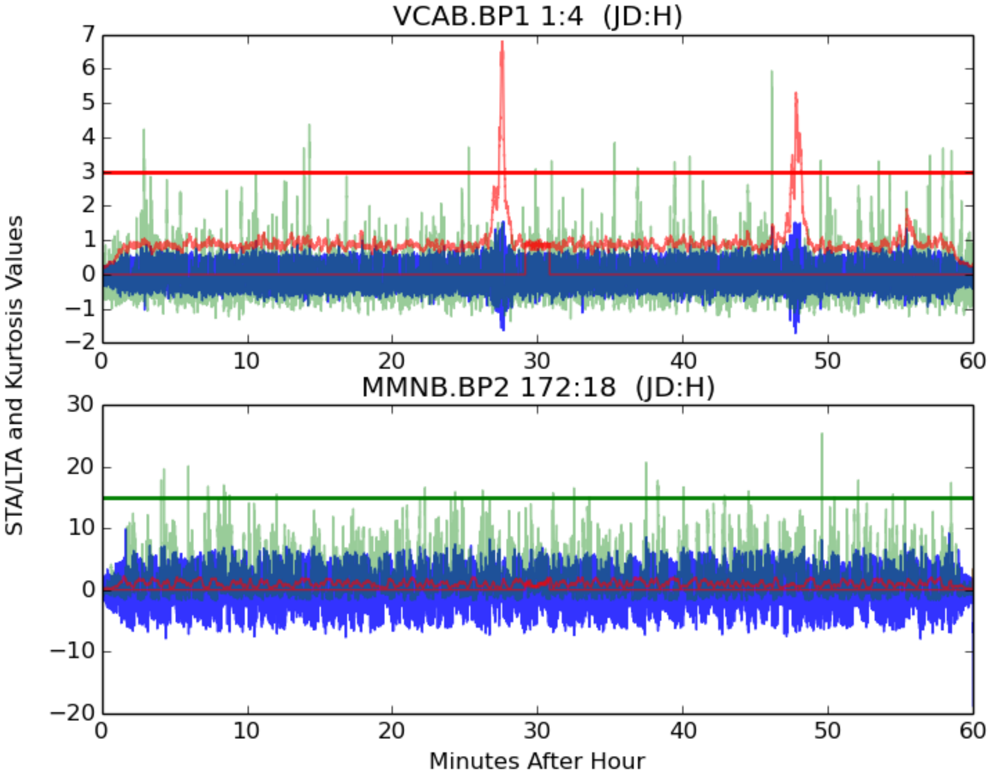

Figure 2.2.7: Detecting instrumental irregularities among individual HRSN components. Top panel: time series recorded by HRSN station VCAB during January 1, 2003 between 04:00-05:00 UTC and plotted in blue. Associated kurtosis and STA/LTA functions overlain in green and red, respectively. Bottom panel: time series recorded by HRSN station MMNB during June 23, 2003 between 18:00-19:00 UTC. Associated kurtosis and STA/LTA functions overlain in green and red, respectively. Note that the time series in the bottom panel has been zoomed in to display its uncharacteristic noise features. Thresholds in each figure plotted as thick horizontal lines, coloured green for kurtosis and red for STA/LTA.

series that fails one or the other is segregated from the final data set. Samples of each of these cases are shown in the panels of Figure 2.2.7. The signal in the top panel of Figure 2.2.7 passes the kurtosis test but fails the STA/LTA test, whereas the opposite situation occurs in the the bottom panel. By considering a variety of signals from each of the HRSN stations, we choose static thresholds of 15 for the kurtosis and 3 for the STA/LTA. To be clear, these thresholds are chosen arbitrarily, such that the pass/fail values to be assigned automatically correspond to those that were assigned the vast majority of the time upon visual inspections of time series.

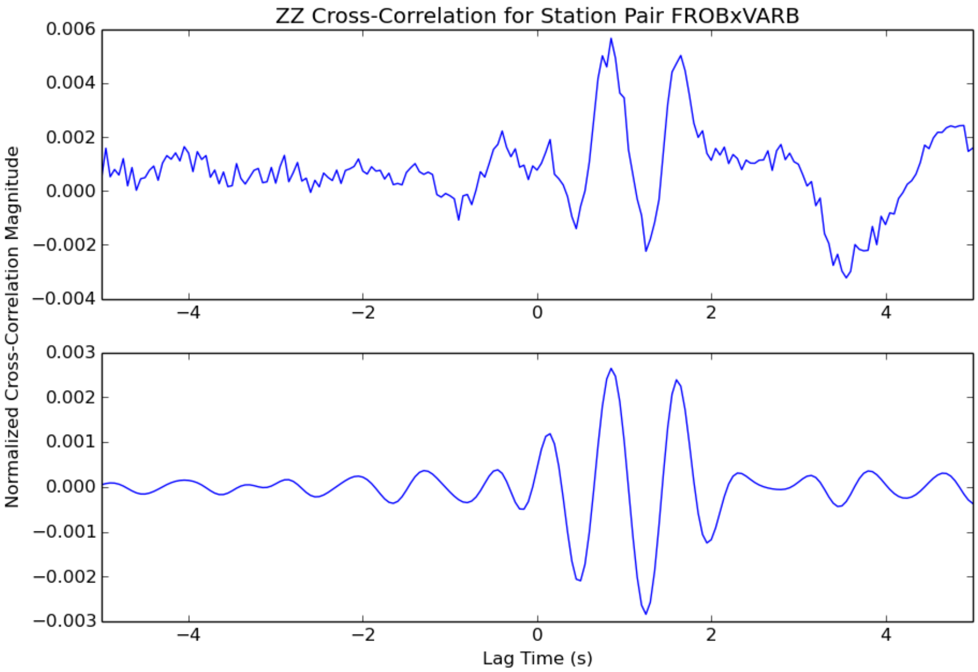

Figure 2.2.8: Noise-based Green's function estimate for HRSN stations FROB and VARB. Top panel: unfiltered, normalized cross-correlation function. Bottom panel: signal bandpass filtered between 1 and 2 Hz. Inter-station distance is 3.97 km.

Vertical Component Cross-Correlation Tests

Carrying out the quality control procedures described in § 2.2.2 results in the creation of four mutually exclusive data sets, to one of which each whitened HRSN signal belongs. With respect to the prescribed order, the first set contains seismograms recorded during times coincident with catalogued tremor activity. Likewise, the second set contains seismograms concurrent with manually detected events, whereas the third and fourth sets consist of signals that pass and fail the combined kurtosis and STA/LTA test, respectively. In other words, the third set consists of signals identified as exclusively containing seismic noise, while the fourth consists of noise recordings interspersed with instrumental irregularities. Using these data sets, we produce four separate Green's function estimates for each station pair in the network. Doing so allows us to compare cross-correlations of signals containing tremors, standard seismic events, pure noise, and instrumental irregularities, respectively.

To produce a Green's function estimate for a single station pair, we cross-correlate hour-long, whitened, time-matching segments between the pair and then stack each temporally matched cross-correlation linearly to produce a final cross-correlation result. A sample cross-

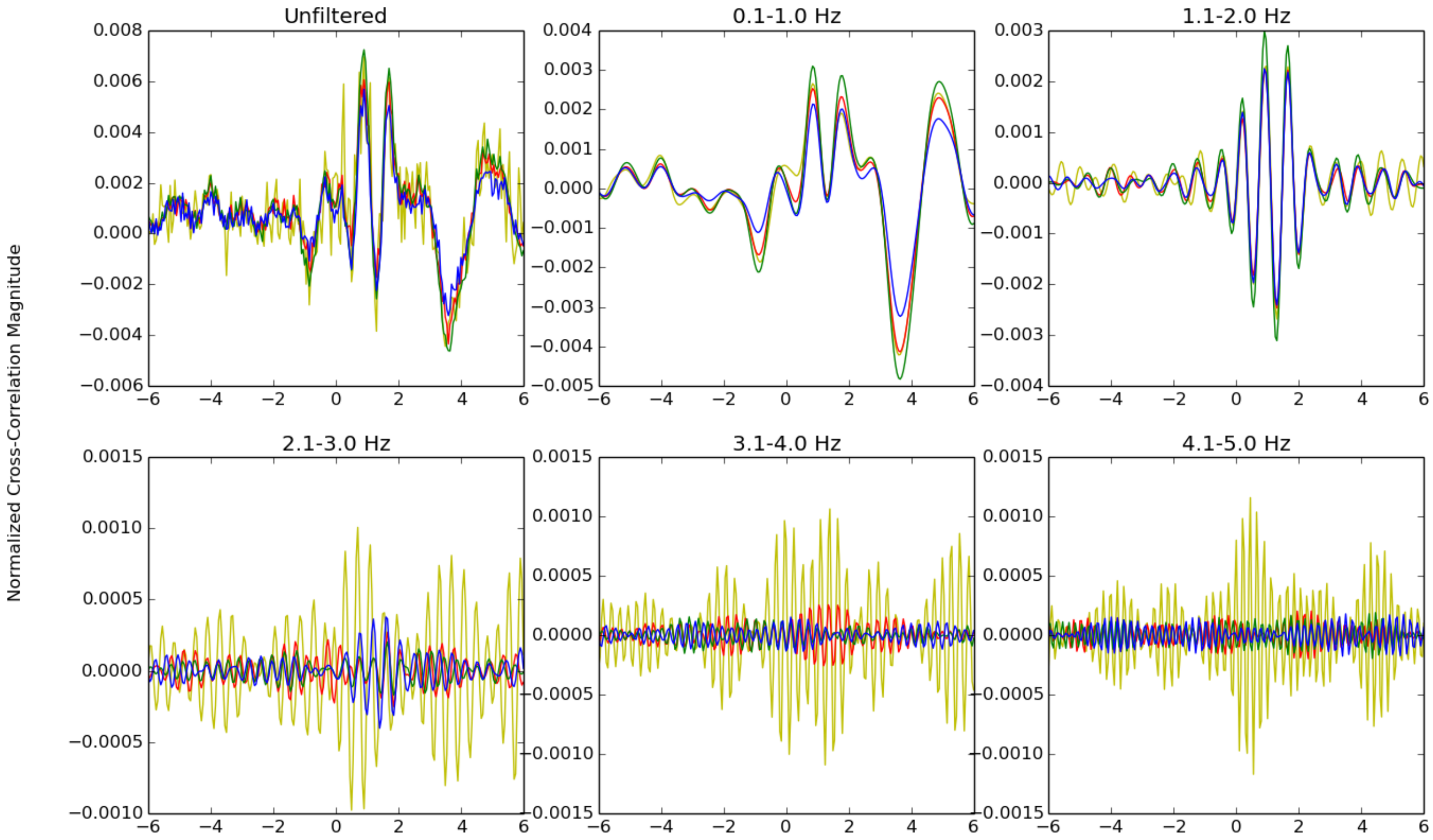

Figure 2.2.9: A comparison of vertical cross-correlation results for HRSN stations FROB and VARB. Results in each frame are bandpass filtered into frequency bands 0.9 Hz wide. Yellow: results produced from the tremor data set. Red: results produced from signals concurrent with manually detected events. Green: results produced from signals interspersed with instrumental irregularities. Blue: results produced from pure noise. Inter-station distance is 3.97 km.

correlation for stations FROB and VARB, produced using the noise data set, is shown in Figure 2.2.8. To compare cross-correlation results produced from each of these data sets, we take results for a particular station pair, filter them within a 0.9 Hz frequency band, overlay each atop one another in the same frame, and repeat for nine distinct frequency intervals. Such a comparison is shown in Figure 2.2.9. Upon examining Figure 2.2.9, we notice that for frequencies below 2 Hz, cross-correlation results are largely the same, apart from minor variations in cross-correlation magnitude. However, beyond 2 Hz, cross-correlation results begin to differ not only in magnitude but in phase as well, with the most drastic variations coming from results produced using the tremor data set. Furthermore, results produced from the tremor data set are consistently the largest magnitude cross-correlation results beyond 2 Hz. Finally, examining Figure 2.2.9 we also notice that the magnitude of noise based cross-correlation results diminishes significantly at frequencies beyond 2 Hz, suggesting that above 2 Hz, body wave signals are much weaker.

Noise Beamforming Analysis

Although we compute ambient noise cross-correlations for each station pair in our network, we must assess the spatial distribution of the noise field incident upon the HRSN during 2003 to determine which of our station pairs conform to the theoretical requirements outlined in § 2.2.1. To do so, we adopt the plane wave beamforming analysis used by Roux (2009) in order to determine the dominant incidence angle of noise propagating among HRSN stations. The beamforming analysis consists of computing the following equation:

$$B(\theta, c) = \frac{1}{\Delta\omega} \int_{\omega_1}^{\omega_2} |\sum_{n=1}^{N} S_n(\omega) e^{i(\frac{\omega}{c}(x_n sin\theta + y_n cos\theta))}|^2 d\omega. \tag{2.3}$$

Equation (2.3) is an array method, which calculates the total energy recorded by an array of N receivers. This is done by summing the energy recorded at each receiver, positioned at coordinates (x_n, y_n) with respect to the centre of the network, due to monochromatic plane waves of angular frequency ω, propagating at phase velocity c, and incident upon the array at angle θ (see Rost and Thomas, 2002 for a detailed account). Finally, $S(\omega)$ represents the complex Fourier component of the time series recorded by the n^{th} receiver. Since Equation (2.3) integrates the array energy over a range of frequencies within a frequency bandwidth of $\Delta\omega = \omega_2 - \omega_1$, it is sometimes referred to as frequency-incoherent beamforming.

All seismic noise, whether caused by ocean waves and atmospheric effects, or local sources such as human activity, dominantly occurs within a narrow frequency band of 0.05-0.2 Hz (Stein and Wysession, 2003). While we seek to recover body waves at frequencies generally beyond 1 Hz through use of ambient noise cross-correlations, it is important to note that we beamform to locate the sources of noise, which, propagating as surface waves, produce body waves mainly through surface-to-body wave scattering. Accordingly, we choose to compute the beamforming function for the HRSN by summing 50 frequencies over the interval 0.05-0.2 Hz exlusively using unprocessed, vertical-component, noise seismograms. Due to the fact that computing the beamforming function can be computationally expensive, rather than beamform the network over the entire year, we choose to beamform 24 distinct hours of noise data recorded by the network and chosen randomly. The beamformer output is displayed in Figure 2.2.10, from which we note a dominant incidence angle of approximately 50 degrees, and weighted toward lower phase velocities. Therefore, Figure 2.2.10 clearly indicates that the noise field incident upon the HRSN during 2003 originated from a backazimuth of 230 degrees and consisted mainly of surface waves propagating at phase velocities ranging from 2-3.5 km/s. Note that this result is in agreement with similar analyses performed on other networks in the Parkfield region, all of which are consistent with seismic noise fields originating from the Pacific ocean (Roux et al., 2005 and references therein; Roux, 2009).

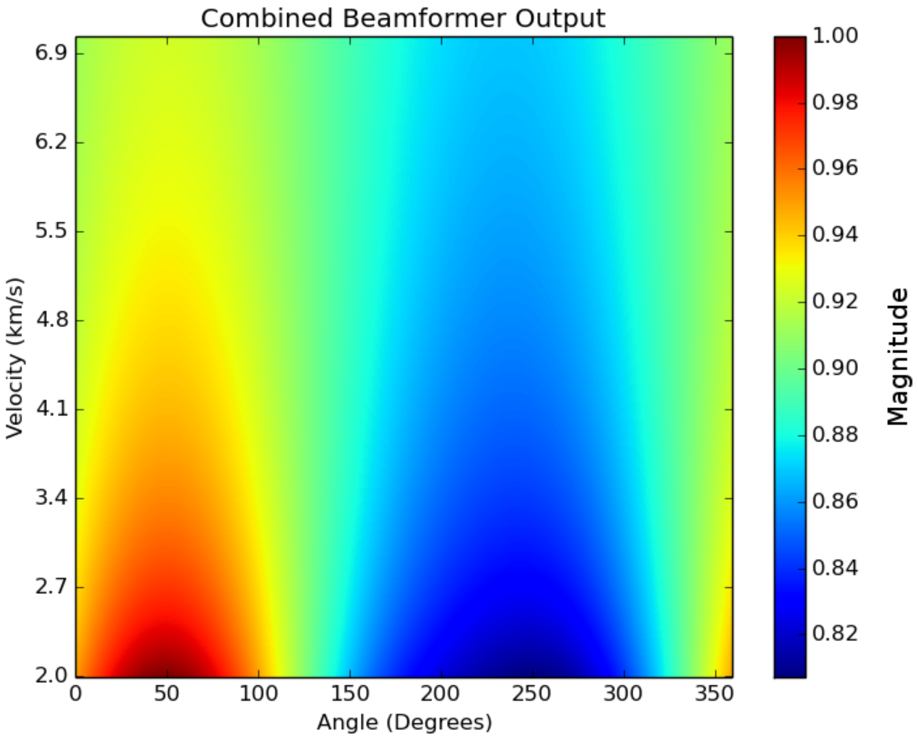

Figure 2.2.10: Result of plane wave beamforming, applied to 24 distinct, unprocessed, network wide HRSN noise recordings chosen at random.

2.3 Results

2.3.1 Vertical Component Cross-Correlation Results

Recalling that we wish to extract P-waves from ambient seismic noise data, we cross-correlate the vertical components (Z-Z) of 12 HRSN stations, which gives a total of 66 potential Green's function estimates. However, due to the observed heterogeneity of the incident noise field during 2003, not all cross-correlation results are reliable (§ 2.2.1). Consequently, consistent with our beamforming result, we reduce 66 vertical-component cross-correlation results to a subset of 17 for which station pairs are constrained to have inter-station orientations within 50 degrees of the dominant incidence angle of the noise field.

These two cases are presented in Figure 2.3.1 and Figure 2.3.2 , respectively. Upon viewing Figure 2.3.1, we note that there are no discernible move out trends among cross-correlation results, for which coherent seismic phases seem absent. However, in Figure 2.3.2 we do see a clear trend among cross-correlation results, moreover, one that is consistent with a

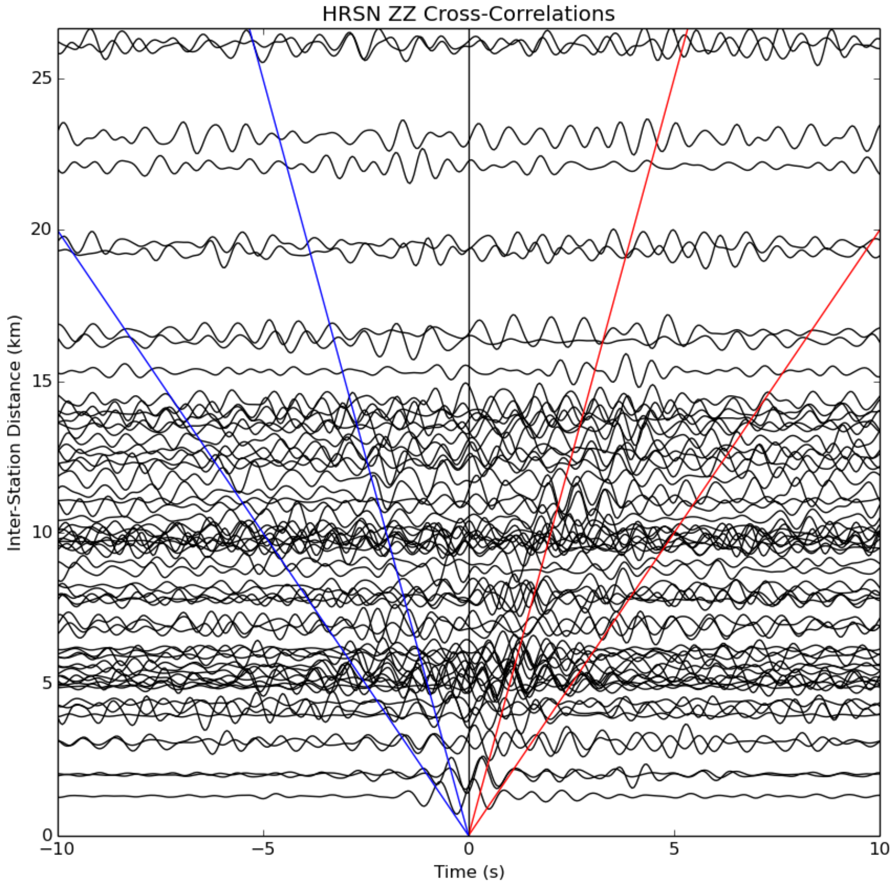

Figure 2.3.1: All HRSN vertical-component cross-correlation results arranged by inter-station distance. Results are filtered between 1-2 Hz and are normalized by their own maxima. Causal move out lines plotted in red for velocities of 5 and 2 km/s, respectively. Corresponding acausal move out lines in blue.

move out expected for P-waves travelling along straight rays at a speed of 5 km/s. Thus, given that the results displayed in Figure 2.3.2 are filtered between 1-2 Hz, and that their apparent propagation velocity is 5 km/s, these results strongly suggest the detection of P-waves propagating between HRSN station pairs.

Note that knowledge of the backazimuth of the noise field allows us to orient cross-correlation results in a consistent manner. This is possible since, in the presence of a heterogeneous noise field, ambient noise cross-correlations are asymmetric with respect to the zero lag-time. We choose to arrange cross-correlation results such that the causal parts represent waves propagating from northeast to southwest.

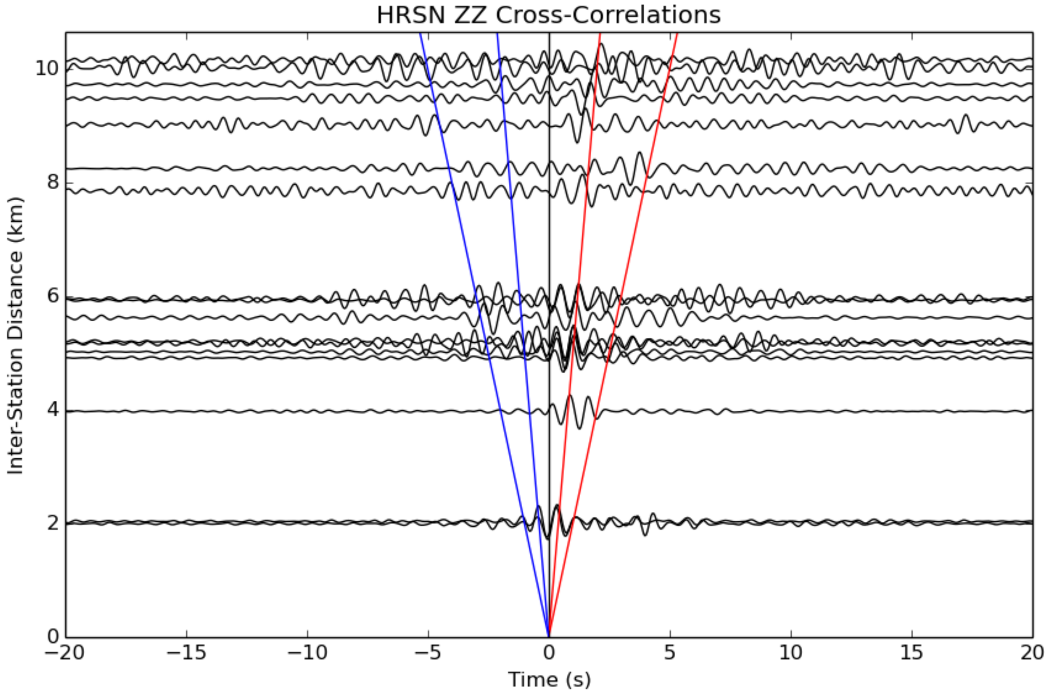

Figure 2.3.2: All HRSN vertical-component cross-correlation results arranged by inter-station distance. Results are filtered between 1-2 Hz and are normalized by their own maxima. Causal move out lines plotted in red for velocities of 5 and 2 km/s, respectively. Corresponding acausal move out lines in blue.

2.3.2 A Preliminary Step Toward Body Wave Tomography

Using the 17 cross-correlation results discussed in § 2.3.1, we may obtain a simple crustal P-wave velocity map of the SAF near Parkfield, CA. As inter-station distances are known precisely, we may estimate P-wave travel-times between each associated station pair. Such a procedure is akin to a first order approximation of the crustal P-wave velocity structure in this region and represents a preliminary step toward a tomographic procedure. However we stress that this result is based on a straight ray approximation and does not take into geometric spreading, scattering, or multipathing effects.

In § 2.1.1, we mentioned that the Green's function between two receivers can be approximated by the ensemble average of cross-correlations of the ambient seismic noise field recorded at both receivers (Equation (2.1)). A more precise version of this statement is that the cross-correlation of ambient noise recorded between two receivers approximates the Green's function for the receiver pair, convolved with the auto-correlation of the noise field

(Wapenaar, 2004). In this case, we write

$$G(A, B, t) * N(t) \approx \langle V(A, t) \otimes V(B, t) \rangle \tag{2.4}$$

which is Equation (2.1), only, modified on the left hand side to include the convolution, represented by the $*$, of the Green's function with the auto-correlation of the noise field, represented by $N(t)$. Therefore, according to Equation (2.4), if we assume that the Green's function for the receiver pair takes the form of a sequence of delta-like impulses, with varying amplitudes and polarities, such that the position of each pulse represents the arrival time of a unique seismic phase, then the maximum of the resulting convolution will approximate the precise location of the first body wave arrival, or, direct P-wave, since the auto-correlation is zero-phase (Figure 2.3.3). In other words, the arrival time of the direct P-wave propagating between a given station pair can be picked from the maximum of the cross-correlation function for that pair rather than the first break. Note that this is also the method used by Nataka et al. (2015) to pick P-wave arrival times from ambient noise cross-correlations. As noted by Nataka et al. (2015), while this method of picking P-wave arrival times is justified, it is not immune to error, especially when convolving finite bandwidth signals. In this case the convolution operation is prone to producing strong side lobes on either side of wave arrivals in ambient noise cross-correlation results. Therefore, in addition to picking P-wave arrival times based on the maximum of cross-correlation functions, we also take into account travel-time estimates of nearby station pairs to produce consistent estimates when appropriate. Finally, to estimate P-wave arrival times as accurately as possible, we also take into account the three-dimensional spatial arrangement of HRSN stations when calculating inter-station distances. While deviations between the two-dimensional surface projection of inter-station distances and the true distances are minimal in this case, the correction is nonetheless straightforward to implement.

Performing the above steps, we obtain 17 P-wave travel-time and velocity measurements between various HRSN station pairs, which are listed in Table 2.3.1. Among these results two velocity estimates appear anomalously slow for P-waves, namely, the velocity estimates of the direct arrivals between stations RMNB and CCRB, and stations RMNB and FROB. In the case of the first station pair, if we instead base our travel-time pick on the third highest peak in the cross-correlation result we obtain a more reasonable velocity estimate. However, due to the distance between the first and third peaks in this cross-correlation result we exclude this result from the final analysis. In the second case, we attempt a travel-time

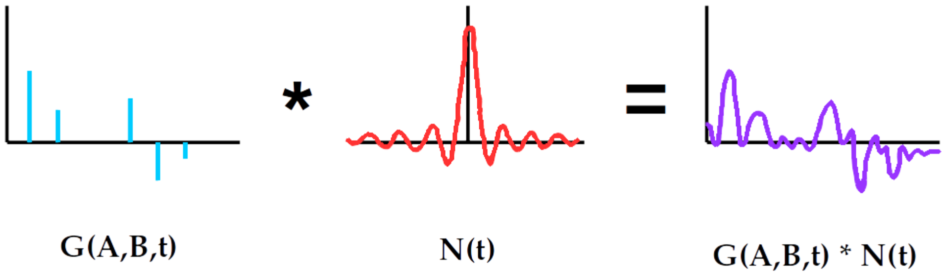

Figure 2.3.3: Cartoon representation of convolving the Green's function between receivers at positions A and B with the noise field autocorrelation function.

Station Pair	2D Inter-Station Distance (km)	3D Inter-Station Distance (km)	Travel-Time Estimate (s)	Velocity Esitmate (km/s)
FROB x JCNB	5.9230	5.9230	1.20	4.93
FROB x JCSB	4.9097	4.9101	0.95	5.16
FROB x VARB	3.9687	3.9689	0.85	4.66
JCNB x JCSB	1.9922	1.9937	0.30	6.64
JCNB x VARB	2.0328	2.0336	0.35	5.81
JCNB x VCAB	9.4755	9.4781	1.70	5.57
JCSB x VCAB	9.0163	9.0213	1.50	6.01
MMNB x CCRB	5.0160	5.0171	1.00	5.01
MMNB x RMNB	5.1956	5.2162	1.00	5.21
MMNB x VCAB	5.1600	5.1663	1.00	5.16
RMNB x CCRB	8.2420	8.2616	3.75 (1.90)	2.20 (4.34)
RMNB x FROB	10.0206	10.0416	3.40 (2.60)	2.95 (3.86)
RMNB x SMNB	9.7225	9.7362	1.90	5.12
RMNB x VCAB	10.1549	10.1631	1.90	4.72
SCYB x CCRB	5.9479	5.9478	2.15	4.96
SCYB x SMNB	5.6179	5.6258	1.20	5.11
VCAB x VARB	7.8495	7.8545	1.35	5.81

Table 2.3.1: Travel-time and velocity estimates from ambient noise cross-correlation results.

pick based on the closest peak to the positive 5 km/s move out line. However, we feel that this new value, independently of producing an unreasonable velocity estimate, is forced, and we therefore reject this result from our final analysis as well. Nonetheless, the alternate travel-time picks and velocity estimates for these station pairs are displayed in Table 2.3.1 in brackets, and furthermore, each travel-time pick between station pairs listed in Table 2.3.1 is displayed in Figure 2.3.4.

Among the selected set of 15 results the average estimated P-wave velocity is 5.32 km/s with a standard deviation of 0.54 km/s. Furthermore this average is consistent with the results of Roux et al. (2005) who also observed P-waves propagating at approximately

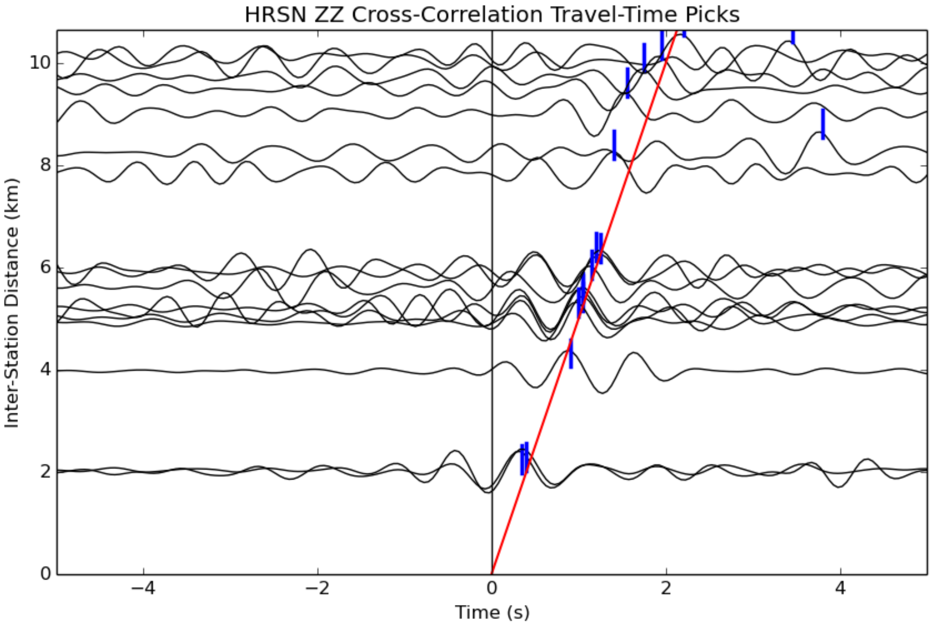

Figure 2.3.4: Travel-time picks for each of the 17 results are listed in Table 2.3.1. Picks are indicated by the blue tick marks. A 5 km/s moveout line is plotted in red for reference.

5 km/s from ambient noise cross-correlations, but among a different network of stations in the Parkfield region. From the final set of 15 travel-time and velocity measurements, we plot estimated P-wave velocities on a map (Figure 2.3.5). The velocities displayed in Figure 2.3.5 are uniform overall; however, we identify two noteworthy patterns. First, we note that velocity lines radiating outward from station VCAB increase with clockwise rotation. Secondly, from the results displayed, we also note that faster velocities seem to occur in the southern portion of the network, that is, among stations south of JCNB, whereas velocities north of JCNB tend to slower values.

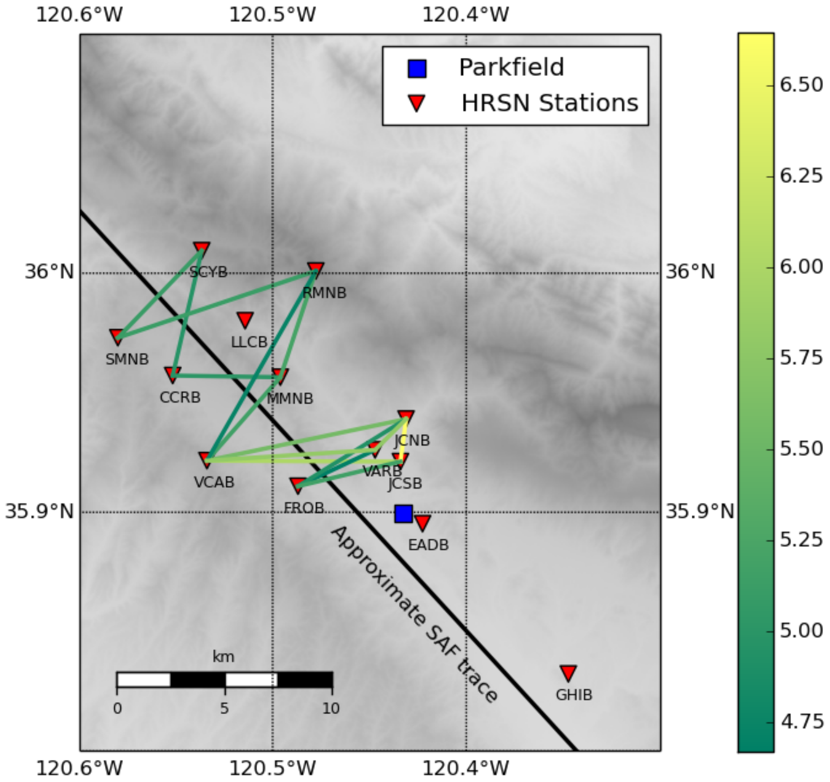

Figure 2.3.5: Crustal P-wave velocity map of the SAF near Parkfield, CA, derived from 15 ambient noise interferometry results and based on a straight ray approximation. Velocity units are km/s.

2.4 Discussion

2.4.1 Comparison of Results and Limitations

We compare the results obtained in this study and plotted in Figure 2.3.5 to a recent and comprehensive, large scale, three-dimensional tomographic model of the SAF proposed by Zeng et al. (2016) for P- and S-wave velocities. While Zeng et al. (2016) demarcate several high velocity bodies and low velocity zones resolved by their model, their region of study is considerably larger than ours, and therefore, the map boundaries in Figure 2.3.5 represent a small subregion within their model (approximately within Y nodes [-9,3] and X nodes [-8,5], see Zeng et al. (2016)). Within this subregion containing the HRSN, their model reveals that the main velocity feature of the SAF is a simple velocity contrast across the fault, extending to 15 km in depth for P-waves, wherein waves propagating on the southwest side of the fault are generally faster than those on the northeast side. This velocity contrast corresponds to the major difference in geology between rocks of the dominantly granitic Salinian block on the southwest side of the fault and rocks of the sedimentary Franciscan Complex on the northeast side.

Apart from two anomalously slow velocity estimates, the remaining 15 P-wave velocities obtained in this study are in excellent agreement with this model, which, in the region of the HRSN, predicts velocities of 5-6 km/s for shallow propagating P-waves (less than 10 km depth) on the southwest side of the fault and velocities of 4-5 km/s for waves propagating on the northeast side. Furthermore, from the S-wave portion of their model, a mid-crustal, southwest dipping, low velocity zone, co-located with the northernmost HRSN stations is also imaged. This low velocity zone appears to bridge the SAF and may explain why velocities between more northern HRSN station pairs were found to be slower compared to stations further south. In addition to these bulk trends, an interesting small scale feature observed in Figure 2.3.5 is the increase in velocity with clockwise rotation for results radiating outward from station VCAB. Such a pattern likely represents either small scale structural heterogeneities with regard to the SAF or local seismic anisotropy.

While the velocity estimates of observed P-waves obtained in this study demonstrate that ambient noise interferometry is a viable approach for body wave tomography studies of the SAF in the Parkfield region, there are limitations. Most notably, the heterogeneity of the noise field severely limits the number of reliable cross-correlation results we can actually obtain. In our case, we are reduced from 66 potential Green's function estimates to 17 usable cross-correlation results, a reduction of nearly 75 %. Therefore the spatial arrangement of stations with regard to the ambient noise field is crucial for successful ambient noise tomography studies of the SAF in the future. A further, but less drastic limitation of this method is also encountered when estimating the maximum value of cross-correlation results in order to pick wave arrival times. When doing so, the accuracy of each travel-time pick is directly limited by the sampling rate of the data. This is graphically demonstrated in Figure 2.4.1 in which the true time of the cross-correlation maximum is likely between 1.0 and 1.05 s. Thus, in the case of HRSN stations, travel-time picks are limited to 0.05 s increments, resulting in a velocity uncertainty of roughly 1 %. Additionally, we reiterate that the velocity results produced in this study and plotted in Figure 2.3.5 are based on a straight ray approximation, and therefore, a tomographic inversion procedure is necessary in order to account for effects

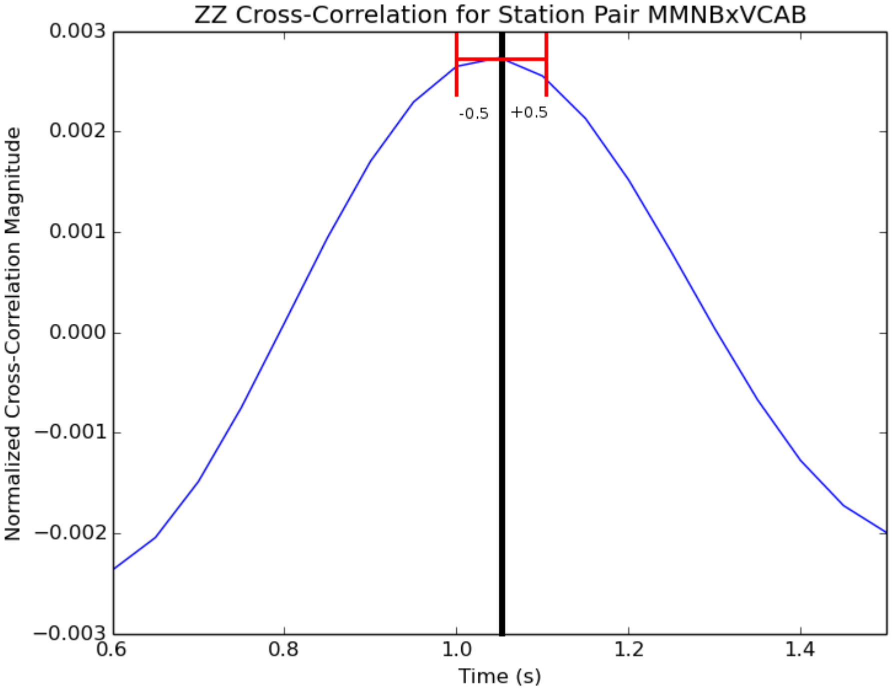

Figure 2.4.1: A graphical illustration of the origin of the uncertainty on velocity estimates obtained in this study. The travel-time pick for the station pair is plotted in black.

such as geometric spreading, scattering, or multipathing. Nonetheless, the major advantage of using ambient noise data to recover body wave travel-times for potential tomographic purposes still remains, in that travel-times obtained in this manner yield more accurate source positions than earthquake-based tomography models. Moreover, by attempting to map P-wave velocities, even assuming a straight ray approximation, patterns of seismic anisotropy and/or small scale structural heterogeneities already begin to emerge.

2.4.2 Data Rejection Volume vs Cross-Correlation Convergence

Further regarding our data pre-processing, rather than pursue a time-domain normalization technique, we implemented various quality control measures in an attempt to segregate pure noise data from data containing either seismic events or instrumental irregularities. However, this alternative approach to time-domain normalization raises an important question of its own. By partitioning HRSN data into four discrete sets, the size of each set, and ultimately, the set of pure noise, depends upon the strictness of the quality control measures implemented. Yet, it is generally known that the greater the number of cross-correlations

Data Set	Number of Hour-Long Segments	Proportion of Raw %
Raw	256,752	-
Manually Detected Events	111,591	43.46
Irregular	86,564	33.71
Noise	53,653	20.90
Catalogued Tremor	4,944	1.93

Table 2.4.1: Classification of 2003 HRSN data.

stacked between a given station pair, the better the convergence of the total cross-correlation function to the inter-station Green's function (Bensen et al., 2007). Therefore, it is natural to explore the point at which a reduction in the size of the data set to be cross-correlated negatively impacts potential Green's function estimates. To further examine this relationship, in Table 2.4.1 we present the breakdown of 2003 HRSN data after applying our quality control scheme, which effectively classifies approximately 80 % of all data from 2003 as unusable. Thus, it would seem that implementing such strict quality control measures severely limits the convergence rate of our cross-correlation results to the inter-station Green's function. However, Seats et al. (2012) find that the mean number of days each station pair requires to converge to within 90 % of its equivalent year-long noise correlation function depends on the pair distance. Accordingly, as a rule of thumb, for small scale networks such as the HRSN, cross-correlating non-overlapping, short window lengths on the order of 1 hour should not require more than five days of data to converge to 90 % of the equivalent year-long noise correlation function. Therefore, even though our quality control measures relegate most HRSN data from 2003 to non-noise categories, on average, we retain enough noise data between station pairs such that our noise based cross-correlation results converge to reasonably good Green's function estimates.

2.4.3 Cross-Correlation Variations as a Function of Frequency and Data Set

According to the theoretical requirements for ambient noise interferometry, outlined in § 2.2.1, seismic events and instrumental irregularities affect cross-correlation results in such a way as to nullify the ability of ambient noise interferometry to produce usable Green's functions estimates. Recall, therefore, that the purpose of time-domain normalization or data rejection schemes is to obviate the nullifying effects posed by such signals to cross-correlation results. However, we find that Green's function estimates, produced using different whitened data sets, some of which include these undesirable elements, do not vary significantly, neither in phase nor magnitude, when filtered below 2 Hz (Figure 2.2.9), whereas results were found to vary above this threshold and most notably for those estimates produced using the tremor data set. From these observations we draw a number of preliminary conclusions. First, due to the fact that the predominant frequency content of tremor signals ranges from 1-15 Hz (Nadeau and Guilhem, 2005), it seems natural that tremor signals do not impact cross-correlation results until higher frequency signals are considered. Therefore, tremor signals likely have minimal impact on surface waves obtained using ambient noise interferometry, as surface waves are low-frequency signals (<1 Hz). Second, given that the choice of data set does not significantly affect cross-correlation results filtered to frequencies below 1 Hz, this suggests that the only required step in estimating Green's functions for surface wave tomography may be to simply whiten signals by reconstructing them from their phase spectra. Third, since tremor signals significantly affect cross-correlation results at frequencies above 1 Hz, we conclude that accounting for tremor activity is necessary when attempting to perform ambient noise body wave tomography in regions where tremors occur. Furthermore, this places impetus on the need to identify and locate tremor events in regions where body wave tomography is to be performed using ambient noise interferometry.

2.5 Conclusions

The primary aim of this study has been to demonstrate improvements to ambient noise cross-correlation data processing methods, specifically, to those methods targeted toward the recovery of body waves propagating between receivers in the vicinity of the SAF, near Parkfield, CA. The cornerstone of our data processing workflow is the whitening method we employ, in which we reconstruct seismic records exclusively from their phase spectra. By means of this method we are able to circumvent issues of varying amplitude responses between stations, should they exist. Furthermore, by whitening signals this way, we are also able to pre-process data efficiently, since we can effectively replace routine data processing operations, such as de-trending, demeaning, and pre-filtering of signals with a single step. We have also demonstrated that this whitening method is effective at enhancing low-magnitude seismic events contained in signals. Apart from these peripheral benefits, our data processing workflow was ultimately developed to approximate inter-station Green's functions between seismic receivers via ambient noise interferometry. Given that Green's functions between receivers contain information concerning the arrival times of various seismic wave types, a phase-based reconstruction of seismic signals is a natural choice in order to spectrally whitening signals (see Appendix E).

Applying our data pre-processing methods and using as much noise data as possible recorded by the HRSN during 2003, we compute 66 inter-station Green's function estimates. However, due to the heterogeneity of the incident noise field during this time frame, and the theoretical requirement of a homogeneous noise field in order to perform ambient noise interferometry, our usable results are substantially limited from 66 to 17. We further reduce the number of results obtained in this study from 17 to 15 after carefully considering individual cross-correlation results. Finally, among these 15 results we observe P-waves propagating between various HRSN station pairs, for which we estimate travel-times and velocities. Using these these observed P-waves we compute an average velocity of 5.32 km/s with a standard deviation of 0.54 km/s, assuming a straight ray approximation.

The fundamental data for seismological studies of the earth's interior are the travel times of seismic waves.

Stein and Wysession (2003)

3

Conclusion

Seismic tomography procedures are among the most powerful methods we have for resolving structures within the Earth's interior. However, traditional tomographic methods are heavily constrained by spatially and temporally sporadic sources. In this research we have shown that by using ambient seismic noise data, one may potentially extend the effective reach of tomographic methods to further image small scale structures of the SAF to a high level of detail. However, the 15 P-wave results obtained in this study are not sufficient to do so. Thus, the next natural step would be to adapt the analysis employed in this study to other networks in the Parkfield region in order to obtain a greater volume of P-wave results. Such an approach would be a straightforward way to increase ray path coverage across the fault in order to develop a robust tomographic model from ambient noise interferometry. Beyond this, what remains to be seen is the possibility of observing higher frequency body wave phenomena from ambient noise cross-correlations, such as reflections and wave mode conversions (P to S). Being able to observe such phenomena through the use of ambient noise data would dramatically improve our ability to characterize complex subsurface Earth structures. Observing such features will further require cross-correlating multiple station components to retrieve S-waves (e.g. RR, RT, ZR, etc), and using data from receivers with higher sampling rates (>20 Hz) to recover body wave reflections. Finally, an interesting avenue for this type of work is its applicability to time-lapse studies of the structural properties of seismic faults, in which detailed tomographic profiles are produced for any given fault of interest using discrete data sets from different time periods. In producing a variety of tomographic maps this way, one would be able to observe structural features of faults changing in time.

A

List of Figures

B

List of Tables

C

References

[1] Aki, K. (1980). Presidential address to the Seismological Society of America. *Bull. Seism. Soc. Am.*, 70:1969–1976.

[2] Aki, K. and Richards, P. (1980). *Quantitative Seismology: theory and methods.* W.H. Freeman, San Francisco.

[3] Audet, P. (2015). Layered crustal anisotropy around the San Andreas Fault near Parkfield, California. *Journal of Geophysical Research: Solid Earth*, 120:3527–3543.

[4] Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A., Lin, F., Moschetti, M., Shapiro, N., and Yang, Y. (2007). Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophys. J. Int.*, 169:1239–1260.

[5] Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R., and Larose, E. (2008). Postseismic Relaxation Along the San Andreas Fault at Parkfield from Continuous Seismological Observations. *Science*, 321:1478–1481.

[6] Brenguier, F., Shapiro, N. M., Campillo, M., and Nercessian, A. (2007). 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations. *Geophysical Research Letters*, 34.

[7] Campillo, M. and Paul, A. (2003). Long-Range Correlations in the Diffuse Seismic Coda. *Science*, 299:547–549.

[8] Eberhart-Phillips, D. and Michael, J. (1993). Three-Dimensional Velocity Structure, Seismicity, and Fault Structure in the Parkfield Region, Central California. *Journal of Geophysical Research*, 98:15,737–15,758.

[9] Lin, F.-C., Moschetti, M., and Ritzwoller, M. (2008). Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps. *Geophys. J. Int.*, 173:281–298.

[10] Lin, F.-C., Ritzwoller, M., and Snieder, R. (2009). Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophys. J. Int.*, 177:1091–1110.

[11] Lin, G., Thurber, C., Zhang, H., Hauksson, E., Shearer, P., Waldhauser, F., Brocher, T., and Hardebeck, J. (2010). A California statewide three-dimensional seismic velocity model from both absolute and differential times. *Bulletin of the Seismological Society of America*, 100:225–240.

[12] Miyazawa, M., Snieder, R., and Venkataraman, A. (2008). Application of seismic interferometry to extract P- and S-wave propagation and observation of shear-wave splitting from noise data at Cold Lake, Alberta, Canada. *Geophysics*, 73:D35–D40.

[13] Nadeau, R. M. and Dolenc, D. (2005). Nonvolcanic tremors deep beneath the San Andreas fault. *Science*, 307:389.

[14] Nadeau, R. M. and Guilhem, A. (2005). Nonvolcanic Tremor Evolution and the San Simeon and Parkfield, California, Earthquakes. *Science*, 325:191–193.

[15] Nadeau, R. M. and Guilhem, A. (2009). Nonvolcanic Tremor Evolution and the San Simeon and Parkfield, California, Earthquakes. *Science*, 325:191–193.

[16] Nataka, N., Chang, J., Lawrence, J., and Boué, P. (2015). Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry. *Journal of Geophysical Research: Solid Earth*, 120:11591173.

[17] P., R., Sabra, K., Gerstoft, P., and Kuperman, W. (2005). P-waves from cross-correlation of seismic noise. *Geophysical Research Letters*, 32.

[18] Peng, Z. and Gomberg, J. (2010). An integrated perspective of the continuum between earthquakes and slow-slip phenomena. *Nature Geosciences*, 3:599 – 607.

[19] Prieto, G., Lawrence, J., and Beroza, G. (2009). Anelastic Earth structure from the coherency of the ambient seismic field. *Journal of Geophysical Research: Solid Earth*, 114.

[20] Rost, S. and Thomas, C. (2002). Array seismology: Methods and applications. *Reviews of Geophysics*, 40:1008.

[21] Roux, P. (2009). Passive seismic imaging with directive ambient noise : application to surface waves and the San Andreas Fault in Parkfield , CA. *Geophys. J. Int.*, 179:367–373.

[22] Seats, K. J., Lawrence, J. F., and Prieto, G. A. (2012). Improved ambient noise correlation functions using Welch's method. *Geophysical Journal International*, 188:513–523.

[23] Shapiro, N., Campillo, M., Stehly, L., and Ritzwoller, M. H. (2005). High-Resolution Surface-Wave Tomography from Ambient Seismic Noise High-Resolution Surface-Wave Tomography from Ambient Seismic Noise. *Science*, 307:8–12.

[24] Stein, S. and Wysession, M. (2003). *An Introduction to Seismology, Earthquakes, and Earth Structure*. Blackwell Publishing.

[25] Thurber, C., Roecker, S., Roberts, K., Gold, M., Powell, L., and Rittger, K. (2003). Earthquake locations and three-dimensional fault zone structure along the creeping section of the San Andreas fault near Parkfield, CA: Preparing for SAFOD. *Geophysical Research Letters*, 30:1112.

[26] Thurber, C., Roecker, S., Zhang, H., Baher, S., and Ellsworth, W. (2004). Fine-scale structure of the San Andreas fault zone and location of the SAFOD target earthquakes. *Geophysical Research Letters*, 31:10–13.

[27] Thurber, C., Zhang, H., Waldhauser, F., Hardebeck, J., Michael, A., and Eberhart-phillips, D. (2006). Three-Dimensional Compressional Wavespeed Model , Earthquake

Relocations , and Focal Mechanisms for the Parkfield , California , Region. *Bulletin of the Seismological Society of America*, 96(4):38–49.

[28] Wapenaar, K. (2004). Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation. *Physical Review Letters*, 93(25).

[29] Wapenaar, K., Draganov, D., Snieder, R., Campman, X., and Verdel, A. (2010a). Tutorial on seismic interferometry: Part 1 Basic principles and applications. *Geophysics*, 75(5):75A195—-75A209.

[30] Wapenaar, K. and Fokkema, J. (2006). Green's function representations for seismic interferometry. *Geophysics*, 71(4):SI33.

[31] Wapenaar, K., Slob, E., Snieder, R., and Curtis, A. (2010b). Tutorial on seismic interferometry: Part 2 Underlying theory and new advances. *Geophysics*, 75(5):75A211.

[32] Weaver, R. and Lobkis, O. (2001). Ultrasonics without a Source: Thermal Fluctuation Correlations at MHz Frequencies. *PHYSICAL REVIEW LETTERS*, 87(13):134301.

[33] Weaver, R. and Lobkis, O. (2002). On the emergence of the Greens function in the correlations of a diffuse field: pulse-echo using thermal phonons. *Ultrasonics*, 40:435439.

[34] Zeng, X., Thurber, C., Shelly, D., Harrington, R., Cochran, E., Bennington, N., Peterson, D., Guo, B., and McClement, K. (2016). 3-D P- and S-wave velocity structure and low-frequency earthquake locations in the Parkfield, California region. *Geophys. J. Int.*, 206:1574–1585.

D

Tablulation of Time Series Containing Manually Detected Events

Julian Day (2003)	Hour(s) Containing Event(s) (UTC)
1	0,1,3,5,7,9,11,13,17,19,20,21,22
2	1,3,7,8,11,14,15,16,18,20,22
3	1,2,4,5,6,7,8,10,11,12,13,14,16,17,19,20,21,22
4	0,2,5,6,7,8,9,10,11,12,13,14,15,20,21,22
5	0,5,6,7,12,15,16,17,18
6	0,1,2,5,9,10,11,12,13,14,15,21
7	0,1,4,6,7,8,9,10,14,16,18,22
8	1,2,5,8,9,10,12,14,15,17,18,22
9	0,2,3,4,5,6,7,8,13,14,15,16
10	0,1,4,5,6,7,8,10,12,13,15,18,19,21
11	0,1,2,7,8,9,10,12,14,15,16,17,20,21
12	6,8,10,11,15,22
13	3,7,8,10,11,12,13,14,15,19,20
14	3,6,8,12,13,17,22
15	4,5,6,7,9,10,13,16,22
16	1,3,7,10,12,18,20
17	6,8,10,11,12,22
18	1,5,10,15,17,18
19	0,3,4,7,9,11,13
20	0,2,3,4,5,6,8,9,10,12,14,15,19,21
21	1,2,4,7,10,11,13,14,17
22	1,3,4,6,7,8,17,19,20
23	0,1,6,7,8,10,13,15,18,19,20,21,22
24	2,3,4,5,7,8,9,11,12,13,14,15,16,18,22
25	1,2,3,5,6,7,8,9,10,12,13,14,18,19,20
26	3,5,6,7,9,12,13,16
27	0,9,10,11,12,13,14,18,20,21
28	0,1,2,3,10,11,12,20,22
29	0,1,2,4,5,6,7,8,10,11,13,19,20,21
30	3,4,8,10,11,12,13,15,17,19,20,21
31	0,3,5,7,8,9,12,16,17,18,22
32	1,7,10,14,16,18,20,22
33	0,4,11,16,17,18,19
34	2,8,11,12,16,18
35	1,3,10,11,12,14,19,22
36	0,2,3,6,7,8,10,11,12,13,14,15,19
37	1,2,11,13,14,15,16
38	0,1,2,3,7,9,10,20
39	3,9,10,12,14,19,20,22
40	0,1,2,3,5,6,7,9,14,15,16,21,22
41	0,3,8,9,10,11,12,13,14,15,16,17,18,20,22
42	2,3,4,6,8,9,10,11,13,14,18
43	0,2,3,8,9,10,13,16,20,21,22
44	0,1,4,5,6,9,10,12,15,16,17,20,21,22
45	1,5,6,14,15,16,22
46	0,7,8,9,11,15,16,21
47	0,1,2,3,5,20,21,22
48	4,5,8,10,16
49	11,14,15
50	3,5,12,15,16,17,20,21,22
51	0,1,2,5,6,9,10,14,22
52	0,1,2,4,6,7,9,12,16,18,19,20,22
53	1,3,5,9,11,12,13,14,15,16,17,19
54	0,7,9,13,15,18,19,20,21,22
55	0,2,3,4,7,8,9,14,18,22
56	0,1,2,3,4,5,9,10,11,12,15,16,21
57	1,8,9,10,14,16
58	0,1,3,4,5,6,8,11,12,13,14,15,18,19,21
59	0,1,2,3,4,6,8,10,11,12,13,15,16,22
60	0,5,6,8,9,11,14,18,20
61	0,3,7,8,9,11,13,14,17,18
62	0,6,10,11,13,17,18,20
63	0,1,7,9,10,13,14,15,18,19,21
64	0,5,6,16,18
65	1,2,3,4,5,7,8,10,11,12,13,15,19
66	1,9,13
67	1,4,6,8,11,15,16,18
68	1,2,5,8,14,21
69	1,2,9,19,20,22
70	1,2,3,7,8,9,10,11,17,19
71	2,4,6,8,10,11,13,16
72	6,7,8,9,11,14
73	0,2,3,9,11,12,13,16,18,21
74	0,2,5,7,10,12,14,15,19
75	3,6,10,14,20,21
76	6,13,16,19,23
77	2,4,7,8,11,19
78	4,5,10,11,12,15,20
79	0,3,4,6,11,12,15,17
80	1,7,8,11,13,14,15,16,17,18,19
81	0,1,5,6,7,8,9,11,15,18,22
82	5,8,10,12,16,20,21,22
83	2,3,4,6,7,8,9,11,13,14,15,16,17,18,19,20,22
84	1,3,8,9,10,15,20
85	1,5,7,10,11,15,20
86	2,6,11,14,15,16,19
87	3,4,17
88	0,9,10,14,15
89	0,4,11,13,14,15,17,18,20
90	1,2,3,5,8,11,13,15,16,17
91	1,3,4,6,9,10,11,13,14,18,19,20,21,22
92	0,1,2,3,4,5,6,7,8,9,18,19,21,23
93	2,3,4,6,10,13,15,20
94	2,4,5,6,7,10,12,14,15,16,18
95	1,2,3,7,8,12,13,14,20
96	2,4,5,7,10
97	0,6,9,10,12,14,22
98	2,3,5,6,20,22

Julian Day (2003)	Hour(s) Containing Event(s) (UTC)
99	1,11,12,14,18
100	2,5,7,13
101	3,4,11,16,19,22
102	1,8,9,10,14,15,21
103	5,11
104	8,10,11,18,22
105	4,5,7,8,9,16,18
106	10,11,12,19
107	0,1,3,5,7,14,15
108	0,9,12,13,16
109	-
110	1,3,4,6,7,10,11,12,13,15,21
111	1,3,4,5,6,7,8,9,10,11,12,14,15,16,17,22
112	1,10,11,12,14,15,16,18,19,21
113	1,4,5,6,7,13,21
114	1,2,3,4,5,6,9,10,11,12,13,14,19,22
115	0,6,7,8,10,11,12,15,16,17,22
116	1,6,9,10,13,15,16
117	2,5,6,7,8,10,12,13,16
118	5,6,7,8,9,10
119	0,1,5,6,7,14,15,19
120	0,8,10,15,17
121	0,5,12,14
122	3,4,6,10,14,16,17,21
123	5,6,9,10,12,13,16,21
124	2,4,6,8,12,13,20,21
125	4,5,9,11,12,15,18,20
126	4,5,6,9,11,13,15,19,20
127	7,12,13,17
128	2,4,5,7,12,16,17
129	2,4,17
130	1,5,6,9,11,17,18
131	11
132	10,12,17,18,20,21
133	2,3,8,10,16,17,21
134	1,2,6,11,19,20,22
135	2,15,17,20,22
136	2,5,8,16
137	0,1,2,4,5,6,8,14,18,22
138	0,2,4,6,10,14,17,18,20
139	0,1,2,3,7,8,10,13,14,15,16
140	0,1,2,10,14,16
141	4,7,9,11,16,18,19,20,21
142	16,19,20
143	2,10,11,16,18
144	2,6,8,10,11,13,16
145	1,2,7,8,19,21
146	2,3,7,9,10,12,21
147	-

Julian Day (2003)	Hour(s) Containing Event(s) (UTC)
148	-
149	8,12,18,22
150	0,1,4,6,9,18,19,22
151	3,6,8,13,14
152	9,21
153	8
154	-
155	0,5,11,14
156	0,5,19
157	2,9,10,11,17,20
158	0,1,5,11,13,14,19,
159	22
160	0,22
161	0,2,6,10,11,14,17,19
162	-
163	1,9,11,12,14,16,18,21
164	0,1,3,6,7,13,14,21
165	4,7,13,16,17,18,22
166	4,7,9,11,12,13,17,18,19,20,22
167	2,8,12,21
168	0,7,9,10,12,13,17,20,21
169	3,5,7,9,14,15,17,19,22
170	1,5,6,8,15,22
171	1,6,7,13,20
172	3,6,8,10,12,16
173	1,4,6,7,10,12,17,18
174	4,6,7,8,9,11,12,16,19,22
175	13,16
176	0,1,4,5,7,9,12,14,21
177	1,3,6,9,13,17,18,21
178	2,4,5,6,9,10,11,15,17,18,19
179	1,13,14,16,22
180	2,6,7,8,9,13,14,17,18,19
181	1,2,3,6,10,14,19,22
182	2,5,9,11
183	5,18,19
184	1,3,6,8,10,17
185	4,7,11,17,18,20,21
186	0,1,2,3,4,9,10,11,12,13,16,21,22
187	0,3,4,7,8,9,10,11,16
188	3,9,11,18,19,22
189	0,3,6,7,10,13,18,22
190	1,7,17,20,21
191	0,4,6,7,11,12,17,19
192	0,10,14
193	1,2,5,6,8,9,14,15,19,21
194	1,5,6,8,10,11,12,13,14,16,18,19,20,21
195	0,1,6,8,20,21
196	1,5,6,10,12,13,15,16,17,20,22

Julian Day (2003)	Hour(s) Containing Event(s) (UTC)
197	0,1,2,3,5,8,9,12,16,22
198	4,8,9,10,13,14,20,22
199	0,1,8,9,11,13,14,16,17
200	1,8,11,16,17,18,19,21
201	1,4,5,6,7,8,11,13,19
202	1,3,4,6,7,14,18,19,20,22
203	0,3,4,5,6,7,8,17,18,19
204	4,7,8,9,10,11,12,13
205	0,7,8,12,13,14,19
206	0,4,5,9,12,13,15,16,17,18,19,22
207	0,1,2,4,5,7,10,13,15,20,21
208	1,2,3,4,5,6,7,8,11,12,17,19
209	0,2,4,6,8,12,13,17,19,20,21
210	0,2,9,11,13,16,17,18,20,21,22
211	1,2,5,6,7,8,9,10,11,12,13,19,20
212	1,2,3,5,7,8,9,10,11,12,13,15,18,19,20
213	2,3,4,6,7,12,13,15,18,19,22
214	0,1,6,11,14,16,20
215	2,5,6,12,14,16,21
216	1,3,4,5,6,11,17,18,19,21
217	1,2,3,4,5,7,8,10,11,12,13,15,16,18,19,20
218	0,1,3,4,6,7,9,10,11,12,14,19,21
219	3,5,7,9,11,12,13,16,17,19,20
220	0,4,11,17,18,20,21
221	0,1,6,9,10,12,13,18,21
222	1,3,7,8,9,10,13,14
223	0,1,7,13,14,15,16,18
224	2,4,7,8,10,11,13,14,15,16,20,22
225	1,3,4,7,9,12,13,14,15,19
226	0,1,2,5,14,19,20,21,22
227	0,1,2,6,9,10,12,13,14,15,16,18,20,21,22
228	0,1,2,5,7,8,10,13,20,21
229	1,3,4,6,10,18
230	1,4,5,7,8,9,11,12,13,17,18,19,20,21,22
231	4,7,12,13,14,17,18,20,21,22
232	0,6,7,12,13,17,21,22
233	2,5,9,10,11,12,14,16,20
234	2,5,7,11,12,15,17,20,22
235	0,3,4,6,10,11,12,15,17
236	0,3,4,6,7,8,9,10,11,12,14,16,17,18,21,22
237	1,3,5,6,7,8,9,10,11,12,13,14,15,19,20,22
238	0,2,6,10,11,14,16,17,18,19,20,21,22
239	3,4,5,6,8,12,15,17,22
240	0,1,2,3,6,8,9,10,12,15,16,17,18,19,20,21,23
241	0,1,3,4,5,6,7,8,9,11,12,14,16,18,19,20,21,22
242	0,4,5,7,8,10,13,17,18,20,21
243	4,5,6,7,8,9,11,14,16,18,21,22
244	0,2,3,5,6,7,9,10,14,16,17,18
245	1,5,7,8,9,10,11,12,13,16,18,19,21

Julian Day (2003)	Hour(s) Containing Event(s) (UTC)
246	4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22
247	0,1,3,5,6,7,9,10,11,13,14,15,17,18,19,21
248	0,2,3,4,6,7,8,9,10,11,12,13,22,23
249	0,1,2,3,7,9,10,11,12,13,14,15,16,17,18,19,21
250	0,1,5,8,9,10,12,13,19
251	3,4,5,6,7,8,9,11,12,13,16,17,19,20
252	0,1,2,3,4,5,6,7,9,13,17,19
253	3,4,5,7,8,11,12,13,15,17,21,22
254	0,3,5,6,7,9,10,11,12,14,17,19,20,21,22
255	0,3,4,5,6,7,9,10,12,13,161,17,18,20,21,22
256	1,5,8,9,12,14,16,19,22,23
257	0,1,2,3,4,6,8,9,10,11,12,15,16,17
258	0,1,2,4,5,11,12,19
259	6,7,11,12,16,17,21
260	2,3,5,7,10,11,19,20,21,22
261	1,2,5,6,8,9,10,11,13,15,16,18,20,22
262	2,14,18,20,21
263	0,6,8,9,10,11,12,16,17,18,19,21,22
264	0,1,2,4,5,6,7,8,9,10,11,12,17,19,20,21
265	0,2,4,5,7,12,13,18,20,22
266	1,3,5,6,7,9,11,12,15,18,20,21
267	0,1,4,8,10,11,18,19,21,22
268	0,1,2,3,7,8,9,10,11,13,14,19,22
269	1,3,6,9,11,13
270	2,3,5,6,7,8,10,11,12,14,16,18,19,20
271	0,1,2,4,5,6,8,10,16,19,21
272	1,2,3,5,11,12,13,16,20
273	3,4,6,8,10,11,12,14,18,19,20
274	1,3,4,5,6,7,8,9,10,13,14,16
275	1,2,3,7,9,10,12,15,16,17,18
276	4,5,7,9,10,11,12,13,16,19,21,22
277	0,1,5,6,9,10,11,12,22
278	0,4,6,8,11,19,22
279	2,3,4,6,8,11,12,21,22
280	1,2,3,4,5,6,7,8,9,10,11,12,17,18,20,22
281	2,3,6,9,11,15,16
282	2,5,6,7,10,13,14,22
283	0,1,2,4,5,6,7,8,11,12,21
284	1,3,4,5,9,10,11,12,14,19
285	1,2,3,4,5,6,7,8,10,11,12,13,14,15,16,17,19,22
286	2,8,11,13
287	2,3,4,8,9,10,12,13,18,19,20,22
288	0,2,3,4,5,6,7,8,9,10,11,12,13,15,16,17
289	0,1,3,4,5,6,7,8,9,10,11,12,13,14,16,17,19,20,22
290	0,2,3,4,5,7,8,9,10,12,13,14,15,17,18,19,22
291	2,3,6,7,9,10,11,13,14,16,21
292	4,5,6,7,8,9,12,13,14,15
293	2,3,4,5,7,8,9,10,11,12,13,17,18,19,20,21
294	0,3,4,5,8,10,11,13,15,16,18,19,22

Julian Day (2003)	Hour(s) Containing Event(s) (UTC)
295	1,2,3,4,5,6,8,9,11,12,13
296	3,7,8,10,11,12,13,14,16,20,21
297	2,3,4,5,8,10,12,13,15,16,17,21,22
298	0,2,4,5,6,11,12,13,14,15,17,18,19
299	0,3,4,5,6,9,10,18
300	3,6,7,8,12,13,15,16,17,19,21
301	0,1,2,4,5,6,8,9,10,11,12,13,14,15,18,20,21,22
302	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,18,20,22
303	3,5,7,10,11,12,13,16,18,19,20,22
304	0,2,7,8,11,12,13,15,18,20
305	1,3,4,7,8,10,11,13,14,20,22
306	2,3,4,10,11,12,19
307	1,2,4,9,10,11,12,13,14,15,17,18
308	1,4,8,10,11,12,13,16,17
309	2,3,4,5,6,7,8,9,10,11,12,17,20,22
310	2,4,5,6,9,10,12,13,17,19,21,22
311	0,1,2,3,4,6,7,8,11,13,14,20,21,22
312	0,1,6,10,11,12,13,14,16,20,22
313	2,3,5,9,11,12,13,16
314	1,3,4,9,10,12,13,14,15,17,18,19,20
315	2,3,4,6,9,10,11,14,15,17,19,21
316	0,2,3,4,5,6,8,10,12,13,14,15,16,17,20,22
317	0,1,3,4,5,6,7,8,9,11,14,17,18,19,22
318	1,3,4,6,8,9,10,11,12,13,14,16,17,18
319	3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,22
320	0,1,3,4,5,6,7,8,9,10,11,12,13,15,17,21,22
321	2,4,6,7,8,9,10,13
322	0,1,2,3,4,5,6,7,8,9,11,12,13,14,16,17,19,21,22
323	1,2,3,4,5,6,7,8,9,10,11,12,13,15,19,21,22
324	0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,18,20
325	1,2,4,6,7,8,9,10,11,12,13,14,15,17,19,21,22
326	0,1,2,3,4,5,6,7,8,9,10,11,12,16,20,21,22
327	0,1,4,5,6,7,8,10,11,12,13,17,18,19,20,21,22
328	0,1,4,5,6,7,9,11,12,14,17,21
329	0,1,3,5,7,8,11,12,13,14,17,20
330	0,1,4,5,7,8,9,10,11,12,13,14,17,18
331	0,2,4,10,11,12,13,14,15,18
332	0,1,2,4,5,7,8,9,10,11,12,13,14,16,18,20,22
333	3,4,5,6,7,8,11,13,15,19,22
334	0,1,2,3,4,5,6,7,9,10,11,12,14,15,17
335	2,3,4,5,7,8,10,11,12,13,14,16,17,22
336	0,1,3,4,7,8,9,10,14
337	0,1,2,3,4,5,7,10,11,13,14,21
338	1,3,5,6,7,8,9,10,13,14,20,21
339	0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,21,22
340	1,2,3,4,5,6,7,8,9,10,11,12,13,14,16,18,21
341	0,1,2,3,4,5,6,7,8,9,10,11,13,14,15,16,17,18,19,20,21
342	5,6,7,9,10,11,15,17,18,19,20,23
343	0,1,2,3,5,6,7,8,9,10,11,12,13,14,18,21
344	0,5,8,10,11,13,14,17,20,21,22,23
345	0,1,2,3,4,6,8,9,10,11,14,18,19
346	1,2,6,7,8,9,10,11,13,14,17,19,20
347	1,3,4,5,6,10,11,12,13,14,15,18,19
348	0,5,8,9,10,11,12,13,14,18,19,20,21,22
349	0,3,5,6,9,11,13,18,21,22
350	1,7,8,9,11,14,15,16,17,18,19,20,21,22
351	3,4,6,7,9,10,11,12,13,14,21,22
352	0,3,5,6,7,9,10,11,13,14,18,19,20,22
353	0,1,4,5,6,7,8,9,10,11,12,13,14,15,16,18,20,21
354	0,4,5,7,8,12,14,15,16,17,19,20,21,22
355	0,2,4,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22
356	1,3,4,8,9,11,19,20,21,22
357	full day
358	full day
359	full day
360	full day
361	0,1,2,4,5,6,8,9,11,12,13,14,15,16,17,18,19,20,21,22
362	full day
363	full day
364	0,1,2,4,5,6,7,9,10,11,12,13,14,15,16,17,18,19,20,21,22

Table D.0.1: Hour-long time series in which events were manually detected. Events were detected in 3706 hour-long seismograms.

E

Spectral Whitening: An Intuition Building Excercise

That a seismic time series may be spectrally whitened by reconstructing it exclusively from its phase spectrum may not seem intuitive. However, an intuition for how this operation affects the data can be gained by literally hearing it's effects. We may gain this intuition by audibly comparing a spectrally whitened acoustic signal to its original waveform, since, with respect to digital signal processing, the content of time series data is irrelevant.

In this exercise we use a short recording of a sequence of notes played on a piano as our input signal (waveform obtained with permission from Xavier Serra of Universitat Pompeu Fabra, 2016). The waveform is shown in Figure E.0.1 and features five prominent attacks. Using this signal as our input, we pre-emptively taper it in the time domain and then obtain both its amplitude and phase spectra via a Fourier transform. We then whiten the signal by rebuilding it from the inverse Fourier transform of its phase spectrum. For completeness, we also rebuild a copy of the signal from its amplitude spectrum, as comparing both reconstructions allows for greater insight into the nature of information contained by

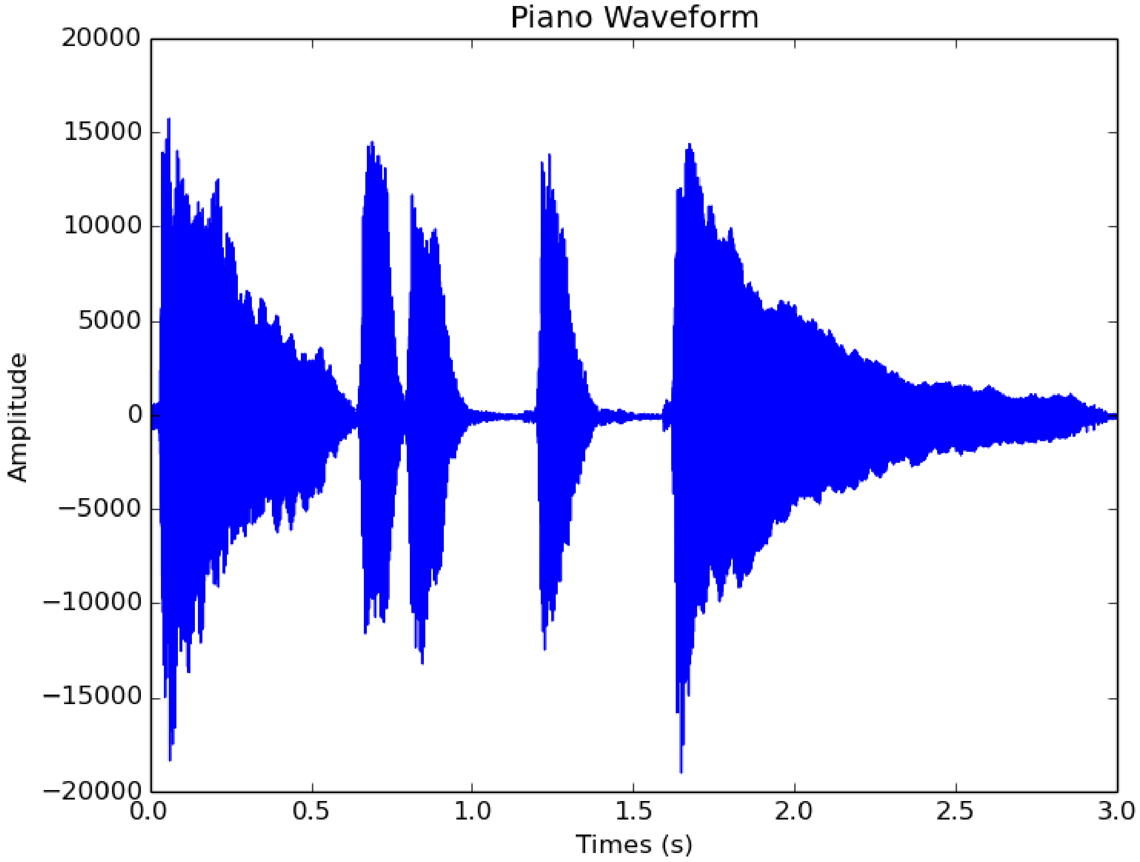

Figure E.0.1: Acoustic waveform produced from a sequence of five notes played on a piano.

the phase and amplitude aspects of a signal. Results are shown in Figure E.0.2, from which we first make several visual observations.

Regarding the phase-reconstructed signal, we note that this signal preserves the precise timing of the attacks in the original waveform. However, the portions of the signal between attacks are characteristically louder, such that the attacks don't stand out as much against the background noise. Additionally, the whitened waveform is significantly diminished in magnitude compared to the original. Regarding the amplitude-reconstructed signal, its most striking feature is its time-symmetric nature, followed by the concentration of signal energy around its endpoints. Part of the reason this occurs is because information concerning attack onset is contained in the phase portion of the original signal, which was discarded in the reconstruction. Thus, the timing of the attacks in the original waveform have been lost and signal energy pools around the endpoints because each frequency component in the signal attains a phase of zero. Finally, the magnitude, or volume, of this signal is much greater

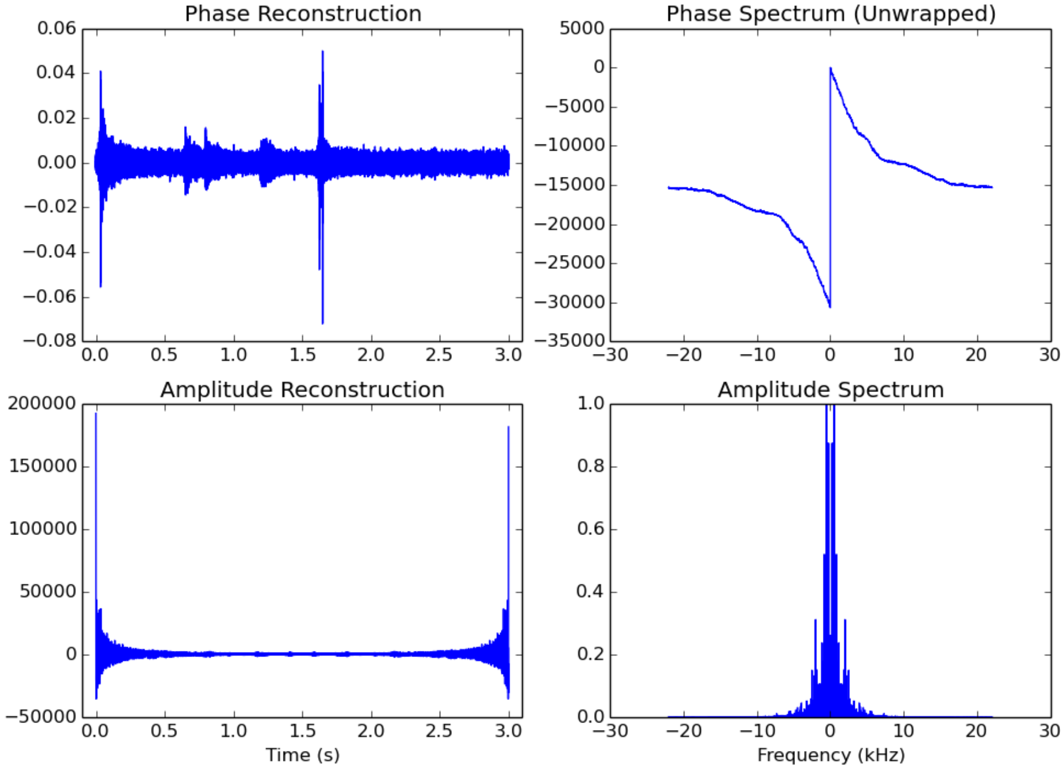

Figure E.0.2: Phase-based and amplitude-based signal reconstructions of the waveform in Figure E.0.1.

than in the original.

If we listen to the phase-reconstructed waveform three distinct features stand out. First, compared to the original, this signal is much quieter. Naturally, this is because amplitude information has been discarded. Second, the phase-reconstructed waveform is suffused with a static hissing due to its uniform amplitude spectrum, and hence sounds much more like white noise. Third, while the onset of the attacks can be heard above the noise floor, their decay is almost immediately lost in the noise. Listening to the amplitude-reconstructed waveform, beyond its time-symmetric nature and it's heightened volume, the signal sounds extremely distorted and disjointed compared to the original.

With regard to seismic data, this exercise demonstrates why whitening via phase-reconstruction is a natural choice for ambient noise interferometry. Qualitatively, we have observed that this operation diminishes the parts of the waveform following an attack (seismic events) and that it boosts the presence of noise in the signal. Another useful feature of this operation is that is preserves attack onset, or timing information, which is the crucial aspect needed to accurately estimate wave travel-times from Green's functions.