

## ABSTRACT

### Seismic Site Characterization through Joint Modeling of Complementary Data Functionals, with Applications to Santo Domingo, Dominican Republic

Martin Schwed, M.S.

Mentor: Jay Pulliam, Ph.D.

We present a strategy for determining seismic “site characterization” through joint modeling of horizontal to vertical spectral ratios (HVSr) and surface wave dispersion, performed via refraction microtremor (ReMi). Fitting of data functionals by synthetics is driven by global optimization. The products of this approach are shear wave velocity profiles for the subsurface, accompanied by marginal posterior probability distributions (PPD) and parameter correlation matrices. Jointly fitting dispersion curves and HVSr functionals via global optimization allows us to assess model reliability, identify parts of the “best-fit” model that are poorly constrained, and guide us toward new data that might improve model constraints. Tools such as the PPD and the parameter correlation matrix assess the relative contribution of each type of data to model constraints. The joint modeling technique is applied to data acquired in an NSF-funded PASI in Santo Domingo, Dominican Republic, as a proof-of-concept survey in a highly built-up urban environment.

Seismic Site Characterization through Joint Modeling of Complementary Data  
Functionals, with Applications to Santo Domingo, Dominican Republic

by

Martin Schwed, B.S.

A Thesis

Approved by the Department of Geology

---

Stacy C. Atchley, Ph.D.

Submitted to the Graduate Faculty of  
Baylor University in Partial Fulfillment of the  
Requirements for the Degree  
of  
Master of Science

Approved by the Thesis Committee

---

Jay Pulliam, Ph.D., Chairperson

---

John A. Dunbar, Ph.D.

---

David Kahle, Ph.D.

Accepted by the Graduate School

August 2014

---

J. Larry Lyon, Ph.D., Dean

Copyright © 2014 by Martin Schwed

All rights reserved

## TABLE OF CONTENTS

Introduction.....	1
Field Application: Santo Domingo, Dominican Republic .....	2
Joint Modeling of Multiple Datasets .....	3
Data Functionals .....	4
Surface Wave Dispersion .....	4
Horizontal to Vertical Spectral Ratios .....	4
Global Optimization Method: Very Fast Simulated Annealing (VFSA).....	5
Uncertainty Estimation .....	6
Marginal Posterior Probability Density (PPD) Functions .....	6
Parameter Correlation Matrices .....	8
Data and Methodology .....	9
Data Collection .....	9
Data Processing .....	9
Refraction Microtremor (ReMi) .....	9
Horizontal-to-Vertical Spectral Ratios (HVSR) .....	10
Optimization through Very Fast Simulated Annealing .....	10
Results .....	11
Este 3 .....	11
Airport 1 .....	13
UASD .....	14
Sur 1 .....	15
Discussion .....	16
Conclusions .....	18
Acknowledgments.....	19
References.....	20
Tables .....	22
Figures .....	23

## LIST OF TABLES

Table 1 .....	22
---------------	----

## LIST OF FIGURES

Figure 1 .....	23
Figure 2 .....	24
Figure 3 .....	25
Figure 4 .....	26
Figure 5 .....	27
Figure 6 .....	28
Figure 7 .....	29
Figure 8 .....	30
Figure 9 .....	31
Figure 10 .....	32
Figure 11 .....	33
Figure 12 .....	34
Figure 13 .....	35
Figure 14 .....	36
Figure 15 .....	37

## Introduction

Seismic “site characterization” describes the amplification and frequencies of ground shaking at particular locations in response to an earthquake (Borcherdt 1970). Investigations to determine site characteristics are therefore critical components, along with source characterization and wave propagation, of earthquake hazard assessment strategies. Seismic site characterization is very local in nature and is typically performed on a dense grid of observation locations in, for example, a metropolitan area (Louie et al. 2011).

Site characterization has been performed for decades using “active” sources, such as explosions, sledgehammer blows to the ground, or specially built shakers (Boore and Asten 2008). These strategies are effective but are often not feasible for urban areas. Recently, a new class of techniques has emerged that use “ambient noise”, low levels of seismic shaking that occur constantly due to natural and human-caused sources. When recorded simultaneously by an optimally configured array of seismic sensors, these signals can be identified and retrieved via computer-intensive cross-correlation of waveforms that were recorded for as little time as 20 minutes. Ultimately, those signals, can be modeled to produce a sufficiently accurate physical characterization of the subsurface

The most widely used site characterization techniques include methods that aim to characterize surface wave dispersion as an indicator of subsurface structure, such as SPatial AutoCorrelation (SPAC) (e.g., Aki 1957; Asten 2001), Refraction Microtremor (ReMi) (Louie 2001), and Multi-channel Analysis of Surface Waves (MASW) (e.g.,

Park, Miller, and Xia 1999). Aside from these, the Horizontal to Vertical Spectral Ratio method (Nakamura 1989) determines the fundamental period of subsurface layers, and has been used widely to estimate the depth to bedrock (Arai and Tokimatsu 2005).

#### Field Application: Santo Domingo, Dominican Republic

This joint modeling technique was applied to a dataset collected during the Pan-American Advanced Studies Institute (PASI), entitled “New frontiers in geophysical research: Bringing new tools and techniques to bear on earthquake hazard analysis and mitigation,” an NSF-funded workshop in Santo Domingo, DR, so the results are intended to serve as a proof-of-concept survey of a highly built-up urban environment (Willemann et al. 2013).

The application to Santo Domingo is challenging, however, for two reasons. First, although the city was founded in the late fifteenth century and has been continuously inhabited since that time, its shallow subsurface structure has not been mapped in detail. Few boreholes have been logged, so “ground truth” is elusive. Second, Santo Domingo was built on a series of limestone terraces, so the soil layer is, in many cases, quite thin (1-3 m) and is underlain widely by limestone in various stages of decomposition and fracture. Studies of amplification have generally focused on basins and relatively thick piles of soil and sediment, very different structures from those found in Santo Domingo. Given the unusual nature, for seismic site characterization, of the city’s subsurface, and the unusual appearance of HVSRs, particularly, we should be open-minded about differences between modeling results in Santo Domingo and results of other studies elsewhere.

Further, for circumstances such as those found in Santo Domingo, it makes sense to develop and apply a modeling algorithm that seeks to match multiple data types via global optimization. Bringing additional types of data to bear has the potential to limit the range of models that fit the observations and global optimization searches the space of possible models broadly to find all models that fit the data acceptably, even if those models are quite different from each other.

### Joint Modeling of Multiple Datasets

Despite having been used widely, both kinds of ambient noise site characterization methods have only recently been used in conjunction with each other (Arai and Tokimatsu 2005). Fitting these distinct functionals (surface wave dispersion curves and horizontal to vertical spectral ratios) jointly should, in principle, provide stronger constraints on model parameters than does either functional alone, and should produce more reliable models of the shallow subsurface.

Optimization strategies for solving nonlinear problems in geophysics have several advantages over linearized inversions. Global optimization allows us to characterize the space of possible models, assess model reliability, identify parts of the “best-fit” model that are poorly constrained, and guide us toward new data that might improve constraints on the model. We employ tools such as the posterior probability distribution and the parameter correlation matrix, which assess the relative contribution of both types of data to model constraints and how to choose optimal weights for each data type.

## *Data Functionals*

### *Surface Wave Dispersion*

Seismic surface waves are inherently dispersive; they propagate at different velocities depending on their wavelength. Depth sensitivity kernels vary by wavelength (and therefore frequency), so after isolating harmonic components and measuring their velocities it is possible to extract relationships between shear wave velocity and depth for the subsurface. Rayleigh waves are particularly sensitive to the shear wave velocity of a medium, and therefore serve as an excellent means for determining shear wave velocity profiles.

A dispersion curve illustrates the relationship between wave velocity and frequency. For all sites in this study, surface wave dispersion curves were generated by applying the ReMi methodology (Louie 2001), employing solely energy from background microtremors.

### *Horizontal to Vertical Spectral Ratios*

The Horizontal to Vertical Spectral Ratio (HVSr or H/V) method (Nakamura 1989) is employed to determine the site “response” to body waves, i.e. the frequencies at which ground motion is amplified and dampened. HVSr are computed by dividing the spectral amplitudes of the horizontal components of ground motion by the spectral amplitudes of the vertical component of ground motion, in order to remove source and Rayleigh wave effects (Huerta-Lopez, Pulliam, and Nakamura 2003). HVSr amplitudes are controlled largely by the shear wave structure of the subsurface and can be modeled using, for example, the “stiffness matrix” approach of Kausel and Roesset (1981).

### *Global Optimization Method: Very Fast Simulated Annealing (VFSA)*

Simulated Annealing (SA) is a numerical algorithm that parallels problems of statistical mechanics involving a large number of atoms in physical systems (Sen and Stoffa 2013). SA models a solid in a heat bath that is brought to a certain initial temperature and allowed to cool gradually. Slow cooling of the physical system will allow for the formation of crystals, i.e., the high mobility of atoms at high temperatures allow them to arrange themselves in the lowest energy state of the collection of atoms, as the material solidifies. In the numerical analogy, a slow cooling schedule allows a large portion of the model space to be examined in search of a global minimum of the fitness function.

In the basic process of Simulated Annealing, a starting model is selected randomly and its fitness is evaluated, a loop then begins in which models are selected or rejected as a function of a probability function (Ingber 1993). As the temperature is lowered, the probability of acceptance of values with greater error is reduced, leading the algorithm to search for solutions more locally and thereby refine solutions found during the broader search at higher temperatures. For proper choices of fitness function and cooling schedule, this process has been shown to identify effectively the global minimum of the fitness function for a variety of geophysical problems (e.g., Sen and Stoffa 1994, Stoffa et al. 1994, Zhao et al. 1996).

Very Fast Simulated Annealing (VFSA) implements a slight change to the probability of model acceptance, leading to an algorithm that finds global minima more efficiently. The change involves the usage of a Cauchy distribution when assigning probabilities to moves within the model space, as opposed to the Boltzmann distribution,

or a flat distribution, as employed by the Metropolis algorithm (Sen and Stoffa 1995). The advantage of the Cauchy distribution is that it has a fatter tail than other statistical distributions used in optimization and therefore a better chance of exiting local minima (Ingber 1989). This means that the cooling schedule, and thus the algorithm's convergence, can be accelerated with respect to the schedule employed in the Metropolis algorithm:

$$T(k) = \frac{T_0}{\ln(k)}$$

$T(k)$  is the temperature at iteration  $k$  and  $T_0$  is the starting temperature. The new, faster cooling schedule in VFSA becomes:

$$T(k) = \frac{T_0}{(k)}$$

Despite the faster convergence of the latter cooling schedule, the properties of the Cauchy distribution allow the VFSA algorithm to escape local minima (Ingber 1989).

### *Uncertainty Estimation*

#### *Marginal Posterior Probability Density (PPD) Functions*

Error functions associated with models often either have broad minima or are multi-valued, indicating that other acceptable solutions are either slightly different or very different from the single, best-fit model, respectively. One can use all the models that were selected and evaluated during multiple searches to compute a marginal posterior probability density (PPD) function and parameter correlation matrix (Sen & Stoffa 1996). The PPD shows the normalized likelihood that a given parameter takes a certain value given the observed data, and is thus a measure of the constraints imposed by

the data on the model parameters and, ultimately, the confidence one should have in the single, best-fit model. The parameter correlation matrix shows the tradeoffs between parameters, and is thus a second measure of confidence in the model. High parameter correlations indicate that model parameters are not constrained independently, so one could offset a change in one parameter with, for example, a change in another parameter and still fit the data equally well. Both the PPD and parameter correlation matrix reflect ambiguities, or uncertainties, of interpretation. The estimation of such uncertainties is particularly necessary in joint modeling problems because they allow us to determine whether, and to what extent, the number of possible models is reduced with the inclusion of each data functional and to evaluate tradeoffs between model parameters (Pulliam & Sen 2005). Perhaps most importantly, these statistical tools can be displayed graphically, so they can be quite useful for evaluating the reliability of results.

PPDs in this work (Figure 1) are represented as functions of shear wave velocity versus depth, in which each layer in the model is represented by a horizontal colored bar. The color within each of those bands represents the probability that the best-fit model will be within a certain range. Blue colors indicate low probabilities, while reds indicate high probabilities. If the model is well constrained for a particular layer, one would see a red region extending over a small range of possible velocities surrounded by blue, indicating that the probability of acceptable models lying outside the red region is small. However, there are several cases exhibited in this paper in which the red regions are broad, and sometimes more than one red region exists per layer, indicating poor constraints or more than one possible solution, respectively.

### *Parameter Correlation Matrices*

Once the PPD is known, it is possible to obtain the covariance through a multidimensional integral, which is then normalized and used to represent tradeoffs between model parameters in a parameter correlation matrix.

Parameter correlation matrices are square and symmetric about the diagonal. For this study, parameter correlation matrices are represented as colored triangle-shaped grids, with the redundant portion omitted. Both the x and y axes of these grids are labeled with a particular parameter for each given layer. Each cell in the grid gives a color representation of the cross-correlation between a pair of parameters, which can range from +1 to -1. Blue colors represent a negative cross-correlation, greens show a relative lack of cross-correlation, and reds show a positive cross-correlation. As models become better constrained, these parameter correlation matrices should become greener. The diagonal represents autocorrelations, which are uniformly equal to one. In every case presented here, the average cross-correlation is displayed on the diagonal of the cross-correlation matrix as a relative measure of parameter correlation, as in, how far from green a particular matrix is, with an average of 0 demonstrating a complete lack of parameter correlation and therefore a completely green matrix. An average of 1 would correspond to an all red/blue matrix. In this study, three of the parameters displayed in the parameter correlation matrices - Poisson's ratio ( $\nu$ ), shear wave velocity ( $\beta$ ), and layer thickness ( $h$ ) - are free parameters, while density ( $\rho$ ) is held constant. For this reason every fourth row and column of the grid is green.

## Data and Methodology

### *Data Collection*

Table 1 and figure 2 show the locations of sites where data were collected for this study, highlighting in yellow sites for which subsurface models are discussed in this paper. Both ReMi and HVSR data were collected at all sites for subsequent separate and joint model optimization.

ReMi data collection utilized a 24-channel array of 4.5 Hz resonance frequency vertical geophones with an 8-meter spacing. At least 20 recordings with durations of 30 seconds were made utilizing an industry-standard DAQ-Link data logger.

Three-component broadband data for use in HVSR processing were collected using Nanometrics Trillium Compact seismometers and Reftek data loggers. These instruments were allowed to record for at least an hour at each location.

### *Data Processing*

#### *Refraction Microtremor (ReMi)*

ReMi data are processed using SeisOpt ReMi software using its standard built-in processing workflow. First, a  $p$ - $\tau$  transform is applied, and then the complex Fourier transform of each  $p$ - $\tau$  trace derived from the original data is computed. Finally, the power spectrum is calculated, which can be plotted with the axes of slowness (or velocity) vs. frequency. This process creates some artifacts, but these have a slope that is vastly different from the dispersion curve and are easily identifiable.

Once a slowness-frequency plot is generated, a dispersion curve is picked manually. In order to compensate for erroneous, fast velocities generated by off-axis noise sources

recorded by a linear array, the lowest apparent phase velocities are picked from the velocity spectral output plot.

#### *Horizontal-to-Vertical Spectral Ratios (HVSr)*

HVSr data processing is conducted using the Geopsy suite of programs. Segments of three-component broadband data with durations of at least 30 minutes are windowed automatically to remove unwanted transients detected in any of the three channels. H/V spectral ratios are calculated for each time window and plotted together, allowing for the user to confirm in the frequency domain that the effects of transients on data processing have been minimized. Because data acquisition for this study coincided with another broadband deployment, the final H/V spectral ratios from all stations in each small local array were averaged together. This step reduces the influence of local and instrumental noise at a single broadband station.

#### *Optimization through Very Fast Simulated Annealing (VFSA)*

Broad ranges of parameters were established for VFSA runs for all locations studied, due to the lack of preexisting geological constraints on shear wave velocities, layer thicknesses, and Poisson's ratios. All VFSA runs were started at a temperature of  $1 \times 10^8$  units, and allowed to cool to a dimensionless temperature of  $1e-21$  units over the course of 10,000 iterations, which provided ample opportunity for the misfit function to stabilize, indicating convergence on a (hopefully) global minimum (Figure 3). The starting temperature was tuned to produce rapid fluctuations in misfit at early iterations, indicating that high probabilities of model acceptance of models allows the search to escape local minima.

The first part of the optimization strategy focused on identifying a best-fit model within a broad parameter space as explained above, this was done to ensure that a single best-fit model could be used to generate forward models for both functionals that would produce minimal misfit and, especially in the case of HVSR, exhibit similar dominant frequency peaks as the observed data. After an acceptable model has been found, the parameter space is reset to include a  $\pm 20\%$  variation in all the parameters, and a set of 30 separate VFSA runs is started to search this new model space. This total of 300,000 iterations allows for a sufficient exploration of the parameter space to generate PPDs and parameter correlation matrices, as explained in section 3.

The VFSA algorithm employed allows for a large number of iterations to be run in short periods of time, producing statistical measures of uncertainty rapidly. When jointly optimizing both functionals, the average time per iteration averages approximately 1.2 seconds.

## Results

### *Este 3*

The left panel of figure 4 shows the final model (solid black line) obtained from joint optimization of both HVSR and surface wave functionals, bounded by the allowed search space shown by the blue dashed lines. The center and right panels superimpose the observed HVSR and surface wave data with the results of the forward calculation employing the best-fit model. There is general agreement in the HVSR peaks up to a frequency of approximately 3 Hz at which point the calculated H/V curve drops off. Whereas it still maintains a good correlation with the highest frequency peak in the

observed data in terms of frequency range, the amplitude does not match. The right panel shows a good fit between the observed and calculated surface wave dispersion curves. The average shear wave velocity of the top 30 m of this best-fit model is 954.52 m/s, placing this location within the National Earthquake Hazards Reduction Program (NEHRP) site class B range.

Uncertainty assessments for the Este 3 site subsurface model are shown in figures 5 and 6. The marginal posterior probability distribution (PPD) plots (figure 5) for the Este 3 site show a good progression in constraints from using single functionals to using both jointly in the optimization scheme. The high probability regions for each layer in the HVSR-only optimization are quite broad, especially in the shallower layers. Also, uncertainties in the SW-only optimization below 100 m are very large, and no single model stands out as most likely. When both functionals are employed in the joint scheme almost all fourteen layers' shear wave velocities are well constrained.

The parameter correlation matrix for HVSR-only optimization of the Este 3 site model (figure 6, left panel) shows strong tradeoffs between parameters in the deeper layers. Shear wave velocities ( $\beta$ ) in layers 9-14 seems particularly sensitive to variations in shear wave velocities of nearby layers. These tradeoffs are partially reduced by the joint optimization process, especially in layers 8-11. This indicates that including constraints induced by fitting surface waves reduces some parameter tradeoffs, particularly in shear wave velocities.

The surface wave-only optimization parameter correlation matrix also shows tradeoffs in shear wave velocities, but in this case they are mostly negative tradeoffs and occur in

shallower layers, where surface waves over the frequency range observed can constrain the subsurface model.

### *Airport 1*

The left panel of figure 7 shows the final model (solid black line) obtained from joint optimization of both HVSR and surface wave functionals, bounded by the allowed search space shown by the blue dashed lines. The center and right panels superimpose the observed HVSR and surface wave data with the results of the forward calculation employing the best-fit model. Overall, the shape of the calculated HVSR curve (red line, center panel) shows the same character as the observed HVSR curve. However, there is a slight offset in the frequency of these peaks, implying that the model used for forward calculations still does not match the subsurface accurately. The right panel shows a good fit between the observed and calculated surface wave dispersion curves. The average shear wave velocity of the top 30 m of this best-fit model is 976.80 m/s, placing this location within the NEHRP site class B range.

Statistical assessments for the Airport 1 site subsurface model are shown in figures 8 and 9. The marginal posterior probability distribution (PPD) plots (figure 8) for the Airport 1 site show a good progression in constraints from using single functionals to using both jointly in the optimization scheme. High probability regions for each layer in the HVSR-only optimization are quite broad, especially in the shallower layers. Also, uncertainties in the SW-only optimization below 100 m are large, and no single model explains the observations uniquely. When both functionals are employed in the joint scheme almost all the shear wave velocities in the 9 layers used in the optimization are well constrained, with the exception of a large layer (layer 7) extending from just above

1000 m through 1800 m, which appears to show poorer constraints in the joint modeling regime than in modeling with individual functionals.

The parameter correlation matrix for HVSR-only optimization of the Airport 1 site model (figure 9, left panel) shows strong tradeoffs between parameters in deeper layers. Shear wave velocities ( $\beta$ ) in layers 5-9 seem particularly sensitive to variations in shear wave velocities of nearby layers. These tradeoffs are partially reduced by the joint optimization process, especially in shallower layers. The parameter correlation matrix for the joint inversion provides insight on the poor constraints seen in layer 7 as explained above, showing strong tradeoffs with parameters in nearby layers.

#### *UASD*

The final model obtained for the Universidad Autonoma de Santo Domingo (UASD) site is shown by the solid black line in figure 10 (left panel); shear wave velocity search bounds are shown by the dashed blue lines. The model exhibits a monotonically increasing shear wave velocity with depth for the first 6 layers, followed by an abrupt drop in shear wave velocity in the seventh layer and a more constant shear wave velocity below. This impedance contrast emerged in the modeling as a necessity to fit the observed HVSR curve. The  $V_{s30}$  identified from joint modeling of both data functionals for the UASD site is 792.21 m/s, placing it in NEHRP site class B.

The calculated HVSR curve for the UASD site matches the overall character of the observed data, but fails to replicate some of the smaller changes in H/V ratios, especially at frequencies below 1 Hz.

As with the other sites, the misfit between the observed and calculated surface wave dispersion curves is minimal. In this case, there is a small departure between the

calculated and observed curves at the low end. This departure may be explained by the fact that the best-fit model satisfied constraints of HVSR rather than SW at the depths that would affect the very low frequency surface wave dispersion data.

The PPD plots for the UASD site (figure 11) show the clearest improvement in model constraints for the shallow subsurface when using both functionals jointly in the optimization strategy. In the top 6 layers, broad peaks are narrowed, and multimodal probability distributions are reduced. This improvement is also noticeable in layers 8 and 9, while layers extending below these to the bottom of the model show little or no improvement.

The parameter correlation matrices for the UASD site (figure 12) show similar features as those generated for other sites. Strong tradeoffs are seen between parameters, especially in deeper layers. These are mitigated slightly when using both functionals jointly for optimization, but many strong tradeoffs remain, especially between shear wave velocities ( $\beta$ ) of nearby layers.

### *Sur 1*

The final model obtained from joint optimization of both HVSR and SW functionals is shown in figure 13, left panel, accompanied by the allowed shear wave velocity search space, shown by the blue dashed lines. Also shown in figure 13 are comparison plots showing observed versus calculated HVSR curves (center panel) and observed versus calculated surface wave dispersion curves (right panel). The calculated HVSR curve shows good alignment with the observed data, namely in terms of the position of the dominant and secondary peaks. Some artifacts exist in the calculated HVSR curve, especially at frequencies above 1 Hz. The  $V_{s30}$  identified from joint

modeling of both data functionals for the Sur 1 site is 1059.23 m/s, placing it in NEHRP site class B.

As with other survey sites presented in this paper, the misfit between the observed and calculated surface wave curves is minimal, providing confidence in the subsurface model over the depths that are capable of affecting the range of surface wave dispersion frequencies observed in this study.

The PPD and parameter correlation matrix analysis of constraints (figures 14 and 15, respectively) for the Sur 1 site are not as convincing as those calculated for other sites. There is clear evidence in the PPDs that multimodal probability distributions were not reduced, and that in certain cases regions of high probabilities within certain layers were broadened.

These observations are reinforced by the parameter correlation matrices for the Sur 1 site (figure 15), which show stronger tradeoffs between more parameters when both functionals were employed jointly in optimization. As with other sites, shear wave velocities seem particularly subject to tradeoffs.

## Discussion

Overall, we are successful in fitting both HVSR and surface wave dispersion curve functionals to generate subsurface models that produce minimal misfits between the observed and calculated data. In some cases the character of the calculated HVSR curve differs in certain details from the observed data. We note cases in which the frequencies of calculated peaks do not exactly match those of the observed peaks. In some cases the frequencies of calculated and observed peaks match well, but with differing amplitude ratios. However, since Nakamura (2010) suggests that matching the

amplitude of horizontal to vertical spectral ratios is less critical for constraining subsurface structure than the frequencies of dominant peaks, we maintain confidence in this result.

All models produced in this study extend well beyond the depths usually studied for the purposes of seismic site characterization. Such studies usually extend to 30 m depth, or sometimes a few hundred, meters, yet we present subsurface models extending to depths of over 5,000 m in some cases. This is because we discovered that many of the lowest frequency peaks in observed HVSR data could only be matched by including model data at depths of several thousand meters. However, the PPDs estimates often show that parameters constraints at large depths are poor, so we do not regard the models to be reliable at depths greater than, for example 1500-2000 meters.

It is clear from the PPDs and parameter correlation matrices presented that the surface wave dispersion data presented in this paper have low sensitivity at depths greater than 100 m. PPDs computed with SW dispersion data alone show very poor constraints in layers at greater depths, illustrated by broad regions of high probability and multimodal solutions. Parameter correlation matrices for the SW-only dependent optimizations are misleading because they appear to show a lack of strong tradeoffs between parameters, especially in the case of the Sur 1 site. For many of the deeper layers at all 4 sites presented in this work, this merely represents the fact that tradeoffs do not exist between model parameters at depth because surface wave dispersion at the frequencies available to us is not sensitive to these depths. Due to this lack of sensitivity, a change to one parameter need not be compensated by a change to another parameter, resulting in green

parameter correlation matrices. This observation emphasizes that parameter correlations must be interpreted in conjunction with PPDs to assess constraints and model reliability.

Further, the statistical assessment tools shown here, the PPD and parameter correlation matrix, indicate that sensitivities of the two functionals employed in this study are concentrated in different depth ranges. Jointly modeling both functionals improves the reliability of models over the 0-1500 m depth range, as indicated by the PPDs shown here, but truly “complementary” functionals would have significant sensitivity to the same parameters, in the same depth ranges. Using lower frequency surface wave dispersion curves and/or high modes should produce greater overlap between the functionals’ sensitivities, providing better constraints for joint optimization. In the study presented here we did not have such dispersion data available to us.

Lastly, we note that, while modeling both functionals jointly modestly increases parameter independence, with respect to modeling functionals individually, in parts of the models, geophysical modeling is typically plagued by parameter tradeoffs. For example, data acquisition geometries are often limited to acquiring data from the Earth’s surface and data functionals often reduce complex waveform affects to a sparse set of summary observations, so a given parameter can be difficult to distinguish in one layer from an adjacent, or even distant, layer. Any success at reducing tradeoffs is therefore important.

## Conclusions

Incorporating model uncertainties is essential in geophysical applications and could benefit strategies for seismic site characterization. Adding additional data types, or “functionals”, should, in principle, improve constraints on model parameters, and therefore model reliability, but improvements are not guaranteed and strategies for conducting joint optimization are not obvious. Applying the sort of global optimization method described in this paper to the seismic site characterization problem, perhaps with the incorporation of other functionals or a broader frequency range of surface wave dispersion data could lead to a better understanding of subsurface characteristics in poorly constrained regions without significant increases in field or computational costs.

## Acknowledgments

We thank R.B. Herrmann and the GEOPSY consortium for providing us with their software. We thank the Incorporated Research Institutions for Seismology (IRIS) for its roles in acquiring, archiving, and disseminating the data used in this study and for organizing the PASI in Santo Domingo. This research was supported by funding from the National Nuclear Security Administration (Contract DE-AC52-09NA29327) and the W.M. Keck Foundation. This thesis is to be submitted for publication in *Geophysics* as Schwed, Pulliam, Sen, Huerta-López, Louie, Willemann, Schmitz, and Polanco Rivera, **Seismic Site Characterization through Joint Modeling of Complementary Data Functionals, with Applications to Santo Domingo, Dominican Republic.**

## References

- Aki, Keiiti. 1957. "Space and Time Spectra of Stationary Stochastic Waves, with Special Reference to Microtremors." *Bulletin of the Earthquake Research Institute of Tokyo University* 35 (3): 415–56.
- Arai, Hiroshi, and Kohji Tokimatsu. 2005. "S-Wave Velocity Profiling by Joint Inversion of Microtremor Dispersion Curve and Horizontal-to-Vertical (H/V) Spectrum." *Bulletin of the Seismological Society of America* 95 (5): 1766–78. doi:10.1785/0120040243.
- Asten, Michael W. 2001. "The Spatial Auto-Correlation Method for Phase Velocity of Microseisms—Another Method for Characterisation of Sedimentary Overburden: In Earthquake Codes in the Real World, Australian Earthquake Engineering Soc." In *Proceedings of the 2001 Conference, Canberra, Paper*. Vol. 28.
- Boore, David M., and Michael W. Asten. 2008. "Comparisons of Shear-Wave Slowness in the Santa Clara Valley, California, Using Blind Interpretations of Data from Invasive and Noninvasive Methods." *Bulletin of the Seismological Society of America* 98 (4): 1983–2003. doi:10.1785/0120070277.
- Borcherdt, R. D. 1970. "Effects of Local Geology on Ground Motion near San Francisco Bay." *Bulletin of the Seismological Society of America* 60 (1): 29–61.
- Huerta-Lopez, Carlos, Jay Pulliam, and Yosio Nakamura. 2003. "In Situ Evaluation of Shear-Wave Velocities in Seafloor Sediments with a Broadband Ocean-Bottom Seismograph." *Bulletin of the Seismological Society of America* 93 (1): 139–51. doi:10.1785/0120010276.
- Ingber, L. 1989. "Very Fast Simulated Re-Annealing." *Mathematical and Computer Modelling* 12 (8): 967–73. doi:10.1016/0895-7177(89)90202-1.
- — — 1993. "Simulated Annealing: Practice versus Theory." *Mathematical and Computer Modelling* 18 (11): 29–57. doi:10.1016/0895-7177(93)90204-C.
- Kausel, Eduardo, and José Manuel Roësset. 1981. "Stiffness Matrices for Layered Soils." *Bulletin of the Seismological Society of America* 71 (6): 1743–61.
- Louie, J., S. Pullammanappallil, A. Pancha, T. West, and W. Hellmer. 2011. "Earthquake Hazard Class Mapping by Parcel in Las Vegas Valley." In *Structures Congress 2011*, 1794–1805. American Society of Civil Engineers.
- Louie, John N. 2001. "Faster, Better: Shear-Wave Velocity to 100 Meters Depth from Refraction Microtremor Arrays." *Bulletin of the Seismological Society of America* 91 (2): 347–64. doi:10.1785/0120000098.

- Nakamura, Yutaka. 1989. "A Method for Dynamic Characteristics Estimation of Subsurface Using Microtremor on the Ground Surface." *Railway Technical Research Institute, Quarterly Reports* 30 (1).
- — — 2010. "Comment on 'Microtremor Measurements in the Nile Delta Basin, Egypt: Response of the Topmost Sedimentary Layer' by E. A. Fergany and S. Bonnefoy-Claudet." *Seismological Research Letters* 81 (2): 241–43.  
doi:10.1785/gssrl.81.2.241.
- Park, Choon B., Richard D. Miller, and Jianghai Xia. 1999. "Multichannel Analysis of Surface Waves." *GEOPHYSICS* 64 (3): 800–808. doi:10.1190/1.1444590.
- Sen, M. K., and P. L. Stoffa, 1994, Bayesian inference, Gibbs sampler and uncertainty estimation in nonlinear geophysical inversion, EAGE 56th Conference and Technical Exhibition, Vienna, Austria.
- Stoffa, P. L., M. K. Sen, C. Varela and R. K. Chunduru, 1994, Geophysical applications of global optimization methods, EAGE 56th Conference and Technical Exhibition, Vienna, Austria.
- Sen, M. K., and P. L. Stoffa. 1995. *Global Optimization Methods in Geophysical Inversion*. Elsevier.
- Sen, Mrinal K., and Paul L. Stoffa. 2013. *Global Optimization Methods in Geophysical Inversion*. Cambridge University Press.
- Willemann, R. J., J. Pulliam, E. Polanco, J. N. Louie, C. Huerta-Lopez, M. Schmitz, M. P. Moschetti, V. Huerfano Moreno, and M. Pasyanos. 2013. "Bringing New Tools and Techniques to Bear on Earthquake Hazard Analysis and Mitigation." *AGU Fall Meeting Abstracts* 43 (December): 0782.
- Zhao, Lian-She, Mrinal K. Sen, Paul Stoffa, and Cliff Frohlich. 1996. "Application of Very Fast Simulated Annealing to the Determination of the Crustal Structure beneath Tibet." *Geophysical Journal International* 125 (2): 355–70.  
doi:10.1111/j.1365-246X.1996.tb00004.x.

Tables

Site	Latitude (°)	Longitude (°)	Elevation (m)	Vs30 (m/s)
Airport 1	-69.9687	18.4720	67	976.80
Airport 2	-69.9691	18.4706	68	1137.09
Este 1	-69.8699	18.4782	37	878.39
Este 2	-69.8683	18.4796	37	1137.45
Este 3	-69.8669	18.4791	37	954.52
Sur 1	-69.9813	18.4354	58	1059.23
Sur 2	-69.9803	18.4357	43	1433.45
UASD	-69.9141	18.4619	19	792.21

Table 1. Table of locations and elevations of survey sites in the city of Santo Domingo, with their Vs30 values derived from the best model in joint optimization.

## Figures

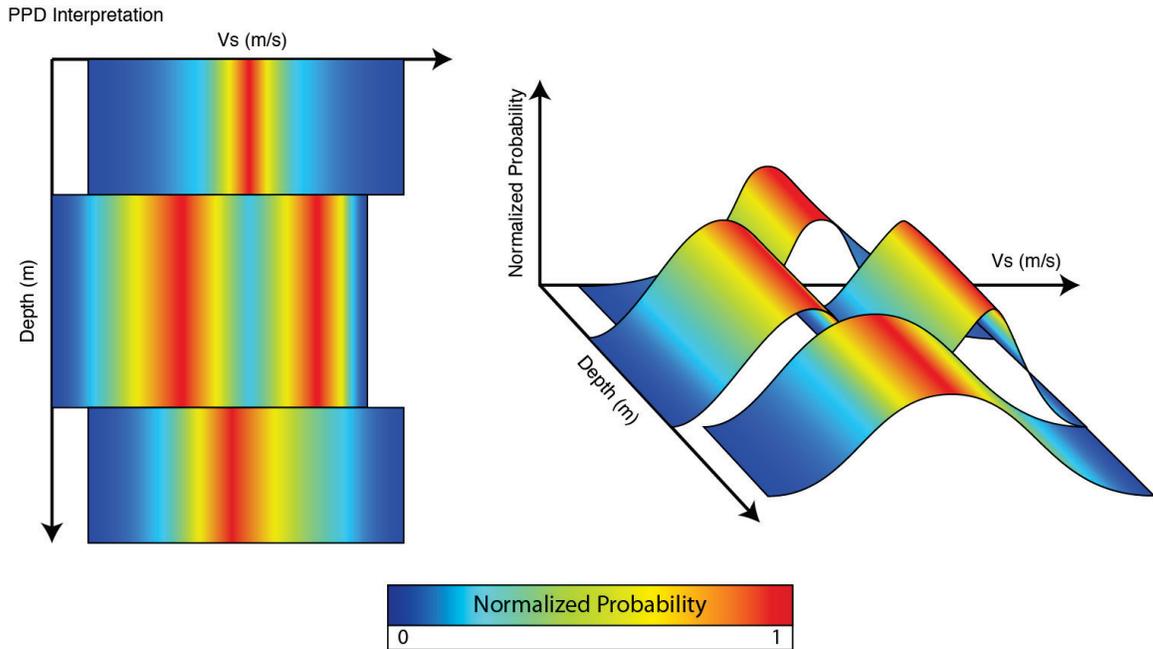


Figure 1. Explanation of marginal Posterior Probability Density (PPD) plots. Colored regions indicate search bands with probability of acceptable parameter values ranging from blue (low) to red (high). The example shown here indicates that  $V_s$  is well-constrained in the first layer, relatively poorly in layer 3 (compared to the first layer), and ambiguous in layer 2. Multiple peaks indicate that very different values can fit the data nearly as well. An alternative explanation is shown at right in which the color within each band represents a normalized probability on the vertical axis.

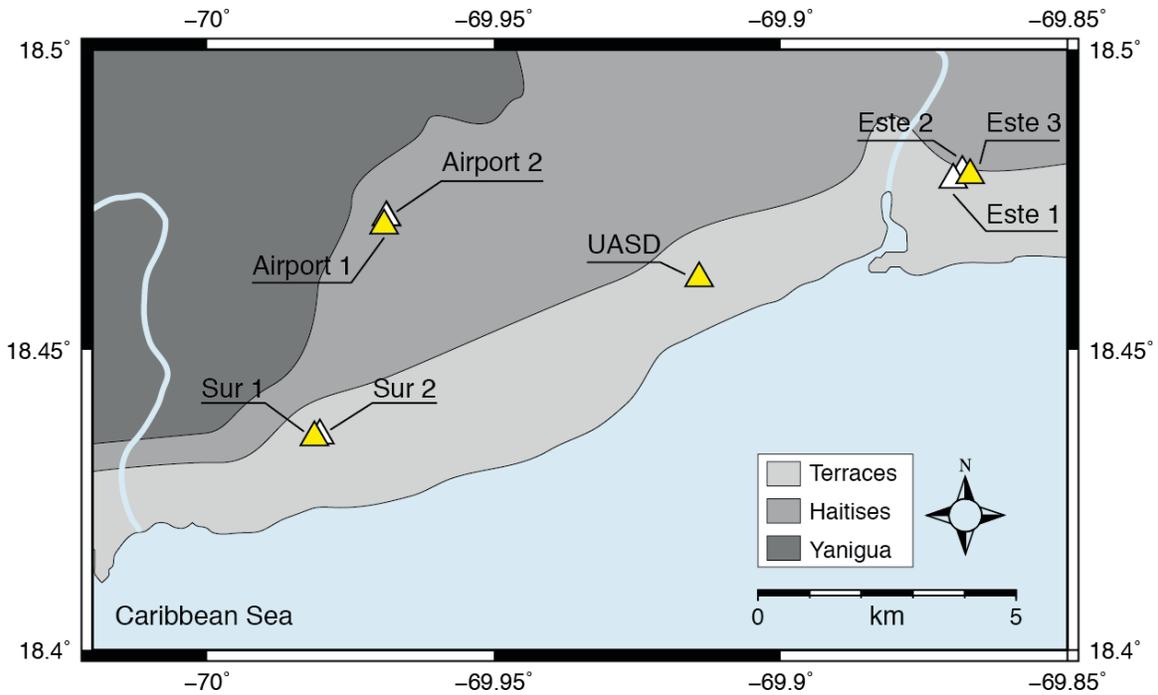


Figure 2. Distribution of the deployment locations (white triangles), and those highlighted in this paper (yellow triangles) throughout the city of Santo Domingo. Also indicated are basic lithologic units as differentiated by the Dominican National Geological Survey's Sheet 62713. Starting in the south there are carbonate terraces, followed by the Haisites formation, consisting of reefal limestones. Furthest northwest in this focus region is the Yanugya formation, which could consist in this region of both loams and limestones or sands, silts, and gravels.

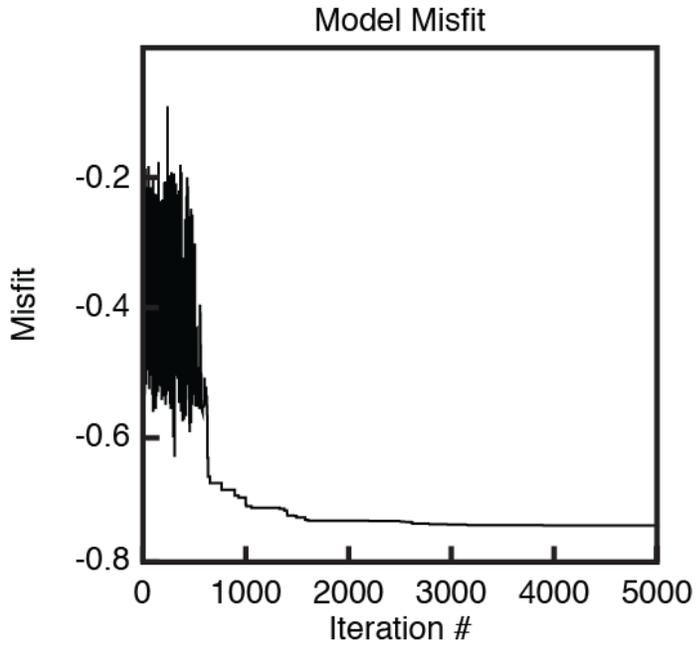


Figure 3. Plot of model misfit as a function of iteration number for a joint inversion run at the Airport 1 site. The misfit shows a large amount of fluctuation before approximately iteration number 750, which indicates that a high enough starting temperature was chosen in the cooling schedule to allow for models with higher misfits to be accepted and thus conduct a broad search of the model space.

### Este 3 Functional Fits

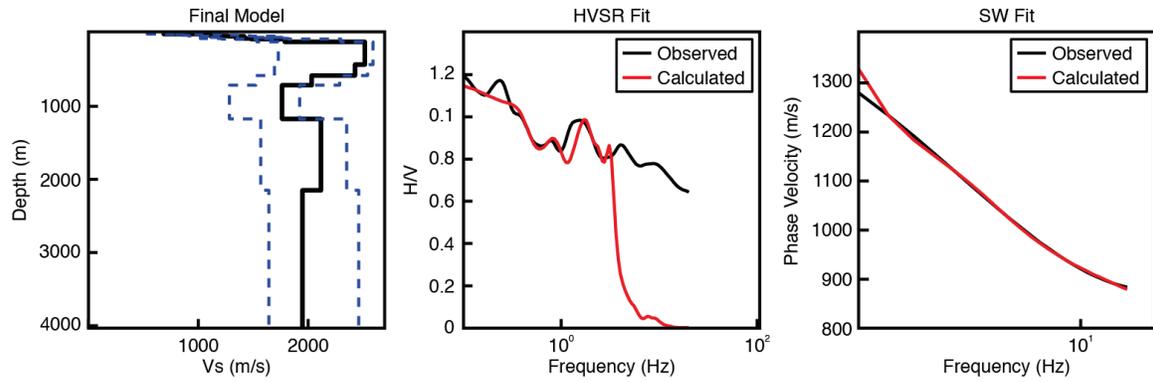


Figure 4. The final best-fit model for the Este 3 site (solid black line, right panel), as well as the allowed search space indicated by the blue dashed lines. The center and right panels show the match between the observed and calculated HVSR and surface wave dispersion functionals, respectively. Note that the PPDs (figure 5) reflect probabilities (rather than certainties) that a model parameter takes on particular values and that the single, best-fit model need not (and, in fact, rarely will) follow the high probability regions through the PPD.

Este 3 PPDs

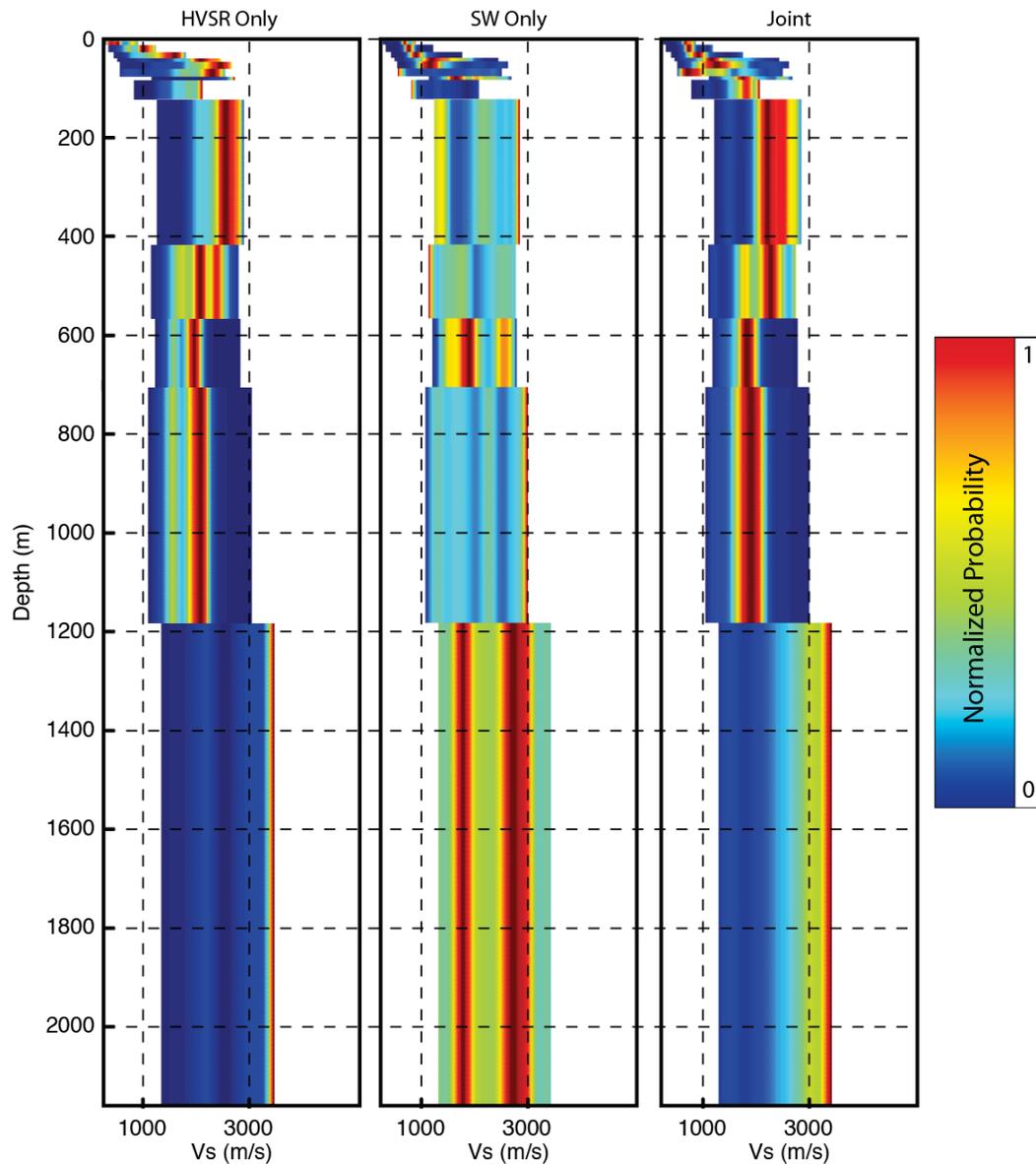


Figure 5. Plots of marginal Posterior Probability Density for the Este 3 site. The left and center plots show the PPD for the optimization of single functionals (HCSR and surface waves, respectively), and the right plot shows the PPD for joint optimization. Normalized probability is represented by a color scale from blue to red, with blue showing the lowest probability density and red showing the highest probability density. Note that the HCSR modeling shows broad and/or multi-valued peaks for shallow layers but narrower, single-valued peaks for deeper layers whereas the (fundamental mode) SW dispersion results show the opposite case. These PPDs display graphically the fact that HCSR constrains deeper layers and SW constrains shallower layers, suggesting that jointly modeling the two functionals should produce tighter constraints, and more reliable models, than modeling either functional alone. The PPD for joint modeling (right) confirms this expectation.

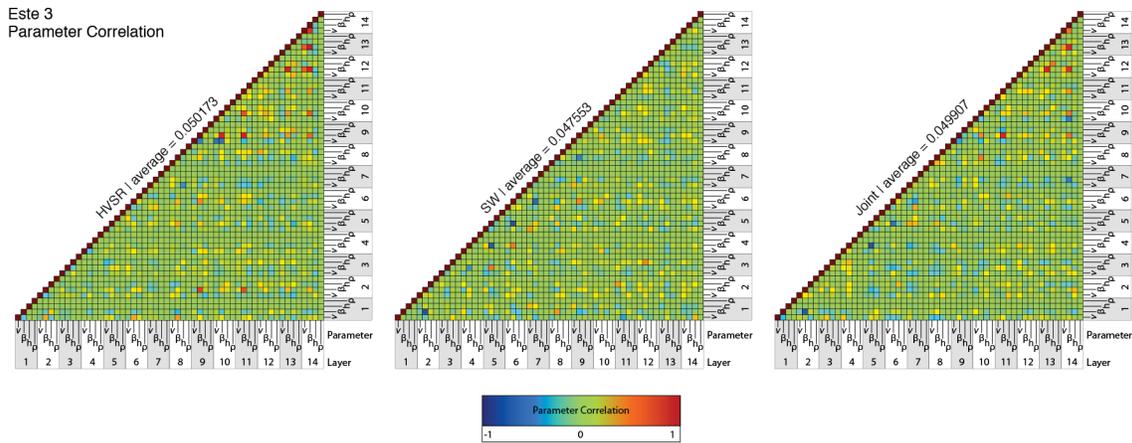


Figure 6. Parameter correlation matrices for the Este 3 site. The left and center plots show the correlation matrix for the optimization of single functionals (HVSF and surface waves, respectively), and the right plot shows the correlation matrix for joint optimization. Average parameter correlation values are computed for each, and displayed along the diagonal. Note that, because density is fixed, correlations between density and other parameters are identically zero but not meaningful. Results for many of the parameter correlation matrices suggest that they underestimate parameter dependence for cases in which constraints, as indicated by PPD results, are poor. When constraints are poor, sensitivity to parameters in certain layers can be low, so tradeoffs between those parameters are also low. This is misleading, in the sense that model reliability is also low. Parameter correlation matrices must therefore be evaluated in the context of PPDs and model reliability can only be assessed with the help of both tools.

Airport 1 Functional Fits

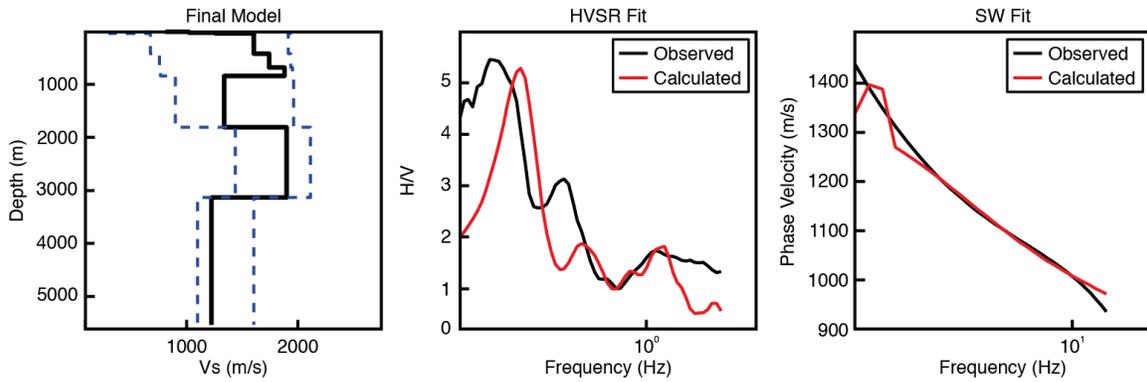


Figure 7. The final best-fit model for the Airport 1 site (solid black line, right panel), as well as the allowed search space indicated by the blue dashed lines. The center and right panels show the match between the observed and calculated HVSF and surface wave dispersion functionals, respectively. The final best-fit model exhibits a monotonic increase in shear wave velocity with depth for the first 800 m, and then large scale fluctuations at and below 1000 m. The best-fit model is well within the allowed parameter search space indicated by the blue dashed lines. There is good agreement between the observed and calculated HVSF curves, although with some slight offset in the calculated and observed frequencies of the fundamental peak and a secondary peak at a higher frequency. The calculated and observed surface wave dispersion curves match well, albeit with some discrepancies in lower frequencies.

Airport 1 PPDs

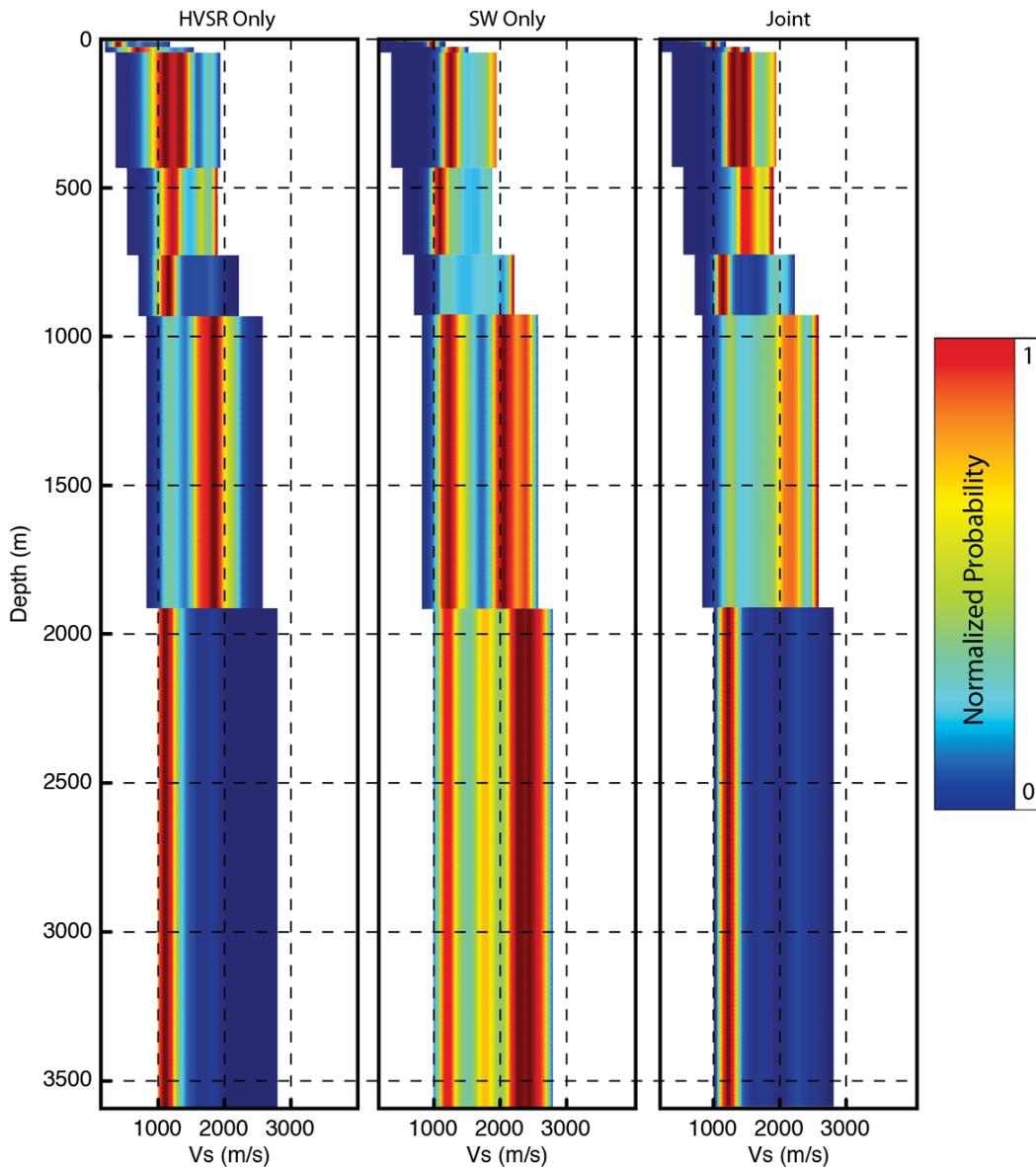


Figure 8. Plots of marginal Posterior Probability Density for the Airport 1 site. The left and center plots show the PPD for the optimization of single functionals (HCSR and surface waves, respectively), and the right plot shows the PPD for joint optimization. Normalized probability is represented by a color scale from blue to red, with blue showing the lowest probability density, and red showing the highest probability density. The PPD for HCSR shows broad peaks in several layers, notably in the layer between 75 and 400 m. This layer becomes better constrained when employing both data functionals for the optimization. Many of the deeper layers are very poorly constrained in the surface wave dispersion optimization, but appear to be better constrained in the HCSR-only optimization. When jointly optimized, constraints show a marked improvement.

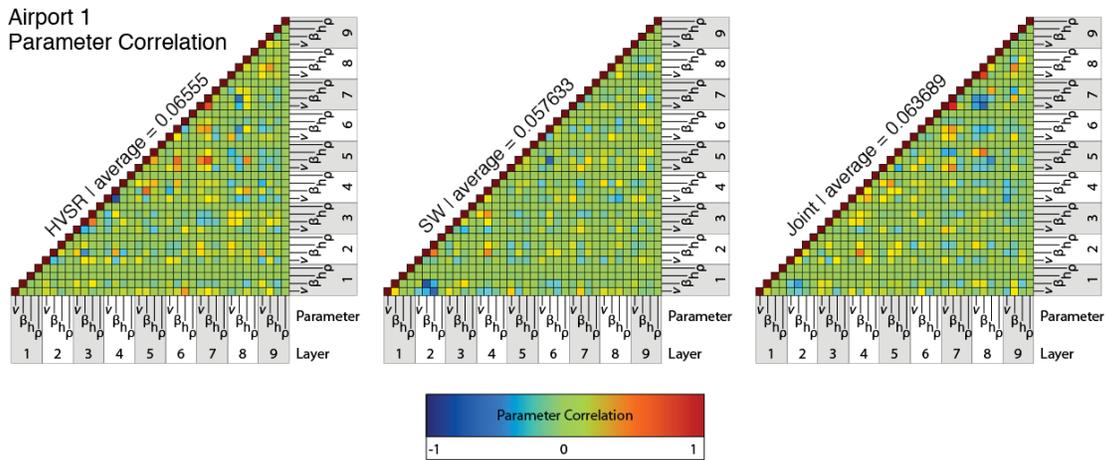


Figure 9. Parameter correlation matrices for the Airport 1 site. The left and center plots show the correlation matrix for the optimization of single functionals (HVSER and surface waves, respectively), and the right plot shows the correlation matrix for joint optimization. Average parameter correlation values are computed for each, and displayed along the diagonal. The strongest tradeoffs in the HVSER parameter correlation plot appear between shear wave velocities of adjacent or nearby layers, especially at and below layer 4. In the case of the surface wave parameter correlation plot, the largest tradeoffs are seen between shear wave velocities and Poisson's ratios in shallow layers. Joint optimization mitigates some of these strong tradeoffs, but because surface wave data is not helping to constrain the deeper parts of the model, the high parameter correlation values remain for deeper layers, apparently as a consequence of HVSER optimization.

UASD Functional Fits

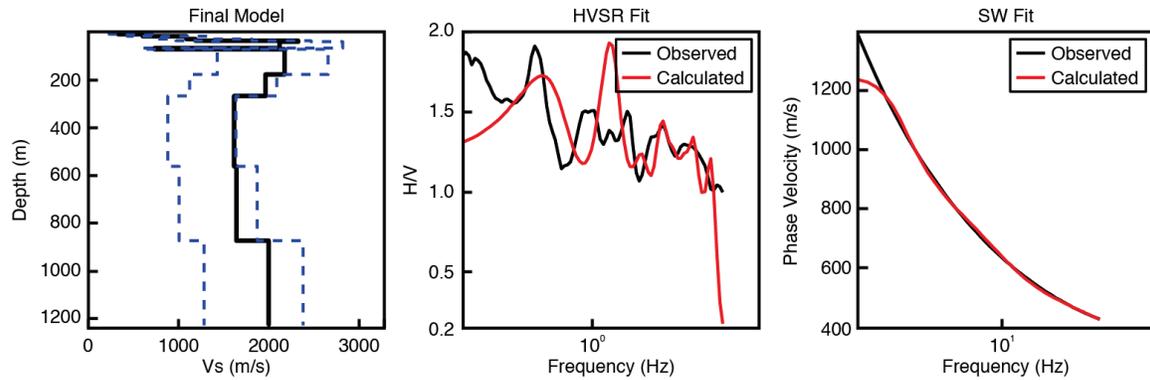


Figure 10. The final best-fit model for the UASD site (solid black line, right panel), as well as the allowed search space indicated by the blue dashed lines. The center and right panels show the match between the observed and calculated HVSF and surface wave dispersion functions, respectively. Although the best-fit model exhibits some very large shear wave velocity fluctuations over a small depth range, the observed and calculated HVSF and surface wave dispersion curves show a good agreement overall, with some exceptions in the lower frequency range of both.

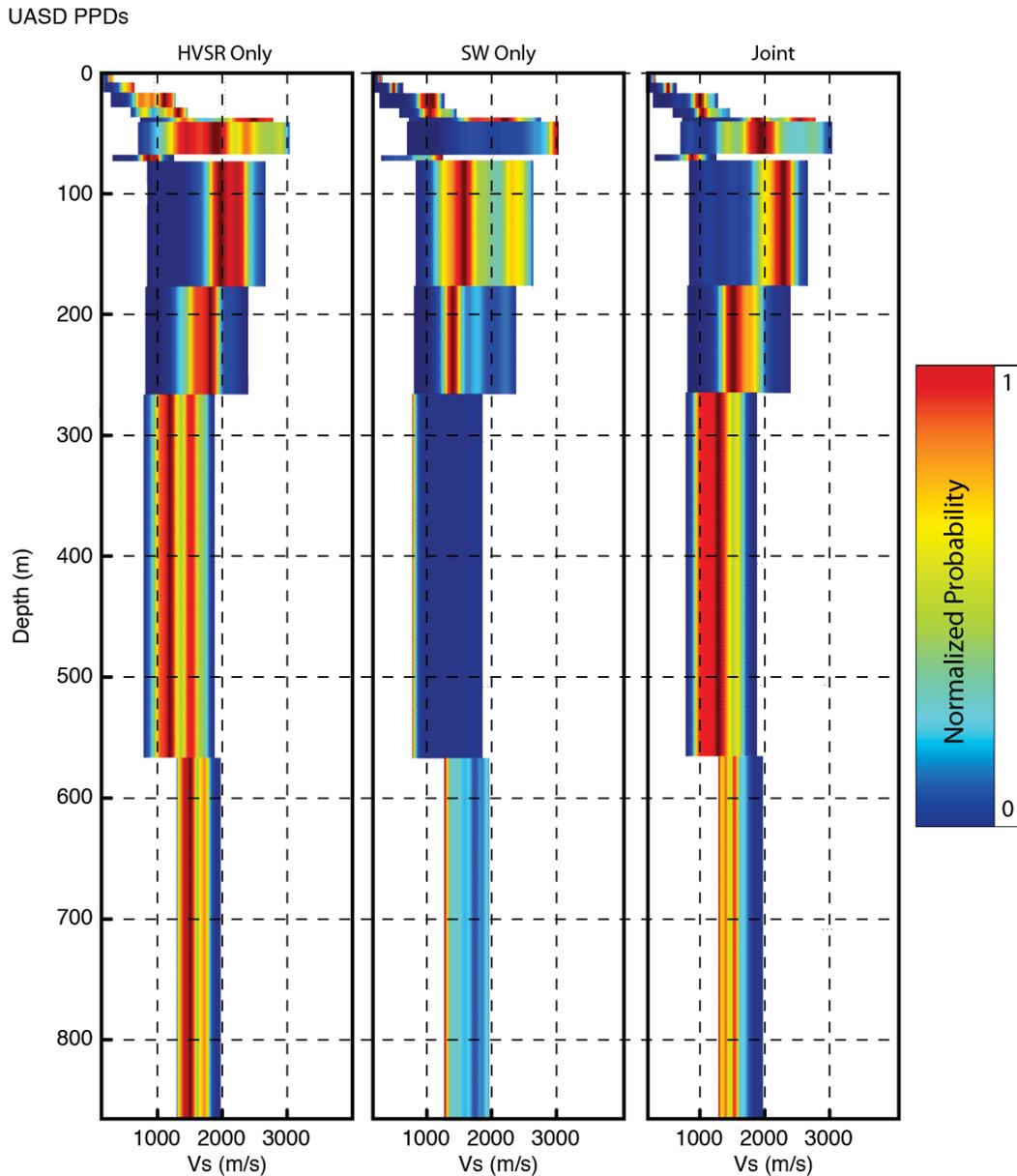


Figure 11. Plots of marginal Posterior Probability Density for the UASD site. The left and center plots show the PPD for the optimization of single functionals (HVSr and surface waves, respectively), and the right plot shows the PPD for joint optimization. Normalized probability is represented by a color scale from blue to red, with blue showing the lowest probability density, and red showing the highest probability density. Up to a depth of 75 m the optimization employing both data functionals is clearly able to better constrain the possible range of shear wave velocities. As with the other sites, the HVSr PPD exhibits many multimodal and broad distributions, and the surface wave PPD shows no constraints with depth.

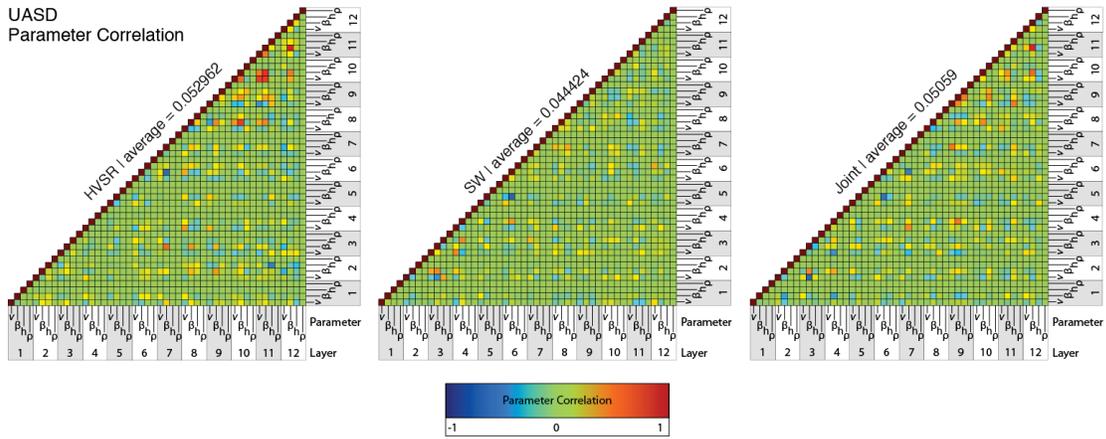


Figure 12. Parameter correlation matrices for the UASD site. The left and center plots show the correlation matrix for the optimization of single functionals (HVSR and surface waves, respectively), and the right plot shows the correlation matrix for joint optimization. Average parameter correlation values are computed for each, and displayed along the diagonal. As is seen in the other sites, the parameters exhibiting the greatest amounts of correlation are shear wave velocities and Poisson's ratios in the deeper layers of the HVSR-only parameter correlation matrix. These are again reduced somewhat in the joint inversion.

Sur 1 Functional Fits

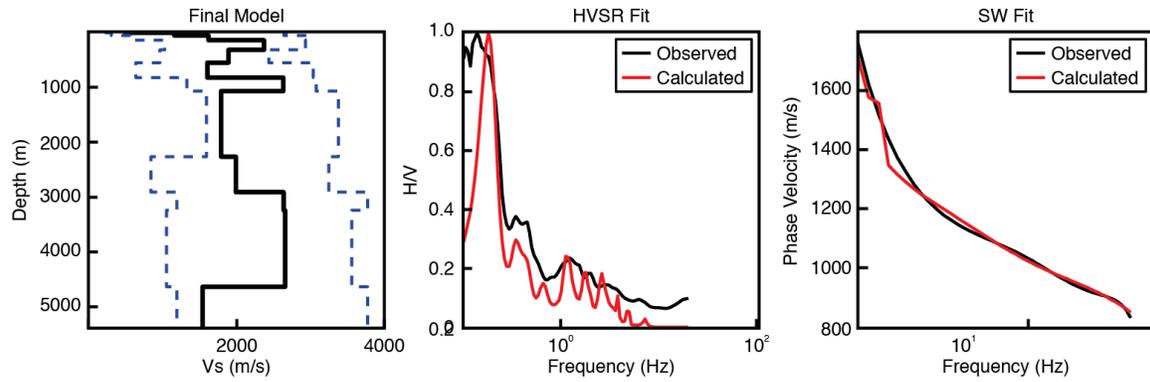


Figure 13. The final best-fit model for the Sur 1 site (solid black line, right panel), as well as the allowed search space indicated by the blue dashed lines. The center and right panels show the match between the observed and calculated HVSR and surface wave dispersion functionals, respectively. The shear wave velocity varies considerably with depth, and besides 3 layers at the top that exhibit a monotonically increasing velocity with depth, do not follow any particular pattern. The fits of observed versus calculated HVSR and surface wave dispersion curves are very good for this site.

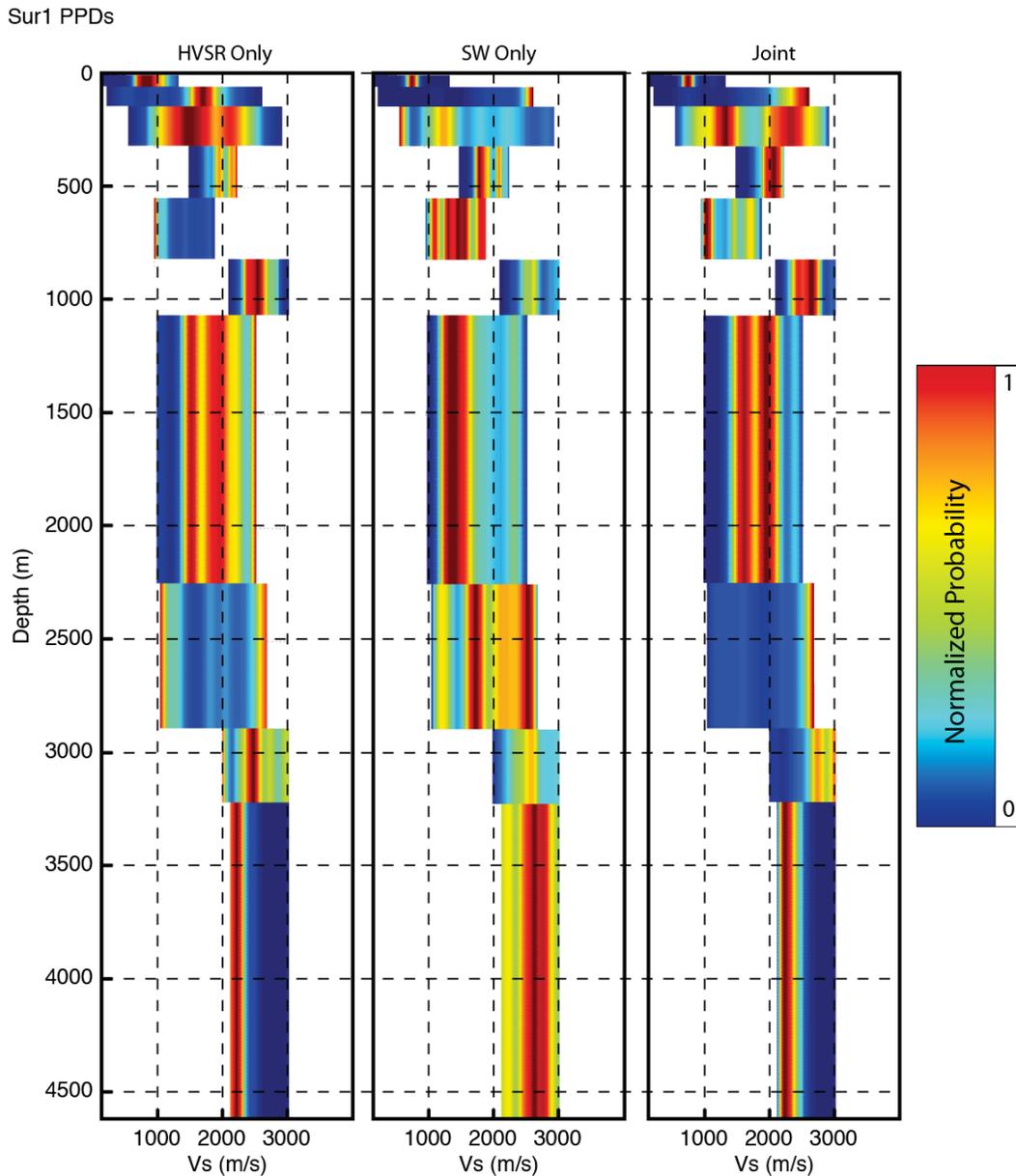


Figure 14. Plots of marginal Posterior Probability Density for the Sur 1 site. The left and center plots show the PPD for the optimization of single functionals (HVSr and surface waves, respectively), and the right plot shows the PPD for joint optimization. Normalized probability is represented by a color scale from blue to red, with blue showing the lowest probability density, and red showing the highest probability density. In this case, there is not a clear improvement in the PPD when using both functionals jointly in the optimization. Broad distributions are not further constrained, and in certain layers the joint optimization PPD shows more uncertainty in shear wave velocity than in the HVSr or surface wave dispersion optimizations alone.

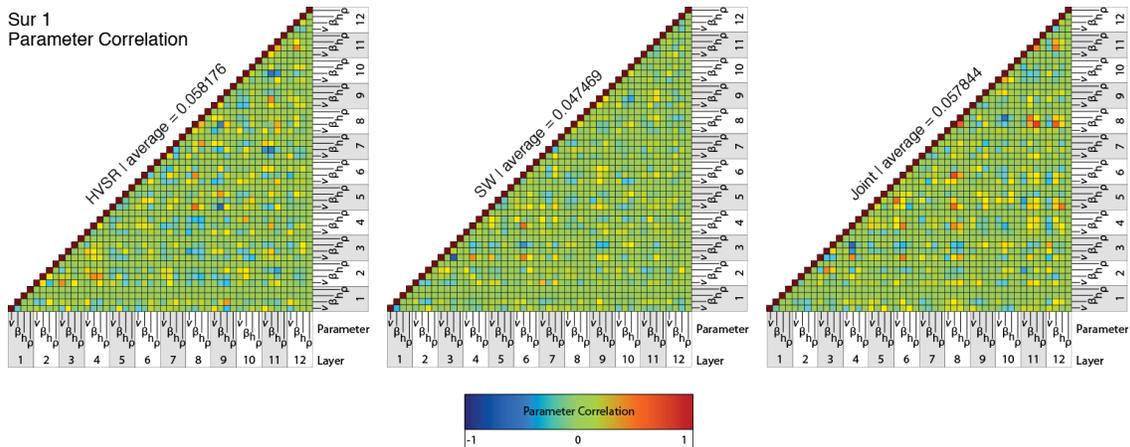


Figure 15. Parameter correlation matrices for the Sur 1 site. The left and center plots show the correlation matrix for the optimization of single functionals (HVSR and surface waves, respectively), and the right plot shows the correlation matrix for joint optimization. Average parameter correlation values are computed for each, and displayed along the diagonal. The parameter correlation matrix for the joint optimization in this case appears to simply inherit the strong tradeoffs from HVSR and surface wave dispersion alone, instead of showing a reduction in them as is seen in other sites.