Towards Gaussian Process Regression Modeling of Simulation-Based
Ground Motion Coherency Functions

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in
Civil and Environmental Engineering
by
Stephen Waldvogel

Dr. Elnaz Seylabi/ Thesis Advisor

May 2022
THE GRADUATE SCHOOL

We recommend that the thesis prepared under our supervision by

Stephen Waldvogel

entitled

Towards Gaussian Process Regression Modeling of Simulation-Based Ground Motion Coherency Functions

be accepted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Elnaz Ezmailzadeh Seylabi, Ph.D. 
Advisor

David McCallen, Ph.D. 
Committee Member

Michael H. Gardner, Ph.D. 
Graduate School Representative

David W. Zeh, Ph.D., Dean 
Graduate School

May, 2022
Abstract

As the design of long-span structures will be subjected to different seismic loading across the entirety of its span (a.k.a multi-support excitation), determining how those ground motions will change is critical for seismic analysis of distributed infrastructure. Due to the lack of densely recorded ground motion databases, ground motions are commonly interpreted from known to unknown points along the span of the structure using ground motion coherency. The coherency is a statistical approach for measuring the similarity of ground motions between two geospatial points. Several methods for approximating coherency have been developed using empirical and semi-empirical functions. These models are developed mainly based on a limited number of dense arrays and simplified wave propagation theories and, therefore, lack the capabilities of modeling source, path, and site complexities and fail to provide consistently accurate values. In the past, many studies have been performed to see how physical parameters may influence coherency loss of ground motions, and many have found importance from numerous site, path, and source effects. This research aims to utilize the results of broadband deterministic earthquake simulations to explore how different physical parameters influence the loss of ground motion coherency as a function of frequency and how machine learning can be used to provide forward progress towards developing non-ergodic and physics-based coherency models.

SW4, a fourth-order finite difference code, is used for broadband deterministic earthquake simulations in the Bay Area region and for generating dense arrays of ground motions for coherency analysis. Three of the SW4 simulations investigated are magnitude 4.5 point-source events, and the fourth is a simulated magnitude 6.5 fault rupture event. A machine learning technique based on Gaussian Process Regression (GPR) is used to train non-parametric
coherency models and measure the correlation between different physical parameters including frequency, spacing, $V_{S30}$, and station and source positions on coherency. In total, eight GPR models were developed to measure coherency across the region of interest. The first model looks at the first SW4 point source simulation and forms coherency predictions using input features longitude, latitude, separation distance, and frequency, conditioned against a training set composed of observed data from the same simulation. Models 2 through 5 investigate 5-feature model accuracy for all three point-source events and the simulated magnitude 6.5 fault rupture event, respectively. These five features include station longitude and latitude, separation distance, frequency, and $V_{S30}$. Like model 1, models 2 through 5 are conditioned against training sets composed of observed data from their respective simulations. Models 6 through 8 investigate the 6-feature model accuracy of each point source event and includes input features of station longitude and latitude, separation distance, frequency, $V_{S30}$, and station distance from the point-source epicenter. Unlike the first 5 GPR models, models 6 through 8 are conditioned against a training set composed of data from all three point sources to see how well they performed when the training set consisted of greater uncertainty. Results found accurate fitting in all eight models, with a better fit observed on the 5-feature models when compared to 4-feature and 6-feature predictions.
Acknowledgments

This research was supported by the Southern California Earthquake Center (SCEC) through award numbers 20064 and 21078. SCEC is funded by NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047. I am grateful for the financial support that has made this research possible.

A special thank you to Dr. Elnaz Seylabi for providing me support, mentoring, and career path advice over the course of my graduate program. Dr. Seylabi has pushed me towards achievements I would never have thought possible. I am honored to have worked beside her throughout my educational career at the University of Nevada, Reno.

I also thank Dr. Arben Pitarka from Lawrence Livermore National Laboratory and Dr. David McCallen for facilitating large-scale deterministic ground motion simulations in the Bay Area region used for the development of coherency data sets. Your support has given me the tools necessary for performing my research.

To my committee members Dr. David McCallen and Dr. Michael Gardner, thank you for the support in the reviewal and completion of my research. I am blessed to have assistance from both of you as I finish my graduate degree.

Last but certainly not least, a much needed thank you to my family. The never-ending encouragement and support you have provided me through my studies will always be remembered. Having each of you present in my life has given me the strength to finish my studies.
# Table of Contents

Abstract ................................................................................................................................. i
Acknowledgments.................................................................................................................. iii
Table of Contents .................................................................................................................. iv
List of Tables ......................................................................................................................... vi
List of Figures ....................................................................................................................... vii
1 Introduction ....................................................................................................................... 1
  1.1 Problem description ...................................................................................................... 1
  1.2 Objectives and Scope of Work .................................................................................... 2
  1.3 Organization of Thesis ............................................................................................... 4
2 Ground Motion Coherency ................................................................................................. 6
  2.1 An overview ............................................................................................................... 6
  2.2 Empirical and semi-empirical coherency models ....................................................... 8
  2.3 Source, path, and site effects on ground motion coherency ....................................... 17
3 Gaussian Process Regression Modeling ........................................................................... 26
4 Simulation-Based Coherency Function Analysis ............................................................. 40
  4.1 Coherency Function Analysis ...................................................................................... 42
  4.2 Coherency Function Results ....................................................................................... 46
5 Gaussian Process Regression Model Development for Simulation-based Coherency Functions ......................................................................................................................... 76
  5.1 Software Packages ..................................................................................................... 76
  5.2 Model Development .................................................................................................... 79
    5.2.1 Data Set, Model Setup, and Input Requirements .................................................. 79
    5.2.2 Model Training .................................................................................................... 81
  5.3 Model Predictions ..................................................................................................... 115
    5.3.1 4-Feature Model Predictions for Event 1 ............................................................. 117
    5.3.2 5-Feature Model Predictions for Event 1 ............................................................. 128
    5.3.3 5-Feature Model Predictions for Event 2 ............................................................. 138
    5.3.4 5-Feature Model Predictions for Event 3 ............................................................. 148
    5.3.5 5-Feature Model Predictions for Event M6.5 ...................................................... 158
    5.3.6 6-Feature Model Predictions for Event 1 ............................................................. 168
    5.3.7 6-Feature Model Predictions for Event 2 ............................................................. 178
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.8</td>
<td>6-Feature Model Predictions for Event 3</td>
<td>188</td>
</tr>
<tr>
<td>5.4</td>
<td>Discussion of Model Predictions</td>
<td>198</td>
</tr>
<tr>
<td>6</td>
<td>Summary and Conclusions</td>
<td>201</td>
</tr>
<tr>
<td>6.1</td>
<td>Summary</td>
<td>201</td>
</tr>
<tr>
<td>6.2</td>
<td>Conclusions</td>
<td>202</td>
</tr>
<tr>
<td>7</td>
<td>References</td>
<td>204</td>
</tr>
<tr>
<td></td>
<td>Appendix A: Informative MATLAB Scripts</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>Appendix B: Informative Python Syntax</td>
<td>216</td>
</tr>
<tr>
<td></td>
<td>Appendix C: Supplemental Coherency Analysis Results</td>
<td>221</td>
</tr>
<tr>
<td></td>
<td>Appendix D: Supplemental GPR Results</td>
<td>230</td>
</tr>
</tbody>
</table>
List of Tables

Table 2-1: Site categorization for rock and soil arrays (*from Schneider et al., 1992*) .................. 24
Table 2-2: Ground motion records used for understanding the effects of soil composition on lagged coherency (*from Schneider et al., 1992*). ................................................................. 24
Table 4-1: MATLAB Coherency Results ......................................................................................... 45
Table 4-2: Station distances to each ground motion event. .............................................................. 67
Table 5-1: Summary of converged kernel hyperparameters from each model ......................... 90
List of Figures

Figure 2-1: Configuration of the SMART-1 Array, Taiwan (from “The SMART-1 Array”). .................................................. 11
Figure 2-2: Lotung LSST placement with respect to the SMART-1 array (from “The SMART-1 Array”). .................................................. 11
Figure 2-3: Lotung LSST Surface Accelerometers (from “The LLSST Array”). ................................................................. 12
Figure 2-4: Mean arctanh(Coherency) calculated from the Lotung LSST Array (from Abrahamson, 1990). .................................................. 13
Figure 2-5: Wave passage and scattering effects of strong ground motions (from Zerva, 2009). .................................................. 19
Figure 2-6: Four-layer canyon used for ground motion coherency analysis (from Bi et al., 2010). .................................................. 21
Figure 2-7: Comparison of the mean lagged coherency between the surface motions (j,k) with that of the incident motion on the base rock: (a) horizontal out-of-plane motion; (b) horizontal in-plane motion; and (c) vertical in-plane motion (from Bi et al., 2010). .................................................. 22
Figure 2-8: Modulus of the site amplification spectral ratio of two local sites: (a) horizontal out-of-plane motion; (b) horizontal in-plane motion; and (c) vertical in-plane motion. (from Bi et al., 2010). .................................................. 23
Figure 3-1: Illustrative example of mean function distribution across objective function (from Ding, 2018). .................................................. 29
Figure 3-2: A Gaussian prior formed from the product of two one-dimensional input features using SE-ARD kernel (from Duvenaud, 2014). .................................................. 30
Figure 3-3: 1-dimensional noisy data set used for training. .................................................. 32
Figure 3-4: Optimized kernel predictions from noisy sin(x) function. .................................................. 33
Figure 3-5: Model fit from optimized kernel predictions. .................................................. 34
Figure 3-6: Kernel predictions from noisy sin(x) function with length scale set to 5.0. .................................................. 35
Figure 3-7: Model fit from kernel predictions with length scale set to 5.0. .................................................. 35
Figure 3-8: Kernel predictions from noisy sin(x) function with length scale set to 0.1. .................................................. 36
Figure 3-9: Model fit from kernel predictions with length scale set to 0.1. .................................................. 37
Figure 3-10: 2-dimensional noisy data set used for training. .................................................. 38
Figure 3-11: Predictions from training on f(x,y) = y² - x² with addition of noise. .................................................. 38
Figure 3-12: Predictions versus true values of f(x,y) = y² - x² .................................................. 39
Figure 4-1: Region of interest and simulated earthquake events from SW4. .................................................. 42
Figure 4-2: Coherency averaging at master node ‘i’ with respect to separation distance. .................................................. 44
Figure 4-3: a) Folder naming convention with respect to station node b) File naming convention for coherency results with respect to distance evaluated. .................................................. 46
Figure 4-4: Average shear wave velocities for the top 30 meters of soil within the region of interest. .................................................. 50
Figure 4-5: Velocity history of Event 1 across the region of interest. .................................................. 51
Figure 4-6: Velocity history of Event 2 across the region of interest. .................................................. 52
Figure 4-7: Velocity history of Event 3 across the region of interest. .................................................. 53
Figure 4-8: Velocity history of the magnitude 6.5 event across the region of interest. .................................................. 54
Figure 4-9: Coherency results from event one with station separation distance equal to 200 meters. .................................................. 59
Figure 4-10: Coherency results from event one with station separation distance equal to 600 meters. ................................. 59
Figure 4-11: Coherency results from event one with station separation distance equal to 1000 meters. ........................................ 60
Figure 4-12: Coherency results from event two with station separation distance equal to 200 meters. ........................................ 60
Figure 4-13: Coherency results from event two with station separation distance equal to 600 meters. ........................................ 61
Figure 4-14: Coherency results from event two with station separation distance equal to 1000 meters. ........................................ 61
Figure 4-15: Coherency results from event three with station separation distance equal to 200 meters. ........................................ 62
Figure 4-16: Coherency results from event three with station separation distance equal to 600 meters. ........................................ 62
Figure 4-17: Coherency results from event three with station separation distance equal to 1000 meters. ........................................ 63
Figure 4-18: Coherency results from event M6.5 with station separation distance equal to 200 meters. ........................................ 64
Figure 4-19: Coherency results from event M6.5 with station separation distance equal to 600 meters. ........................................ 65
Figure 4-20: Coherency results from event M6.5 with station separation distance equal to 1000 meters. ........................................ 65
Figure 4-21: Mean coherency loss across the region of interest with respect to frequency. ...... 66
Figure 4-22: Coherency function results for station S026.074_03 ................................................ 68
Figure 4-23: Coherency function results for station S026.006_03 ................................................ 69
Figure 4-24: Coherency function results for station S005.028_03 ................................................ 70
Figure 4-25: Coherency function results for station S002.006_03 ................................................ 71
Figure 4-26: Coherency function results for station S002.055_03 ................................................ 72
Figure 4-27: Coherency function results for station S010.005_03 ................................................ 73
Figure 4-28: Coherency function results for station S019.053_03 ................................................ 74
Figure 4-29: Coherency function results for station S026.041_03 ................................................ 75
Figure 5-1: Hyperparameter optimization of SE-ARD kernel variance ....................................... 82
Figure 5-2: Hyperparameter optimization for SE-ARD length scale, longitude. ......................... 83
Figure 5-3: Hyperparameter optimization for SE-ARD length scale, latitude. ............................ 84
Figure 5-4: Hyperparameter optimization for SE-ARD length scale station separation distance. 85
Figure 5-5: Coherency decay as a function of distance at 'S020.058_03' .................................... 86
Figure 5-6: Hyperparameter optimization for SE-ARD length scale, frequency ........................... 86
Figure 5-7: Hyperparameter optimization for SE-ARD length scale, Vs30. ................................. 87
Figure 5-8: Hyperparameter optimization for SE-ARD kernel length scale, radius to the epicenter. ........................................ 88
Figure 5-9: Hyperparameter optimization for SE-ARD kernel likelihood .................................. 89
Figure 5-10: Model fit of training data (25% of data set) for event 1 4-feature model using a generic kernel ........................................................................... 92
Figure 5-11: Model fit of testing data (75% of data set) for event 1 4-feature model using a generic kernel. ................................................................. 93
Figure 5-12: Model fit of training data (25% of data set) for event 1 4-feature model using an optimized kernel. ................................................................................. 94
Figure 5-13: Model fit of testing data (75% of data set) for event 1 4-feature model using an optimized kernel. ................................................................................. 95
Figure 5-14: Mean-squared error comparison from generic and optimized kernels used for predictions on the training and testing sets for event 1 ................................................................. 96
Figure 5-15: Model fit of training data (25% of data set) for event 1 5-feature model using an optimized kernel. ................................................................................. 98
Figure 5-16: Model fit of testing data (75% of data set) for event 1 5-feature model using an optimized kernel. ................................................................................. 99
Figure 5-17: Model fit of training data (25% of data set) for event 2 5-feature model using an optimized kernel. ................................................................................. 100
Figure 5-18: Model fit of testing data (75% of data set) for event 2 5-feature model using an optimized kernel. ................................................................................. 101
Figure 5-19: Model fit of training data (25% of data set) for event 3 5-feature model using an optimized kernel. ................................................................................. 102
Figure 5-20: Model fit of testing data (75% of data set) for event 3 5-feature model using an optimized kernel. ................................................................................. 103
Figure 5-21: Model fit of training data (25% of data set) for event M6.5 5-feature model using an optimized kernel. ................................................................................. 104
Figure 5-22: Model fit of testing data (75% of data set) for event M6.5 5-feature model using an optimized kernel. ................................................................................. 105
Figure 5-23: Mean-squared error comparison from the five model training sets .............................................. 106
Figure 5-24: Mean-squared error comparison from the five model testing sets .............................................. 106
Figure 5-25: Model fit of testing data (25% of data set) for event 1 6-feature model using an optimized kernel. ................................................................................. 108
Figure 5-26: Model fit of training data (75% of data set) for event 1 6-feature model using an optimized kernel. ................................................................................. 109
Figure 5-27: Model fit of testing data (25% of data set) for event 2 6-feature model using an optimized kernel. ................................................................................. 110
Figure 5-28: Model fit of training data (75% of data set) for event 2 6-feature model using an optimized kernel. ................................................................................. 111
Figure 5-29: Model fit of testing data (25% of data set) for event 3 6-feature model using an optimized kernel. ................................................................................. 112
Figure 5-30: Model fit of training data (75% of data set) for event 3 6-feature model using an optimized kernel. ................................................................................. 113
Figure 5-31: Mean-squared error comparison from the 6-feature model training sets ....................................... 114
Figure 5-32: Mean-squared error comparison from the 6-feature model testing sets ....................................... 115
Figure 5-33: 4-feature model predictions for event 1 at station 'S026.074_03' .............................................. 117
Figure 5-34: 4-feature model predictions for event 1 at station 'S026.006_03' .............................................. 118
Figure 5-35: 4-feature model predictions for event 1 at station 'S002.006_03' .............................................. 118
Figure 5-36: 4-feature model predictions for event 1 at station 'S005.028_03' ........................................ 119
Figure 5-37: 4-feature model predictions for event 1 at station 'S002.055_03' ........................................ 119
Figure 5-38: 4-feature model predictions for event 1 at station 'S010.005_03' ........................................ 120
Figure 5-39: 4-feature model predictions for event 1 at station 'S019.053_03' ........................................ 120
Figure 5-40: 4-feature model predictions for event 1 at station 'S026.041_03' ........................................ 121
Figure 5-41: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 200 meters ................................................................................................................................. 122
Figure 5-42: Difference between 4-feature predictions and true coherency values for event 1 with a station separation distance of 200 meters ........................................................................ 123
Figure 5-43: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 600 meters ................................................................................................................................. 124
Figure 5-44: Difference between 4-feature predictions and true coherency values for event 1 with a station separation distance of 600 meters ........................................................................ 125
Figure 5-45: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 1000 meters ...................................................................................................................... 126
Figure 5-46: Difference between 4-feature predictions and true coherency values for event 1 with a station separation distance of 1000 meters ........................................................................ 127
Figure 5-47: 5-feature model predictions for event 1 at station 'S026.074_03' ........................................ 128
Figure 5-48: 5-feature model predictions for event 1 at station 'S026.006_03' ........................................ 128
Figure 5-49: 5-feature model predictions for event 1 at station 'S002.006_03' ........................................ 129
Figure 5-50: 5-feature model predictions for event 1 at station 'S005.028_03' ........................................ 129
Figure 5-51: 5-feature model predictions for event 1 at station 'S002.055_03' ........................................ 130
Figure 5-52: 5-feature model predictions for event 1 at station 'S010.005_03' ........................................ 130
Figure 5-53: 5-feature model predictions for event 1 at station 'S019.053_03' ........................................ 131
Figure 5-54: 5-feature model predictions for event 1 at station 'S026.041_03' ........................................ 131
Figure 5-55: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 200 meters ................................................................................................................................. 132
Figure 5-56: Difference between 5-feature predictions and true coherency values for event 1 with a station separation distance of 200 meters ........................................................................ 133
Figure 5-57: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 600 meters ................................................................................................................................. 134
Figure 5-58: Difference between 5-feature predictions and true coherency values for event 1 with a station separation distance of 600 meters ........................................................................ 135
Figure 5-59: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 1000 meters ................................................................................................................................. 136
Figure 5-60: Difference between 5-feature predictions and true coherency values for event 1 with a station separation distance of 1000 meters ........................................................................ 137
Figure 5-61: 5-feature model predictions for event 2 at station 'S026.074_03' ........................................ 138
Figure 5-62: 5-feature model predictions for event 2 at station 'S026.006_03' ........................................ 138
Figure 5-63: 5-feature model predictions for event 2 at station 'S002.006_03' ........................................ 139
Figure 5-64: 5-feature model predictions for event 2 at station 'S005.028_03' ........................................ 139
Figure 5-65: 5-feature model predictions for event 2 at station 'S002.055_03' ........................................ 140
Figure 5-66: 5-feature model predictions for event 2 at station 'S010.005_03' ........................................ 140
Figure 5-67: 5-feature model predictions for event 2 at station 'S019.053_03' ........................................ 141
Figure 5-68: 5-feature model predictions for event 2 at station 'S026.041_03' ........................................ 141
Figure 5-69: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 200 meters ........................................................................................................ 142
Figure 5-70: Difference between 5-feature predictions and true coherency values for event 2 with a station separation distance of 200 meters ........................................................................ 143
Figure 5-71: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 600 meters ........................................................................................................ 144
Figure 5-72: Difference between 5-feature predictions and true coherency values for event 2 with a station separation distance of 600 meters ........................................................................ 145
Figure 5-73: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 1000 meters ........................................................................................................ 146
Figure 5-74: Difference between 5-feature predictions and true coherency values for event 2 with a station separation distance of 1000 meters ........................................................................ 147
Figure 5-75: 5-feature model predictions for event 3 at station 'S026.074_03' ........................................ 148
Figure 5-76: 5-feature model predictions for event 3 at station 'S026.006_03' ........................................ 148
Figure 5-77: 5-feature model predictions for event 3 at station 'S002.006_03' ........................................ 149
Figure 5-78: 5-feature model predictions for event 3 at station 'S005.028_03' ........................................ 149
Figure 5-79: 5-feature model predictions for event 3 at station 'S002.055_03' ........................................ 150
Figure 5-80: 5-feature model predictions for event 3 at station 'S010.005_03' ........................................ 150
Figure 5-81: 5-feature model predictions for event 3 at station 'S019.053_03' ........................................ 151
Figure 5-82: 5-feature model predictions for event 3 at station 'S026.041_03' ........................................ 151
Figure 5-83: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 200 meters ........................................................................................................ 152
Figure 5-84: Difference between 5-feature predictions and true coherency values for event 3 with a station separation distance of 200 meters ........................................................................ 153
Figure 5-85: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 600 meters ........................................................................................................ 154
Figure 5-86: Difference between 5-feature predictions and true coherency values for event 3 with a station separation distance of 600 meters ........................................................................ 155
Figure 5-87: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 1000 meters ........................................................................................................ 156
Figure 5-88: Difference between 5-feature predictions and true coherency values for event 3 with a station separation distance of 1000 meters ........................................................................ 157
Figure 5-89: 5-feature model predictions for event M6.5 at station 'S026.074_03' ................................ 158
Figure 5-90: 5-feature model predictions for event M6.5 at station 'S026.006_03' ................................ 158
Figure 5-91: 5-feature model predictions for event M6.5 at station 'S002.006_03' ................................ 159
Figure 5-92: 5-feature model predictions for event M6.5 at station 'S005.028_03' ................................ 159
Figure 5-93: 5-feature model predictions for event M6.5 at station 'S002.055_03' ................................ 160
Figure 5-94: 5-feature model predictions for event M6.5 at station 'S010.005_03' ................................ 160
Figure 5-95: 5-feature model predictions for event M6.5 at station 'S019.053_03' ................................ 161
Figure 5-96: 5-feature model predictions for event M6.5 at station 'S026.041_03' ................................ 161
Figure 5-97: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 200 meters. ............................................................ 162
Figure 5-98: Difference between 5-feature predictions and true coherency values for event M6.5 with a station separation distance of 200 meters. ............................................................ 163
Figure 5-99: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 600 meters. ............................................................ 164
Figure 5-100: Difference between 5-feature predictions and true coherency values for event M6.5 with a station separation distance of 600 meters. ............................................................ 165
Figure 5-101: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 1000 meters. ............................................................ 166
Figure 5-102: Difference between 5-feature predictions and true coherency values for event M6.5 with a station separation distance of 1000 meters. ............................................................ 167
Figure 5-103: 6-feature model predictions for event 1 at station 'S026.074_03'........................................ 168
Figure 5-104: 6-feature model predictions for event 1 at station 'S026.006_03'........................................ 168
Figure 5-105: 6-feature model predictions for event 1 at station 'S002.006_03'........................................ 169
Figure 5-106: 6-feature model predictions for event 1 at station 'S005.028_03'........................................ 169
Figure 5-107: 6-feature model predictions for event 1 at station 'S002.055_03'........................................ 170
Figure 5-108: 6-feature model predictions for event 1 at station 'S010.005_03'........................................ 170
Figure 5-109: 6-feature model predictions for event 1 at station 'S019.053_03'........................................ 171
Figure 5-110: 6-feature model predictions for event 1 at station 'S026.041_03'........................................ 171
Figure 5-111: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 200 meters. ............................................................ 172
Figure 5-112: Difference between 6-feature predictions and true coherency values for event 1 with a station separation distance of 200 meters. ............................................................ 173
Figure 5-113: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 600 meters. ............................................................ 174
Figure 5-114: Difference between 6-feature predictions and true coherency values for event 1 with a station separation distance of 600 meters. ............................................................ 175
Figure 5-115: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 1000 meters. ............................................................ 176
Figure 5-116: Difference between 6-feature predictions and true coherency values for event 1 with a station separation distance of 1000 meters. ............................................................ 177
Figure 5-117: 6-feature model predictions for event 2 at station 'S026.074_03'........................................ 178
Figure 5-118: 6-feature model predictions for event 2 at station 'S026.006_03'........................................ 178
Figure 5-119: 6-feature model predictions for event 2 at station 'S002.006_03'........................................ 179
Figure 5-120: 6-feature model predictions for event 2 at station 'S005.028_03'........................................ 179
Figure 5-121: 6-feature model predictions for event 2 at station 'S002.055_03'........................................ 180
Figure 5-122: 6-feature model predictions for event 2 at station 'S010.005_03'........................................ 180
Figure 5-123: 6-feature model predictions for event 2 at station 'S019.053_03'........................................ 181
Figure 5-124: 6-feature model predictions for event 2 at station 'S026.041_03'........................................ 181
Figure 5-125: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 200 meters. ............................................................ 182
Figure 5-126: Difference between 6-feature predictions and true coherency values for event 2 with a station separation distance of 200 meters. ................................................................. 183
Figure 5-127: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 600 meters. ................................................................. 184
Figure 5-128: Difference between 6-feature predictions and true coherency values for event 2 with a station separation distance of 600 meters. ................................................................. 185
Figure 5-129: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 1000 meters. ................................................................. 186
Figure 5-130: Difference between 6-feature predictions and true coherency values for event 2 with a station separation distance of 1000 meters. ................................................................. 187
Figure 5-131: 6-feature model predictions for event 3 at station ‘S026.074_03’. ................................................................. 188
Figure 5-132: 6-feature model predictions for event 3 at station ‘S026.066_03’. ................................................................. 188
Figure 5-133: 6-feature model predictions for event 3 at station ‘S002.066_03’. ................................................................. 189
Figure 5-134: 6-feature model predictions for event 3 at station ‘S005.028_03’. ................................................................. 189
Figure 5-135: 6-feature model predictions for event 3 at station ‘S002.055_03’. ................................................................. 190
Figure 5-136: 6-feature model predictions for event 3 at station ‘S010.005_03’. ................................................................. 190
Figure 5-137: 6-feature model predictions for event 3 at station ‘S019.053_03’. ................................................................. 191
Figure 5-138: 6-feature model predictions for event 3 at station ‘S026.041_03’. ................................................................. 191
Figure 5-139: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 200 meters. ................................................................. 192
Figure 5-140: Difference between 6-feature predictions and true coherency values for event 3 with a station separation distance of 200 meters. ................................................................. 193
Figure 5-141: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 600 meters. ................................................................. 194
Figure 5-142: Difference between 6-feature predictions and true coherency values for event 3 with a station separation distance of 600 meters. ................................................................. 195
Figure 5-143: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 1000 meters. ................................................................. 196
Figure 5-144: Difference between 6-feature predictions and true coherency values for event 3 with a station separation distance of 1000 meters. ................................................................. 197
1 Introduction
1.1 Problem description

The built environment characterizes cultural importance, advancements in physical systems, and marks of community demand. While civil infrastructure is crucial to communities, many are susceptible to seismic events. Understanding the characteristics behind ground motions and their interaction with the built environment helps improve the way civil infrastructure is designed to avoid severe damages in future earthquake events.

A seismic event is an occurrence of spatially varying ground motions (SVGMs), which come in the form of waves propagating and scattering through the media affected by site, path, and source parameters. Ideally, these effects should be modeled through ground motion coherency that measures the correlation of wave responses between two points. However, in the current practice, no physics-based models are capable of accounting for all-inclusive coherency effects on long-span structures such as bridges, tunnels, and pipelines, which has led to abridged notions for considering these effects (Zerva et al., 2018). One way coherency effects are currently accounted for is by using simplified planar wave theories or empirical models. These methods do not account for all coherency effects as vertical/inclined plane wave assumptions can only account for site and wave passage effects, and empirical models are rough estimates based on limited data from research centers such as the Lotung START-1 array located in Taiwan. As a result, current design approaches may lead to uncertainty of bridge response on the order of 100%-1000% (Zerva, 2018).

Advancements in high-performance computing has increased the capabilities for running ground motion simulations at higher resolutions. Today, deterministic broadband simulations are
practical to the frequency of 5 Hz and are expected to be practical up to 10 Hz within the next decade. These higher resolution simulations provide the opportunity to further study physical properties affecting SVGMs such as station separation distance, frequency, \( V_{S30} \), and radius to the epicenter. It is important to inspect these properties, especially when analyzing long-span structures such as bridges, pipelines, and tunnels, because observed ground motions may change significantly over the footprint of distributed infrastructure. In this regard, this research explores the possibilities of using broadband deterministic ground motion simulations and machine learning to form more accurate physics-based models for coherency and enhance the abilities to model SVGMs to ensure proper analysis of long-span structures.

1.2 Objectives and Scope of Work

To this end, with support from the Southern California Earthquake Center (SCEC) and Lawrence Livermore National Laboratory, physics-based coherency analyses have been conducted for the San Francisco Bay Area to further the understanding of how different site, path, and source effects contribute to the changes in SVGMs observed in deterministic ground motion simulations. According to USGS, the San Francisco Bay Area has a 72% chance of exceeding a magnitude 6.7 earthquake before 2043, making this region a critical area of understanding for SVGM behavior (Aagaard, 2016).

More specifically, the objectives of this thesis are to: (1) generate earthquake simulation-based coherency data sets within a prescribed region of the San Francisco Bay Area at frequencies ranging from 0.00 – 3.50Hz, (2) implement GPR to train and model coherency results within the region of interest using physical parameters as input features, and (3) better understand the
influence physical parameters used as model input features have on coherency values between geospatial points.

To achieve the first objective, four SW4 simulations conducted in the San Francisco Bay Area are used to formalize coherency data sets through a set of MATLAB codes. SW4 is a time-domain fourth-order accurate finite-difference code based on the summation-by-parts principle for accurate and efficient simulation of seismic wave propagation (Petersson et al., 2015). Three of the SW4 simulations investigated are magnitude 4.5 point-source events, and the fourth is a simulated magnitude 6.5 fault rupture event (Pitarka et al., 2021, & Graves et al., 2016). Coherency was developed using a magnitude-squared coherence approach, considering the power spectral density functions at two stations and the cross power spectral density between the stations.

The second objective was tackled by taking results from the coherency function analysis of simulations and assigning input features to each observation, including station longitude, latitude, separation distance, frequency, $V_{S30}$, and radius to the epicenter. With known physical parameters capable of being used as model input features, GPR was used to form predictions across each data set individually. The first magnitude 4.5 source event was modeled using four features and five features to see if there was any model improvement from the inclusion of $V_{S30}$ as a model input parameter. Since there were some improvements in model accuracy after modeling event one on five features, the other three events were modeled with the inclusion of $V_{S30}$ as a fifth input feature. After 5-feature modeling, radius to the epicenter was considered as a sixth input feature for the three point-source events. The training set was comprised of coherency
results from each of the three point-source simulations to measure modeling performance with greater model uncertainty.

After modeling each event using GPR, objective three was reached by analyzing the results from optimized hyperparameters inside of the kernels used to form predictions on the data sets. These hyperparameters provided insight into the relationships formed between input features and expected coherency output, providing pertinent information on how site, source, and path effects influenced the evolution of coherency in the frequency domain.

1.3 Organization of Thesis

The thesis is organized into six chapters and followed by Appendices A, B, C, and D. Following the introduction, Chapter 2 delivers a literature review on the importance of coherency, particularly for long-span structures, modeling techniques, and physical parameters that affect levels of ground motion coherency between two geospatial points. Chapter 3 provides an in-depth look at the mathematics behind GPR with included examples. Chapter 4 details how coherency was generated using an analytical approach. This chapter furthers the understanding of the second chapter’s literature review of coherency, and the results presented in this chapter are used for GPR modeling in Chapter 5. Chapter 5 uncovers modeling approaches with GPR and how it was used to formulate predictions based on the four data sets generated from SW4 simulation results. This chapter implements the theory discussed in Chapter 3 using the coherency data sets formulated in Chapter 4. Chapter 6 synthesizes the work performed, conclusions made from this research, and future work. Appendix A supports Chapter 4 with MATLAB scripts used to generate the coherency data set along with syntax for the understanding of the functions used. Appendix B supports Chapter 5 with Python functions used
to implement GPR and includes documentation for an understanding of Python modules. Appendix C provides more results from coherency analysis that are not presented in Chapter 4, and Appendix D provides more results from GPR that are not presented in Chapter 5.
2 Ground Motion Coherency

2.1 An overview

Ground motions produce spatial variability during seismic events, which causes sporadic variation of ground interaction across a given region. SVGMs are commonly characterized through properties such as coherency, apparent wave velocity, and amplitude variability (Mezouer, 2010). It is crucial to inspect these attributes, especially when analyzing long-span structures such as bridges, pipelines, and tunnels because observed ground motion accelerations will vary due to the site, path, and source characteristics. Depending on a structure’s importance and size, the effects of ground motion coherency on small footprints can be ignored in many cases because it is not expected for seismic waves to change significantly over small regions. But, when considering structure footprints on the order of 100 meters or longer, wave propagation may be observed differently from one end of a structure to another. Spatial variability of seismic ground motions is a complex and approximate science due to the number of variables that influence the behavior of seismic waves. These include but are not limited to frequency, shear wave velocity, station separation distance, epicentral distance, wave passage, and soil composition. These variables dictating the evolution of wave propagation across a region will be discussed in this chapter.

Ground motion coherency is represented as either lagged or unlagged coherency. The lagged coherency accounts for amplitude variability of ground motions and is the more common use for engineering applications (Ghiocel et al., 2004). Lagged coherency values vary between 0 and 1, with higher values suggesting a stronger correlation between the two points. Unlagged coherency accounts for amplitude and wave passage variability. In other words, unlagged coherency accounts for the change in wave amplitude between two points while also accounting for the
time lag between points ‘i’ and ‘j’. Unless points ‘i’ and ‘j’ lie along the same plane of wave propagation, it is expected for the wave to reach points ‘i’ and ‘j’ at different times. While unlagged coherency is discussed in the coherency literature review, lagged coherency is used for analysis because of its commonality in engineering applications.

Including lagged coherency into the design of long-span structures is essential for determining appropriate ground motions across the span of the structure, which can change drastically as it travels through a medium. Designing a long-span structure to experience the same ground motions can be detrimental to its performance under seismic loading, and spatial variability cannot be disregarded (Zerva, 2002). Saxena et al. (2000) discuss the relevancy of coherency and how it can negatively affect the structural performance of a reinforced concrete bridge by performing a nonlinear analysis on a highway bridge in two cases. The first case considered homogeneous soil while the second case considered different soil conditions under each bridge support. The conclusion of this study reported that ductility demands for the second case were double those of the first case, with the dominating factor being coherency effects.

Today, several methods for quantifying coherency have been proposed through the forms of empirical, semi-empirical, and analytical functions. Many of these functions account for distance and frequency, although they are derived from limited data incapable of accurately accounting for site-specific SVGMs. Knowing the physical properties that influence coherency will allow for more accurate modeling of ground motion records along the entire span of a structure.
2.2 Empirical and semi-empirical coherency models

Many approaches have been taken to model ground motion coherency. With advancements in technology, computing power, and more extensive collections of real data, researchers within the field have proposed several models for computing ground motion coherency. This section will examine the different approaches and advancements in modeling coherency functions. Two general approaches have shaped this field of research, empirical and semi-empirical modeling. Empirical modeling solely bases predictions against observed evidence, while semi-empirical modeling bases predictions against observed evidence along with theoretical knowledge. While the database of true earthquake ground motions continues to grow, densely spaced empirical records are insufficient to quantify different site, source, and path effects influencing SVGMs.

Coherency is widely understood in terms of spacing and circular frequency. Equation 2-1 denotes an unlagged coherency function with respect to lagged coherency, distance, and circular frequency (Zerva, 2018).

$$\gamma(\Delta x, \omega) = |\gamma(\Delta x, \omega)| \exp \left[ -i \omega \frac{\Delta x}{V_X} \right]$$  \hspace{1cm} Eq. 2-1

Where $\Delta x$ is the distance between two stations coherency is sought, $\omega$ is the circular frequency, $i$ is the imaginary number, $V_X$ is the apparent propagation velocity of waves along the axis in which coherency is being estimated, and $|\gamma(\Delta x, \omega)|$ is the lagged coherency function. Let’s consider two ground motion time series recorded at stations ‘j’ and ‘k’ with spacing $\Delta x$. Equation 2-2 describes the formulation of the lagged coherency between these two stations by determining the power spectral density of each recorded time series and their cross power spectral density.

$$|\gamma(\Delta x, \omega)| = \gamma_{jk}(\omega) = \left| \frac{S_{jk}(\omega)}{\sqrt{S_{jj}(\omega)S_{kk}(\omega)}} \right|$$  \hspace{1cm} Eq. 2-2
In this formulation, \( S_{jk}(\omega) \) is the smoothed cross power spectral density between points ‘j’ and ‘k’, while \( S_{jj}(\omega) \) and \( S_{kk}(\omega) \) are the power spectral densities at points ‘j’ and ‘k’, respectively. \( S_{jj}(\omega) \) and \( S_{kk}(\omega) \) are considered point estimates of motion at stations ‘j’ and ‘k’, which are used to normalize \( S_{jk}(\omega) \) (Zerva, 2002).

Eq. 2-1 is under the assumption of planar wave propagation across homogenous and isotropic soil (Zerva, 2018). While several types of seismic waves exist, shear waves are assumed to transfer almost all energy during a seismic event. Because of this, an apparent shear wave velocity, \( V_x \), is used and all other waves are ignored (Zerva, 2018). It is important to note that the unlagged coherency function above considers wave passage variability through the exponential term, while the lagged coherency function itself can take the form of any empirical or semi-empirical lagged coherency function. Relevant empirical and semi-empirical lagged coherency functions follow.

Many empirical lagged coherency functions have been formulated through data collected at the SMART-1 and Lotung LSST arrays in Taiwan. The SMART-1 array was developed through a joint effort between the University of California at Berkley and the Institute of Earth Sciences in 1980 as the first large-scale digital array of strong-motion seismographs (Abrahamson et al., 1987). Its development was to aid researchers in the field of seismic hazard analysis with strong motion data that would further research being performed. In 1980, 37 stations were placed within three rings with radii of 200 meters, 1000 meters, and 2000 meters, with one of the 37 stations located in the centroid of all three rings, as seen in Figure 2-1. Four additional stations were later placed at distances of 3000, 4000, 5000, and 6000 meters from the central station,
respectively. The SMART-1 array was operational from 1980 to 1991 and recorded 60 seismic events with varying magnitudes, focal depths, and epicenters (“The SMART-1”).

The 1985 Lotung LSST array was constructed between the middle and outer rings of the SMART-1 array seen in Figure 2-2, and similarly consisted of 37 stations (“The LLSST”). This project was funded through the Taiwan Power Company and U.S. Electric Power Research Institute to gain insight into ground motion variability with respect to depth in soil and structure elevation (“The LLSST”). While this site was much smaller than the SMART-1 array, its purpose was to study different ground motion interactions, particularly within soil-structure interaction. Two scaled-down reinforced concrete containments, scaling at 1/4 and 1/12 of typical construction size, were placed above the Lotung LSST array with stations located within each structure, as seen in Figure 2-3. 30 earthquakes were recorded between 1985 and 1991 during the operation of the Lotung LSST. These two arrays of stations served as the backbone for early empirical ground motion coherency modeling because of the consistent data they provided.
Figure 2-1: Configuration of the SMART-1 Array, Taiwan (from “The SMART-1 Array”).

Figure 2-2: Lotung LSST placement with respect to the SMART-1 array (from “The SMART-1 Array”).
Abrahamson et al. (1991a) introduced one of the empirically derived lagged coherency functions for strong ground motions by analyzing 15 of the 60 strong ground motions recorded from the Lotung LSST arrays located in Taiwan. They used data from the 37 original Lotung LSST stations to generate an empirical coherency function for stations less than 100 meters of separation. The initial parametric model proposed by Abrahamson for short separation distances can be seen in Equation 2-3.

\[
\text{tanh}^{-1}\gamma(f, \xi) = (a_1 + a_2\xi) \left[\exp\{(b_1 + b_2\xi)f\} + \frac{1}{3}f^c\right] + k
\]

Eq. 2-3

In this empirical coherency model derived from the Lotung LSST array, \(a_1, a_2, b_1, b_2,\) and \(c\) are parameters estimated by nonlinear regression, \(k\) is the noise level, \(f\) is the frequency of the ground motion in Hertz, and \(\xi\) represents the separation distance between stations in units of meters. Abrahamson used the proposed model above to analyze the change of coherency as a function of station separation distance and frequency. To solve for \(a_1, a_2, b_1, b_2,\) and \(c,\)
coherency values from the 15 selected ground motions were assigned to six different bins in accordance with their station separation distance. Mean coherency and variance were calculated for each event at frequencies 1 to 50 Hz and are seen in Figure 2-4 provided by Abrahamson below. Lagged coherency values are presented in the inverse hyperbolic tangent form. These bins categorize separation distance in the range of 0-10, 10-20, 20-30, 30-50, 50-70, and 70-100 meters, respectively.

Figure 2-4: Mean arctanh(Coherency) calculated from the Lotung LSST Array (from Abrahamson, 1990)

An interesting discovery found within the relationship between distance, frequency, and coherency is that the coherency is generally dictated by the change in frequency when the station separation distance is small, as seen above. As station separation distance increases towards 100 meters, distance proves to have a more significant contribution towards decreased values of lagged coherency. The value of coherency can also be observed converging as frequency increases (Abrahamson, 1991a). From the behavior seen with coherency converging, the unknown $k$ parameter is taken as the median value seen at a frequency of 50Hz. The other values were solved through nonlinear regression and are displayed below in Equation 2-4. This
empirical model provides a 90% confidence interval for coherency values greater than 0.56 at the Lotung LSST site in Taipei, Taiwan (Abrahamson, 1991a).

\[ \tanh^{-1}|\gamma(f, \xi)| = (2.54 - 0.012\xi) \left[ \exp\{-0.0115 - 0.00084\xi\}f + \frac{1}{3}f^{-0.878} \right] + 0.35 \quad \text{Eq. 2-4} \]

Not only did Abrahamson et al. derive an empirical coherency model for short station separation distances using data collected from the LLSST array, but they also took advantage of real data collected from the SMART-1 array to develop horizontal and vertical coherency models for station separation distances 100-1000m. Since this research intends to look at horizontal coherency, the vertical coherency function will not be discussed in detail. Like the coherency function developed on the Lotung LLST array, the following equation is represented purely in terms of frequency and separation distance, as seen in Equation 2-5 (Zerva, 2009).

\[ |\gamma^H(\xi, f)| = \tanh\left\{ \frac{c_1^H(\xi)}{1+c_2^H(\xi)f + c_4^H(\xi)f^2} + (4.80 - c_1^H(\xi)\exp[c_3^H(\xi)f] + 0.35 \right\} \quad \text{Eq. 2-5} \]

Where the functions \( c_1^H, c_2^H, c_3^H, c_4^H \) are given in terms of separation distance, \( \Delta x = \xi \).

\[ c_1^H(\xi) = \frac{3.95}{1+0.0077\xi+0.000023\xi^2} + 0.85\exp[-0.00013\xi] \quad \text{Eq. 2-6} \]

\[ c_2^H(\xi) = \frac{0.4\left[1-(1+\left(\frac{\xi}{180}\right)^3\right]^{-1}}{\left[1+(\xi/190)\right]\left[1+(\xi/180)\right]} \quad \text{Eq. 2-7} \]

\[ c_3^H(\xi) = 3(\exp[-0.05\xi] - 1) - 0.0018\xi \quad \text{Eq. 2-8} \]

\[ c_4^H(\xi) = -0.598 + 0.106\ln(\xi + 325) - 0.0151\exp[-0.6\xi] \quad \text{Eq. 2-9} \]

This model proposed by Abrahamson was developed for distances up to 1000 meters and will be used for comparison against coherency results computed in this research. An important aspect of this empirical coherency function is that the 0.35 value added at the end of the function represents a convergence value seen within the empirical data, like the \( k \) parameter in the empirical function for short separation distances.
Questions have been raised about the legitimacy of empirical coherency functions because of the influence source, path, and site effects have on the spatial variability of ground motions. Coherency models such as the ones proposed by Abrahamson have been developed from a minimal number of true spatial arrays that exist today. These spatial arrays, such as the SMART-1 array in Taiwan, do not have enough variability in site, path, and source characteristics to represent significantly different geographical regions. The lack of dense arrays worldwide and therefore limited accuracy of the resulting empirical models is the point of interest for using simulation-based ground motions to study coherency.

Semi-empirical coherency models were beginning to develop in the late 1970s and 1980s with notable researchers such as B. J. Uscinski, J. E. Luco, and H. L. Wong. Through his work in wave propagation theory, B. J. Uscinski pioneered coherency functions for ground motion wave propagation. In 1977, Uscinski published “The Elements of Wave Propagation in Random Media”, which proposed methods for understanding uncertainty in wave dispersion. One notable equation from his work is the semi-empirical coherency function in Equation 2-10 (Uscinski, 1977).

\[
\langle E_1 E_2^* \rangle = \exp \left\{ -t \left[ 1 - \exp \left( \frac{-\left( \xi^2 + \eta^2 \right)}{r_0^2} \right) \right] \right\} \tag{Eq. 2-10}
\]

Where \( \xi \) is equivalent to \( (x_1 - x_2) \) and \( \eta \) is equivalent to \( (y_1 - y_2) \), which allows for \( (\xi^2 + \eta^2) \) to be represented as the squared distance between two points in space, i.e., \( \Delta x^2 \). This form will be referred to as \( |x - x'|^2 \) for the duration of this discussion as it is commonly referred to in later literature. Within the exponential, \( r_0 \) refers to the length scale of random irregularities along a path. This exponential is set equal to \( \langle E_1 E_2^* \rangle \) which in this case refers to the spatial autocorrelation function of an electric field. Spatial autocorrelation refers to the similarity of two
points with respect to distance and variation of some physical property. So, while Uscinski refers to this in terms of an electric field, the practicality of spatial autocorrelation applies to many areas of science including strong ground motions which is detailed further through the work of Luco and Wong, 1986. Luco and Wong apply this general equation to ground motion lagged coherency modeling as seen in Equation 2-11 by approximating the scaled medium distance of wave propagation, $l$, in Equation 2-11 and assuming that $|x - x'| < r_o$. Assuming that the separation distance between two points is less than the scale length of random path irregularities $r_o$, the two points evaluated are both experiencing some form of randomness.

$$f(|x - x'|, \omega) = \exp \left[ - \left( \frac{\gamma \omega |x-x'|}{\beta} \right)^2 \right]$$

Eq. 2-11

$\beta$ in this formulation represents an apparent shear wave velocity, $|x - x'|$ represents the distance between two points formerly notated as $\sqrt{\xi^2 + \eta^2}$, and $\omega$ represents the circular frequency. It is apparent that $f(|x - x'|, \omega)$ is equivalent to the general spatial autocorrelation function, $\langle E_1 E_2^* \rangle$.

When applying this idea to strong ground motions, coherency is a varying physical property with respect to the distance between two points and frequency. The formulation of $\gamma$, as seen in Equation 2-12, is a factor governed by the standard deviation of elastic properties, wave propagation distance $H$, and the length scale of random inhomogeneities, $r_o$ (Luco et al., 1986).

$$\gamma = \mu \sqrt{\frac{H}{r_o}}$$

Eq. 2-12

Approximation of $l$:

$$l \approx \frac{\omega^2 r_o H \mu^2}{\beta^2}$$

Eq. 2-13

The function proposed by Luco et al. (1986) is still a popular coherency estimation and is used for generating spatially varying ground motion synthetics (Zentner, 2012).
2.3 Source, path, and site effects on ground motion coherency

Ground motion coherency has proven to be difficult to model because of the challenges involved with modeling SVGMs traveling through heterogeneous soils. In this section, the influence of source, path, and site conditions on ground motion coherency will be discussed.

Source:

Source characteristics play an essential role in observed ground motion coherency values. Some of the influencing factors described through research conducted by Abrahamson et al. (1991b) and AfifChaouch et al. (2016) include epicentral distance and earthquake magnitude. These factors and their correlation to ground motion coherency are described below.

AfifChaouch et al. (2016) studied earthquake magnitude's effects on ground motion coherency. To understand this source effect, two earthquake events were used for analysis. The first earthquake event used for this study was the 1980 El-Asnam earthquake which observed a magnitude 7.3 mainshock. The second earthquake used to measure the effects on coherency due to earthquake magnitude was the 1995 Kozani-Greeneva earthquake which featured a magnitude 6.5 mainshock. The resulting study found that decay in coherency began at lower frequencies for the larger magnitude earthquake, El-Asnam, compared to the smaller magnitude earthquake, Kozani-Greeneva (AfifChaouch et al., 2016). The results show that magnitude of rupture is a defining attribute to ground motion coherency and indicates that an increasing rupture area leads to decreasing values of coherency (AfifChaouch et al., 2016). This effect is seen because wave propagation from different fault locations moving in different directions changes the ground
motions observed at two stations, resulting in lower coherency. This is later observed from ground motion simulations used for coherency analysis in the San Francisco Bay Area.

Epicentral distance, or the distance from a station to the surface point above an earthquake's hypocenter, is also concluded to play a significant role in ground motion coherency values. Abrahamson et al. (1991b) studied these effects from two earthquake recordings from the Lotung LSST array to compare near-field and far-field ground motion coherency. Interestingly, Abrahamson concluded that low values of coherency at low frequencies observed in the near-field (5km) suggested a dominant source effect. It was also concluded that low coherency values at high frequencies in the far-field (66km) suggested a dominant path effect (Abrahamson et al., 1991b). Epicentral distance in far-field events was found to have negligible effects at further distances. AfifChaouch et al. (2016) concluded similar results from their study. The team explored ground motion coherency from epicentral distances of 5km (near-field), 33km (intermediate-field), and 65km (far-field) on the 1980 El-Asnam mainshock. They found that coherency loss was negligible in the far-field (65km) for frequencies up to 8Hz, which backs the results found by Abrahamson et al. (1991b) of having path effects, but no source effects in the far-field. AfifChaouch et al. (2016), similarly to Abrahamson et al. (1991b), found significant coherency loss in the near-field, indicating the epicentral distance is an essential characteristic in observed ground motion coherency values.

**Path:**
Path effects on ground motion coherency regard the direction and length of travel of strong ground motion waves. Wave passage, scattering effects, and length of travel will be discussed to
understand how path effects influence strong ground motion coherency. Wave passage refers to the path seismic waves take through a specific site. The path of waves can significantly influence the coherency between two points because of time delays which can also be described as the time difference between a seismic wave observed at two points (Ancheta et al., 2011). Unlagged coherency values account for wave passage variability, such as the example presented in Figure 2-5. Scattering is observed when seismic waves pass through heterogeneous soil, influencing the path that seismic waves take when traveling through a medium.

![Wave passage and scattering effects](image)

Figure 2-5: Wave passage and scattering effects of strong ground motions (*from* Zerva, 2009).

Laib et al. (2015) studied the effects of soil heterogeneities on coherency by modeling two columns of an assumed multi-support structure on top of heterogeneous soils. These two columns simulated common civil infrastructure such as a bridge with a large separation between supports. Laib et al. (2015) introduced stochastic randomness to the soil properties to model heterogeneous soils below the two columns, ensuring each column would interact differently under the same ground motions. The evolution of coherency between the two columns was analyzed for separation distances between 50 and 650 meters. This study shows that heterogeneities within a soil cause irregularity in seismic motions and wave passage between two points, leading to exponential decay in ground motion coherency. This was especially true near
resonant frequencies of the soil as it led to amplification of ground motions (Laib et al, 2015). While soil heterogeneities rely on site characteristics, they are important for understanding the effect of wave scattering and the energy transfer through a medium.

As discussed earlier, Abrahamson et al. (1991b) and AfifChaouch et al. (2016) find that wave path does factor into coherency decay at high frequency, far-field events. Although low frequency, far-field events have little to no effect on coherency because there is less energy transfer to the soil it is traveling through, both studies concluded that when the epicentral distance is close to the site, it can be considered a source effect due to the size of the rupture. However, when the epicentral distance is in the far field, it can be considered a path effect because of the wave’s length of travel.

Site:
Different regions of interest are shown to act differently during seismic events, often due to site conditions. Site conditions pertain to characteristics including soil structure and topography. The soil structure of a particular site plays an essential role in nonlinearity and liquefaction, amplification, wave attenuation, and dispersion. All of which influence the coherency between two points of a free field. Soil structures at sites can vary from large alluvium deposits to surface bedrock which plays a crucial role in the values of ground motion coherency observed. Bi et al. (2010) studied these effects by generating ground motions and modeling soil interaction on a layered canyon site. This study took a set of normally distributed soil properties accounting for shear moduli, density, and damping ratios to develop a four-layer canyon, as seen in Figure 2-6 (Bi et al., 2010). The soil properties defined in the figure are taken as true values of the site for
modeling coherency within this region of interest. A coefficient of variation (COV) was then added to each soil property to see how coherency would change with respect to the values described below. COVs chosen for each soil property were based on practical discrepancies seen in the field.

Monte-Carlo simulations were used to understand the effects ground motions had on the variability of soil composition. Interestingly, random damping ratios and soil densities proved to be insignificant to ground motion coherency, $\gamma_{jk}$ and $\gamma_{j'k'}$ (Bi et al., 2010). Exploring the effects of the damping ratio containing a COV of 40%, the coherency model had little change. Similarly, adding a COV of 5% to the soil density had little effect on the deterministic soil properties where COV is 0. The addition of a 40% COV value on the shear modulus did have an influence on ground motion coherency, especially when coherency was above 3Hz. Not only did it affect the ground motion coherency, but COV levels also increased the variance seen for calculated coherencies.
This study also shed light on the influence of site topography and how ground motion coherency changed with respect to this physical characteristic. Inspecting the influence of topography, the deterministic, or true values described in Figure 2-6 were used (Bi et al. 2010). In other words, the coefficient of variation for random soil properties was set to zero. For comparison, Bi et al. compared how coherency changed as a function of frequency from points $j'$ to $k'$ and points $j$ to $k$ to measure the influence of topography. It was found that the coherency values on the ground surface were smaller than those on base rock (Bi et al. 2010). Further, coherency values between ground surface stations experienced peaks and troughs as frequency increased, highlighting the importance of analyzing multiple modes for design purposes. Low coherency values at the ground surface are correlated to high site amplification ratios, proving a concentration of ground surface motions at a few frequencies (Bi et al. 2010). Below, some of the results concluded from this study are included.

Figure 2-7: Comparison of the mean lagged coherency between the surface motions $(j,k)$ with that of the incident motion on the base rock: (a) horizontal out-of-plane motion; (b) horizontal in-plane motion; and (c) vertical in-plane motion (from Bi et al., 2010).
Figure 2-8: Modulus of the site amplification spectral ratio of two local sites: (a) horizontal out-of-plane motion; (b) horizontal in-plane motion; and (c) vertical in-plane motion. (from Bi et al., 2010).

Figure 2-7 and Figure 2-8 provided by Bi et al. (2010) show the correlation between coherency and spectra ratio between surface ground motions. The spectra ratio is taken as the ratio of a frequency response function between two points on the ground surface, i.e., \( \frac{|H_k(i\omega)|}{|H_j(i\omega)|} \) (Bi et al., 2010). These figures show low coherency between surface ground motions is directly correlated to higher spectra ratios. The larger the spectra ratio, the less correlation there is between observed ground motions at points ‘j’ and ‘k’.

Schneider and Abrahamson also conducted studies regarding site conditions and their effects on ground motion coherency for stations with a separation distance of 100 meters or less (Schneider et al., 1992). In particular, the two researchers were curious about how dependent site characteristics were on spatial variation of ground motions. To observe this site dependency, coherency values collected from ten different locations were compared to Abrahamson’s empirical lagged coherency function seen in Equation 2-3. The sites chosen for this study consisted of five soft-soil and five rock site classes to see if any discrepancies were found in
coherency values due to ground composition. These sites are detailed below in Table 2-1 and the ground motions used for characterizing the effects of soil composition are detailed in Table 2-2.

Table 2-1: Site categorization for rock and soil arrays (from Schneider et al., 1992).

<table>
<thead>
<tr>
<th>Array</th>
<th>Location</th>
<th>Site Class</th>
<th>Surface Class</th>
<th>Spacing (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPRI LSST</td>
<td>Taiwan</td>
<td>Soil</td>
<td>15</td>
<td>3-85</td>
</tr>
<tr>
<td>EPRI Parkfield</td>
<td>CA</td>
<td>Rock</td>
<td>13</td>
<td>10-191</td>
</tr>
<tr>
<td>Chiba</td>
<td>Japan</td>
<td>Soil</td>
<td>15</td>
<td>5-319</td>
</tr>
<tr>
<td>USGS Parkfield</td>
<td>CA</td>
<td>Rock</td>
<td>14</td>
<td>25-952</td>
</tr>
<tr>
<td>Imperial Valley Diff</td>
<td>CA</td>
<td>Soil</td>
<td>5</td>
<td>18-213</td>
</tr>
<tr>
<td>Hollister Diff</td>
<td>CA</td>
<td>Soil</td>
<td>4</td>
<td>61-256</td>
</tr>
<tr>
<td>Stanford (temp)</td>
<td>CA</td>
<td>Soil</td>
<td>4</td>
<td>32-185</td>
</tr>
<tr>
<td>Coalinga (temp)</td>
<td>CA</td>
<td>Rock</td>
<td>7</td>
<td>48-313</td>
</tr>
<tr>
<td>UCSC ZAYA (temp)</td>
<td>CA</td>
<td>Rock</td>
<td>6</td>
<td>25-300</td>
</tr>
<tr>
<td>Pinyon Flat (temp)</td>
<td>CA</td>
<td>Rock</td>
<td>58</td>
<td>7-340</td>
</tr>
</tbody>
</table>

To investigate the effects of soil composition on lagged coherency, Schneider et al. (1992) took a similar approach to Abrahamson (1991) by categorizing soft-soil and rock sites separately for comparison. Since the Lotung LSST array showed that magnitude did not have a large impact on coherency values between stations located less than 100 meters away from each other, it was assumed that coherency values would not see a significant impact from earthquake magnitude. Using a power spectral density approach, mean lagged coherency values were computed for separation distances 15-30 meters and 50-80 meters. The results were compared to those using Abrahamson's empirical equation for short separation distances. Their studies found that coherency values were lower at rock sites when compared to soft-soil sites. This study also
showed that Abrahamson's empirical equation overestimated coherency values for station separation distances less than 100 meters at rock sites (Schneider et al., 1992). This result is most likely due to drastically different soil compositions located at the Lotung LSST array, making it a poor equation for coherency analysis at stiffer soil sites. When comparing mean lagged coherency values from soft-soil sites, results show a good fit to the Abrahamson empirical function at short separation distances. This study highlights the importance of site characteristics and how current empirical functions may fail to capture important parameters that affect its decay through the frequency domain.

In another study, AfifChaouch et al. also studied the effects of shear wave velocity on observed coherency at a site by comparing shear wave velocities on a generic rock site (AfifChaouch et al., 2016). The inspected shear wave velocities were 1500 m/s, 2500 m/s, and 3500 m/s, respectively. This study found that shear wave velocity has a significant effect on ground motion coherency at frequencies greater than 2 Hz (AfifChaouch et al., 2016). Their study also concluded that coherency decayed faster at lower shear wave velocities. Higher velocities correspond to faster arrival times between stations, which strengthens the results concluding faster decay times as shear wave velocities decrease.
3 Gaussian Process Regression Modeling

Defining relationships between spatial and parametric data can be quite challenging within the field of statistics, especially when observing multivariate data. It is critical to take an approach where a suitable integration method is used for understanding the cross relationship of data. Gaussian process regression (GPR) serves as a tool for spatial modeling which incorporates a prior distribution based on training data to form predictions across unseen data points. The following chapter details a conceptual understanding of this form of regression.

GPR is a stochastic method of interpretation, defining the relationship between random variables through multivariate normal distribution (Schulz, 2018). Multivariate normal distribution describes the relation between ‘n’ random variables. It does so by expressing a Gaussian distribution onto higher dimensions. Suppose a Gaussian process is expressed as a normally distributed function of ‘n’ random variables. It can be described that an arbitrary multivariate Gaussian process is determined by the possible linear combinations of these ‘n’ random variables (Wang, 2021). To understand this further, an investigation into the differences between one-dimensional and multi-dimensional distributions will be discussed.

In a classic one-dimensional Gaussian distribution, the mean $\mu$, is found within the set of all real numbers and has a positive-definite variance $\sigma^2$. The distribution is described through a scalar random variable $x$ with the following exponentiated quadratic probability density function seen in Equation 3-1.

$$
\mathcal{N}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\exp[-\frac{1}{2}(x - \mu)^2 / \sigma^2]} \\
$$

Eq. 3-1
This probability density function follows a classic bell-curve shape being centered around the mean. The term in front of the exponentiated quadratic acts as a normalization constant which ensures integration of the functions is equal to one. Inside the exponential, it is seen that the random variable and mean are represented in quadratic form and divided by the variance $\sigma^2$.

Multivariate Gaussian distribution finds its roots within the described one-dimensional distribution. What differs is the number of random variables. Upon exceedance of one random variable, the probability density function takes the form of Equation 3-2 (Wang, 2021).

$$
\mathcal{N}(\mathbf{x} \mid \mathbf{\mu}, \Sigma) = \frac{1}{2^{d/2}|\Sigma|^{-1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{\mu})^T \Sigma^{-1} (\mathbf{x} - \mathbf{\mu}) \right\}
$$

Eq. 3-2

In this representation, $\mathbf{x}$ is a vector of random variables with length $d$, $\mathbf{\mu}$ is a vector of means with length $d$, and $\Sigma$ is the corresponding covariance matrix. The multivariate probability density function also contains a normalization constant to ensure the integration is equal to one. A multivariate Gaussian distribution can similarly be represented in form of Equation 3-3 which shows the relationships between different points and how they are represented through mean and covariance (Wang, 2021).

$$
\begin{bmatrix}
\mathbf{x}_1 \\
\mathbf{x}_2 \\
\vdots \\
\mathbf{x}_n
\end{bmatrix}
\sim
\mathcal{N}
\left(
\begin{bmatrix}
\mathbf{\mu}_1 \\
\mathbf{\mu}_2 \\
\vdots \\
\mathbf{\mu}_n
\end{bmatrix},
\begin{bmatrix}
\sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\
\sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2 \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2
\end{bmatrix}
\right)
\sim
\mathcal{N}(\mathbf{\mu}, \Sigma)
$$

Eq. 3-3

Every value in the mean vector, $\mathbf{\mu}$ is associated with a random variable within the $\mathbf{x}$ vector, and every index of the covariance matrix refers to the correlation between two random variables. As for the measure of spread, instead of using a simple variance, a covariance matrix is used which models the joint variability of all random variables. The values which make up this positive and semi-definite covariance matrix can be represented in the notation of Equation 3-4 (Gut, 2009).

$$
\Sigma_{i,j} = cov[\mathbf{x}_i, \mathbf{x}_j] = E[(\mathbf{x}_i - E[\mathbf{x}_i])(\mathbf{x}_j - E[\mathbf{x}_j])]
$$

Eq. 3-4
In the above notation, \( E \) notates the expected mean value, \( x_i \) represents the \( i^{th} \) random variable, and \( x_j \) represents the \( j^{th} \) random variable. Invoking the Kronecker delta on the covariance matrix returns a vector of variances for each random variable. These values represent the diagonal of the covariance matrix where taking the square root of each value produces the standard deviation.

GPR refers to this joint distribution of variables. This process creates a prior distribution from all possible functions across a continuous domain. In the case of this research, the multivariate data being referred to is across a geospatial domain found within the San Francisco Bay Area. GPR takes account of all potential combinations of random variables and normally distributes them about a mean (Wang, 2021). Before training, there are an infinite number of functions that can define the model because there are no formal observations. As training data is introduced, the model learns what functions may possibly represent the data. Every point seen updates the model and narrows down possible model representations by forming a prior of functions (Wang, 2021). This prior is taken to be the posterior distribution as more data is introduced during training. Every new data point seen during training of the Gaussian model updates the prior for later predictions. While the model is observing data points, it begins to understand how outputs of the unknown function are distributed.

To model the prior, GPR takes advantage of what is known as the kernel trick. A Gaussian kernel can be used to develop the covariance matrix, \( \Sigma \), and mathematically model correlations between random variables. Following is a discussion of the squared exponential kernel, also known as the radial-basis function (RBF) kernel used for coherency predictions within the scope of this research.
The RBF kernel considers the Euclidian distance or the squared distance between two points. It serves as a commonly used kernel in many Gaussian process applications because of its ability to be integrated over many functions and only requires three parameters including the distance between two observations $\|x - x'\|$, length scale $l$, and variance (Duvenaud, 2014). The following formulation seen in Equation 3-5 is for a one-dimensional case and its extension into higher dimensionality follows in Equation 3-6.

$$K(x, x') = \sigma^2 \exp\left[\frac{\|x - x'\|^2}{2l^2}\right]$$  
Eq. 3-5

The length scale $l$ measures the relationship between $\|x - x'\|$ and its output while the variance $\sigma^2$ is a scaling factor for determining the expected values from the model’s mean function. When $x = x'$, the exponential term evaluates to 1, the RBF function is equal to the kernel variance. This variance evaluates how likely the mean function is to capture the expected value. As seen in Figure 3-1, the kernel variance at $x = x'$ provides insight into how many standard deviations ($\sigma$) away from the mean function the expected output $y$ is likely to fall within.

![Figure 3-1: Illustrative example of mean function distribution across objective function (from Ding, 2018).](image)
Extending the use case of the RBF kernel into higher dimensionality accounts for each dimension individually and multiplies them together to account for multiple input features with unique length scales as seen in Equation 3-6. This formulation of the RBF (squared exponential) kernel follows identically to that seen in Equation 3-5, but in this case, accounting for the relevance of each input parameter. SE-ARD refers to the squared exponential automatic relevance determination as it weighs the relationship of each input feature with its corresponding output (Duvenaud, 2014). A large length scale value implies smaller change across dimensions, while a smaller length scale implies a stronger correlation between the input feature and its corresponding output. Figure 3-2 demonstrates how taking the product of higher dimensional function features builds the kernel used for predicting across that space (Duvenaud, 2014).

\[
\text{SE-ARD}(\mathbf{x}, \mathbf{x}') = \prod_{d=1}^{D} \sigma_d^2 \exp\left(-\frac{1}{2} \frac{(x_d-x'_d)^2}{l_d^2}\right) = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{d=1}^{D} \frac{(x_d-x'_d)^2}{l_d^2}\right) \quad \text{Eq. 3-6}
\]

Figure 3-2: A Gaussian prior formed from the product of two one-dimensional input features using SE-ARD kernel (from Duvenaud, 2014).

The kernel trick proves to be beneficial to forming predictions through GPR, although is reliant on the tuning of hyperparameters within the kernel function used. Tuning kernel hyperparameters is essential to GPR because they govern the generalizations made in the model (Wang, 2021). The optimized parameters are extremely important to the accuracy of the model, and improper tuning can lead to misleading results. A sound method of doing so is through maximization of the log marginal likelihood as seen in Equation 3-7 (Wang, 2018).
\[
\log p(y \mid X) = -\frac{1}{2}y^T(K + \sigma_n^2 I)^{-1}y - \frac{1}{2} \log \det(K + \sigma_n^2 I) - \frac{n}{2} \log 2\pi \quad \text{Eq. 3-7}
\]

where \( y \) is a vector of outputs, \( X \) is a matrix of input variables, \( K \) is the covariance matrix, and \( \sigma_n^2 I \) is the noise. Hyperparameter tuning follows Equation 3-8 (Wang, 2018).

\[
\theta^* = \arg\max_\theta \log p(y \mid X, \theta) \quad \text{Eq. 3-8}
\]

where \( \theta^* \) are the tuned hyperparameters on the probability of outputs, given inputs, and untuned hyperparameters. This process continues until log marginal likelihood converges to a maximum.

Training a Gaussian model relies on hyperparameter tuning to form accurate predictions on unseen data sets. The predictions formed on a trained Gaussian model can be represented as Equation 3-9 (Wang, 2018).

\[
f^* \mid X, y, X_*, \theta^* \sim \mathcal{N} \left( \bar{f}^*_\star, \text{cov}(\bar{f}^*_\star) \right) \quad \text{Eq. 3-9}
\]

where \( \bar{f}^*_\star \) is the mean function used to form predictions given the training set \((X, y)\), input variables for the test set \(X_*\), and optimized kernel function hyperparameters \( \theta^* \).

When implementing Gaussian kernels into analysis, it is beneficial to normalize the output about a zero mean with a standard deviation equal to one (Wang, 2018). Normalization allows for data to be bound along the same scale and influence regression at an equal rate. This allows for more accurate observations as the area under the distribution curve is equal to one and each deviation away from the mean is one. After forming predictions on normalized data, the normalization is reversed to report prediction values of the output.

To understand how GPR forms predictions, two examples of its implementation using the kernel trick follows. The first example highlights the GPR use-case for a one-dimensional problem, and the second example extends into the multi-dimensional use case of GPR with the implementation
of the SE-ARD kernel. Traditional GPR models follow a 70-30 split of the entire data set between training and testing sets. This percentage is maintained for each of the following examples.

The first example that will be inspected is: \( f(x) = \sin(x) \). This example will provide an understanding of how the one-dimensional case is implemented as seen in Equation 3-5. To make the example more realistic, noise is added to the data set which can be seen in Figure 3-3. The blue line represents the true \( \sin(x) \) graph while the red points show the noisy data used for training. For this example, 50 noisy data points were generated between 0 and 10, which can be seen along the trace of the noisy data function. These data points form a prior distribution of functions or the possible representations of the data.

![Figure 3-3: 1-dimensional noisy data set used for training.](image)
Following the training and tuning of hyperparameters, 20 random points were used as test observations to assess how well the model learned the noisy data. Below shows the predicted outcome of test points along with the corresponding standard deviation in Figure 3-4. The model captures the true sin(x) function after training on the noisy input, serving as a nice primer to the capabilities of GPR.

![Optimized Hyperparameters for sin(x) Function](image)

**Figure 3-4:** Optimized kernel predictions from noisy sin(x) function.

To measure the model’s accuracy to the true sin(x) function, predictions are fit against corresponding true values and the mean-squared error is calculated. Figure 3-5 shows a good fit to the true sin(x) function which provides evidence for the accuracy of the GPR model.
After optimizing the kernel used for predictions, the x-input RBF length scale was found to be equal to 1.87, and a variance equal to 1.1. This one-dimensional example provides a good opportunity to address the influence of the length scale. Setting the variance equal to its optimized value and changing the length scale used for modeling provides a good understanding of the length scale's influence on predictions of one-dimensional problems. The following results in Figure 3-6 and Figure 3-7 show how predictions change with respect to the length scale being set to 5.0. As discussed earlier, if the length scale is too large for modeling, the mean function used for predicting becomes smoother and underfits the model trend.
Figure 3-6: Kernel predictions from noisy sin(x) function with length scale set to 5.0.

Figure 3-7: Model fit from kernel predictions with length scale set to 5.0.

When compared to results from the optimized length scale seen in Figure 3-4 and Figure 3-5, setting the length scale to 5.0 shows underfitting against the true function. While this length scale can capture the general trend of the sin(x) function, the model fit is poor in comparison to the
optimized kernel length scale. Similarly, the MSE from these comparisons shows a lack of correlation. Compared to the MSE calculated from optimized kernel predictions (0.00328), the MSE calculated from the length scale set to 5.0 (0.15113) is dramatically larger.

Understanding the effect of having a length scale larger than its optimized value leads to the discussion of the contrasting situation, having a length scale smaller than its optimized value. In contrast to the previous underfitting example, if the length scale is smaller than its optimized value, overfitting against the training set occurs as seen in Figure 3-8.

Figure 3-8: Kernel predictions from noisy sin(x) function with length scale set to 0.1.

Figure 3-8 shows the GP mean ‘jitter’ about the training data. This is a clear sign of overfitting which affects the accuracy of predictions compared to the true sin(x) function. Overfitting occurs
when the Gaussian mean function tries too hard to fit the observed training data. This is further seen in Figure 3-9 which shows a poor fit to the true sin(x) values.

As in Figure 3-7, kernel predictions show a weak correlation to the true sin(x) function. When compared to the MSE calculated with the optimized kernel length scale (0.00328), the MSE calculated from the length scale set to 0.1 (0.09177) is much larger and provides insight into how poorly the model performs.

Following the discussion of how the RBF kernel is used in a one-dimensional case, the function $f(x,y) = y^2 - x^2$ is evaluated to better understand how this kernel is used in higher dimensionality. Like the first example, data was generated for x and y between 0 and 10, and a mesh grid was formulated from these values. 30 values for x and 30 values for y, forming a mesh grid of 900 spatial data points comprised of x, y, and output z. This can be seen in Figure 3-10.
Figure 3-10: 2-dimensional noisy data set used for training.

240 random points from the developed mesh grid were pulled to create a testing set, and the other 560 data points were used for training the Gaussian model. As introduced earlier, the dependency of each input feature is evaluated separately, and the product is taken to describe the interaction between all inputs and the corresponding output. Results for modeling can be seen in Figure 3-11 and Figure 3-12.

Figure 3-11: Predictions from training on \( f(x,y) = y^2 - x^2 \) with addition of noise.
As seen in Figure 3-11, the GP model could form relationships between the output and input features, even though a noisy data set was introduced for training. Toggling the noise to be larger or smaller than what has been presented is expected to change the levels of model accuracy. Introducing higher amounts of data to the example is expected to penalize the relationships between input features and output. On the other hand, setting noise levels to lower values is expected to strengthen the relationships between input and output features.

To see how well this model performed, a plot of predictions against true values is investigated. MSE is calculated between predictions and true values of f(x,y). It can be seen in Figure 3-12 that a tight fit is formed between the model predictions and true values of f(x,y). A mean squared error of ~0.8 is calculated, which is low for the magnitude of values seen in this example. The value of MSE is dependent on the scale of values. In the sin(x) example, MSE was much lower than this but can be contributed to the fact that output values are bounded by 1 and minus 1. In this example, the output z is bounded by the values 100 and minus 100, resulting in a greater value of MSE.

![Figure 3-12: Predictions versus true values of f(x,y) = y^2 - x^2](image-url)
4 Simulation-Based Coherency Function Analysis

The San Francisco Bay Area is a highly seismic region of the United States. This region is susceptible to the San Andreas, Hayward, and Paicines faults, all of which contribute to a 72% probability of exceeding a magnitude 6.7 earthquake or greater before 2043 (Aagaard, 2016). The high likelihood of strong ground motions in this heavily populated region pushes the importance of studying wave propagation in the region. Simulated ground motions allow researchers to study probable scenarios that may occur in the future.

For the scope of this research, ground motions have been developed using deterministic methods. Deterministic ground motion fields are computed on individual earthquake simulations with a specified magnitude and location. An individual earthquake simulation can be idealized as a singular point-source or multiple point sources to simulate a fault rupture. For the generation of simulated ground motions in the San Francisco Bay Area, the seismic simulation code SW4 is used. SW4 formulates seismic wave equations through a displacement-based fourth-order finite difference method in the Cartesian coordinate system, making it suitable for simulating wave propagation on a regional scale (Petersson, 2017).

The following chapter will discuss the formulation of coherency data sets representative of the San Francisco Bay Area. A 9km by 24km region located close to the Hayward fault is investigated and detailed in Figure 4-1. A mesh grid was used to generate 2,400 master stations for coherency analysis. At each master station, coherency is evaluated for frequencies ranging from 0.00-3.50Hz with a frequency step size of 0.01Hz, and station separation distances ranging from 100-1000m at a distance step size of 100m. For the purposes of this research, four
simulated deterministic ground motions have been used to generate coherency within the region of interest. These four simulations will be mentioned as Ev1, Ev2, Ev3, and EvM6.5. Ev1, Ev2, and Ev3 are deterministic ground motion simulations of point sources, while EvM6.5 is a simulated fault rupture (Pitarka et al., 2021, & Graves et al., 2016). Ev1 will be highlighted as blue, Ev2 will be highlighted as red, Ev3 will be highlighted as green, and EvM6.5 will be highlighted as cyan. See Figure 4-1 for details on simulation locations. The three point-source simulations are on the order of a 4.5 magnitude earthquake, while the simulated fault rupture is on the order of a 6.5 magnitude earthquake. The 2,400 master nodes, 351 frequencies, and 10 station separation distances generate a total of 8,424,000 values of coherency across the region of interest for each SW4 simulation. The magnitude of data generated from these simulations will prove important for analyzing coherency results and later modeling discussed in Chapter 5.

Section 4.2 builds on the power spectral density model discussed in Chapter 2 by describing the process of generating a coherency dataset using MATLAB. Section 4.3 reviews the results generated from MATLAB and compares them to empirical coherency models for large separation distances, which was also introduced in Chapter 2.
4.1 Coherency Function Analysis

This section will describe the process of generating coherency data within the region of interest. MATLAB was utilized for all coherency calculations, and scripts that were used for evaluation are discussed throughout this section and can be referenced in Appendix A. To compute coherency, the equation below, formerly discussed in Chapter 2 was used.

\[ \gamma_{jk}(\omega) = \frac{S_{jk}(\omega)}{\sqrt{S_{jj}(\omega)S_{kk}(\omega)}} \]  

Eq. 4-1

Coherency analysis is performed using several MATLAB scripts, with the source code titled ‘MainCoherency.m’. Several functions are called from the source code to perform operations on the velocity files recorded from SW4 simulations. A station ID file containing a list of each station’s SW4 identification, longitude, and latitude is read into MATLAB. The corresponding longitude and latitude pairs are used as input arguments to convert decimal coordinates to x-y
coordinates. The Cartesian coordinate system helps track station separation distances for coherency analysis. A ‘for loop’ is initialized within the ‘MainCoherency.m’ source code to run through the 2400 master node stations and record coherency values corresponding to station separation distance and frequency. For each station, delta-x and delta-y vectors are assembled between station ‘i’ and all other stations in the data set. The delta-x vector corresponds to the Cartesian values of longitude, and the delta-y vector corresponds to the Cartesian values of latitude. Each station corresponding to one of the ten station separation distances used for coherency analysis is found and indexed for coherency analysis. A 10-meter margin of error is allowed for indexing stations at distances of interest. So, a station 990m-1010m away from station ‘i’ would be indexed as 1000 meters away. Similarly, a station 90m-110m away from station ‘i’ would be indexed as 100 meters away.

Once station separation distances have been determined, SAC files are called out by generating a file base name corresponding to station ‘i’. The function ‘plotsac.m’ was utilized in the ‘MainCoherency’ source code, pulling velocities in the x, and y directions at station ‘i’. Then, the Tukey window was applied to the velocity histories to ensure the start and end of the recordings taper to zero. A cosine fraction of 0.05 was input into the Tukey window, which tapered 2.5% of the data at the beginning and end of the recordings. Then, between the master node and identified stations, coherency is evaluated in the x and y directions using the mscohere function and the resulting coherencies were averaged for each master station. Figure 4-2 is a visual representation of how this averaging is performed in MATLAB. These averaged coherency values calculated for each master station at respective frequency steps and separation
distances are used for Gaussian model training, detailed further in Chapter 5. An example of what an expected output file should contain can be seen in Table 4-1.

Figure 4-2: Coherency averaging at master node 'i' with respect to separation distance.
Table 4-1: MATLAB Coherency Results

<table>
<thead>
<tr>
<th>Coherency</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.99671</td>
</tr>
<tr>
<td>0.01</td>
<td>0.99672</td>
</tr>
<tr>
<td>0.02</td>
<td>0.99674</td>
</tr>
<tr>
<td>0.03</td>
<td>0.99679</td>
</tr>
<tr>
<td>0.04</td>
<td>0.99688</td>
</tr>
<tr>
<td>0.05</td>
<td>0.99699</td>
</tr>
<tr>
<td>0.06</td>
<td>0.99711</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>3.44</td>
<td>0.73709</td>
</tr>
<tr>
<td>3.45</td>
<td>0.73529</td>
</tr>
<tr>
<td>3.46</td>
<td>0.73354</td>
</tr>
<tr>
<td>3.47</td>
<td>0.73178</td>
</tr>
<tr>
<td>3.48</td>
<td>0.72987</td>
</tr>
<tr>
<td>3.49</td>
<td>0.7276</td>
</tr>
<tr>
<td>3.5</td>
<td>0.72475</td>
</tr>
</tbody>
</table>

Results are directed to appropriate folders and dropped into subfolders titled the same naming convention as the respective station node. Files are titled corresponding to the separation distance describing coherency between stations. Figure 4-3 details the naming convention of the subfolders organized by respective station ID and the files located in each folder, respectively. In total, the region of interest consists of 2400 master nodes, each containing coherency results evaluated from 0.00-3.50Hz at station separation distances 100m-1000m.
4.2 Coherency Function Results

The following section will provide results from the coherency analysis of the SW4 ground motion simulations discussed above. To better understand coherency analysis results and some of the differences between events, the average shear wave velocity to 30 meters depth is provided in Figure 4-4. This velocity profile map gives a general understanding of the soil composition within the ROI. The San Francisco Bay Area is known for having large amounts of soft-soil deposits, which is apparent when looking at the $V_{s30}$ map of the region. The average shear wave velocity across all 2400 stations within the ROI is 316m/s, with minimum values reaching 176.1m/s. According to ASCE 7-16 Table 20.3-1, if shear wave velocities are below 600ft/s (~183m/s), the location is categorized as Site Class E. Low shear wave velocities coincide with softer soils and amplitude amplification effects, which can lead to damaging structures. Within the ROI, some harder soils are present, particularly close to the fault, which can influence wave passage and energy dissipation. It should be noted that in all SW4 simulations the minimum shear wave velocity was set to 500m/s. However, while SW4
simulations define a minimum shear wave velocity of 500 m/s, a stochastic velocity model has been introduced into simulations to capture small-scale heterogeneities present in the domain space. Looking at the position of each SW4 simulation event with respect to the region of interest in Figure 4-1, wave scattering is expected to differ between simulations because the path taken through the ROI is not the same. Figures 4-5 through 4-8 show the snapshots of wave propagation within the region of interest in terms of the amplitude of the particle velocities, which help achieve better insight into how path and site effects may influence coherency through the region of interest. For plotting purposes, half-second intervals of each event are taken. Each simulation took place over a 25-second interval, but snapshots were reduced to the length of ground motion excitation. It is also important to note that the scale between the point source simulations and fault rupture are different because surface velocity values are much higher for the 6.5 magnitude rupture.

Event 1 is in the southeast corner of the region of interest, and ground motions begin around the 9-second interval of the simulation, as seen in Figure 4-5. The source wave travels towards the southwest corner of the region and scatters upwards through the rest of the region. Scattering effects can be seen shortly after the 9.0-second timestamp as velocity contours can be observed throughout the entirety of the region. This is mainly due to soil heterogeneities in the region. As discussed in Chapter 2, soil heterogeneities change the direction of traveling waves which is a significant reason why the site effects can affect coherency values. The 12.5-second timestamp clearly shows how scattering influences the direction of wave travel as the region of interest is filled with dispersed energy from the initial rupture. Trapping effects are also observed due to the soft soils present within the region of interest. Trapping occurs when seismic waves propagate
from stiff to soft soil deposits. As ground motions enter soft soil deposits, velocity is dramatically decreased, leading to the stacking of seismic waves that struggle to release energy and amplify effects. Timestamps 10.5-12.5 seconds show velocity increasing as seismic waves pass through in the northern section of the region of interest. Notably, the 12.5-second timestamp highlights high ground motion velocities in the northeast corner.

While event 2 is of the same magnitude as event 1, the change in location will influence the velocity history differently. Location of the event is considered a source effect and changes ground motion velocities observed due to the new respective path. Compared to event 1, event 2 has a more defined wave propagation as it passes from the northeast to the southwest corner of the region of interest as seen in Figure 4-6. Like event 1, scattering effects are observed shortly after the initial rupture. Trapping is also observed in timestamps 10.0, 10.5, and 11.0 seconds. These timestamps further amplify ground motion velocities in the northern part of the region of interest following the initial seismic wave.

Events 1 and 3 share more commonalities with each other than they do with event 2. Like event 1, the initial wave propagation of event 3 shows lower surface velocities than those of event 2. Event 3 follows a sweeping pattern across the region of interest from east to west which can be observed through timestamps 9.0-11.0 seconds in Figure 4-7. The 10.5 and 11.0-second timestamps are interesting because they show similar wave trapping effects seen in events 1 and 2. Event 1 highlighted ground motion amplification in the northeast corner of the region of interest, which is also observed in the 11-second timestamp of event 3. Event 2 also revealed site
amplification effects in the northern part of the region of interest, and similar amplification effects can be seen in the 10.5 and 11.0-second timestamps in event 3.

The simulated magnitude 6.5 fault rupture duration was longer than events 1, 2, and 3, which is expected because its order of magnitude is much greater. Shear wave velocities were also much more significant in this event, reaching values nearly 100 times those of the point source events. While velocities were much higher in this event, site and path effects were like those of the point source events. Amplitude amplification of ground motion waves can be seen in many of the timestamps but are notably like those seen in the northeast corners of events 1 and 3.
Figure 4-4: Average shear wave velocities for the top 30 meters of soil within the region of interest.
Figure 4-5: Velocity history of Event 1 across the region of interest.
Figure 4-6: Velocity history of Event 2 across the region of interest.
Figure 4-7: Velocity history of Event 3 across the region of interest.
Figure 4-8: Velocity history of the magnitude 6.5 event across the region of interest.
Coherency from each simulation was inspected across the region of interest by plotting its decay at individual frequencies with respect to the station separation distance. For purposes of discussion in this chapter, the change in coherency with respect to frequency is plotted with corresponding station separation distances of 200, 600, and 1000 meters for each ground motion simulation. Coherency analysis for station separation distances of 400m and 800m can be found in Appendix D. The decay of coherency as frequency increases supports empirical and semi-empirical equations discussed in Chapter 2, including Abrahamson 1993, and Luco and Wong 1986. The findings from coherency analysis also support source, site, and path effects contributing a significant role in ground motion coherency observed between stations.

A noticeable trend that will be seen throughout the following figures is that coherency remains closer to 1 when station separation distances are small. While coherency loss is expected at short separation distances, its decay is much slower than its long-distance counterpart. Coherency analysis of the ROI begins at 0.0Hz and continues to 3.5Hz. As discussed in Chapter 2, the increase in frequency influences the decay of coherency. Observed coherency decay for event 1 at a station separation distance of 200 meters presents itself around 0.75Hz as seen in Figure 4-9. Coherency decay begins further away from the event and gets increasingly closer to the source as frequency increases. Note the pocket of larger coherency values surrounding the source. This is an interesting characteristic of event 1, which also presents itself at larger separation distances. This pocket becomes less noticeable but is retained as the frequency approaches 3.5Hz.

Coherency decay is much more rapid for event 1 at a station separation distance of 600 meters when compared to a station separation distance of 200 meters. This is likely due to scattering and
amplitude amplification effects seen within the region of interest. While coherency trends towards zero faster at a 600-meter station separation distance, its decay follows a similar trend as the 200-meter analysis. Coherency loss begins around 0.75Hz and begins near the boundaries of the ROI. A pocket of greater coherency values also develops around the point source event, which is retained as frequency increases to 3.5Hz.

For a station separation distance of 1000 meters, event 1 continues to follow similar trends in coherency decay as compared to the 200 meter and 600-meter separation distances, seen in Figure 4-11. The main difference between the two is how quickly coherency decays within the ROI. Although coherency is lower than those from shorter station separation distances, loss of coherence is seen encroaching towards the source as frequency increases, and a pocket of high coherency values is retained around the point source. As frequency increases, greater amplitude amplification can be observed, causing the energy dissipation within the medium to become scattered and random. With larger station separation distances, it is expected for surface displacements, velocities, and accelerations to become increasingly disconnected.

While the behavior of event 2 shows some similarities to event 1, there are a few differences that should be pointed out. Coherency values of event 2 at a station separation distance of 200 meters highlight how path effects influence coherency. Event 1 shows strong levels of incoherence in the northern section of the ROI, while event 2 shows strong levels of incoherence throughout the entirety of the ROI, especially at larger frequencies. Looking back at Figure 4-4 depicting $V_{s30}$ values across the ROI, along with Figure 4-6, which provides information on the path of event 2, it is observed that the second event travels through a more complex soil structure located in the
northeast section of the ROI. This can be causing further scattering effects within the ROI and stacking of seismic waves, explaining why surface velocities are higher, and coherency values are lower across event 2. Like event 1, a pocket of coherency is observed near the rupture although unlike event 1, this pocket disappears as the frequency approaches 3.5Hz.

Event 2 at a station separation distance of 600 meters follows similarly to the same event at 200 meters but shows a more rapid decay of coherency. As frequency approaches 1.75Hz for event 2 at a station separation distance of 600 meters, levels of coherency begin converging to minimum values, with little change from 1.75-3.5Hz as seen in Figure 4-13.

Following the 200 and 600-meter station separation distances for event 2, the 1000-meter station separation distance furthers the explanation of scattered energy dominating the ROI at higher frequencies. Coherency decays quicker than shorter station separation distances and converges to minimum values as the frequency approaches 1.25Hz which is graphically represented in Figure 4-14. While it appears that coherency loss begins in the southern part of the ROI, it quickly engulfs the entire region at around 0.75Hz.

Event 3, located between events 1 and 2, shows results like those seen from event 1. Coherency begins trending towards 0 around the outside of the ROI and a pocket of higher coherency values forms around the point source as seen in Figure 4-15. Ground motion coherency values also seem to stay higher along the rupture path for event 3 seen in Figure 4-7.
At a station separation distance of 600 meters, event 3 shows a unique characteristic that is not easily seen in other coherency analysis results. A fanning pattern is observed through the region of interest as central stations within the region remain more coherent with each other when compared to stations further away from the source as seen in Figure 4-16. This effect is likely due to the path event 3 takes and shows how scattering effects are more likely to influence ground motions further from the source. The fanning pattern is apparent around the frequency 1.0Hz and becomes more defined as the frequency approaches 3.5Hz.

Event 3 coherency results for a station separation distance of 1000 meters follow similarly to those seen for event 3 with a station separation distance of 600 meters. The fanning pattern is also apparent beginning at a frequency of 1.0Hz and becoming well-defined as the frequency reaches 3.5Hz, as detailed in the results shown in Figure 4-17. The reasoning for this is likely connected to the source location being centralized at the eastern side of the region, making this fanning pattern easier to see. Looking at event 1, this fanning pattern from the source can faintly be seen as the loss of coherency becomes greater in the northern section of the region.
Figure 4-9: Coherency results from event one with station separation distance equal to 200 meters.

Figure 4-10: Coherency results from event one with station separation distance equal to 600 meters.
Figure 4-11: Coherency results from event one with station separation distance equal to 1000 meters.

Figure 4-12: Coherency results from event two with station separation distance equal to 200 meters.
Figure 4-13: Coherency results from event two with station separation distance equal to 600 meters.

Figure 4-14: Coherency results from event two with station separation distance equal to 1000 meters.
Figure 4-15: Coherency results from event three with station separation distance equal to 200 meters.

Figure 4-16: Coherency results from event three with station separation distance equal to 600 meters.
As discussed earlier, the magnitude 6.5 event is different from the other three events because it simulates an entire fault rupture, depicted by the cyan line in Figure 4-18. Since the behavior of this event is not like the others, it is expected for coherency results also to have major differences. These differences are noticeable when first looking at coherency results for this event with a station separation distance of 200 meters. Coherency is observed to decay across the entire region of interest and displays a striping pattern that is not seen in the point source events. This striping pattern shows greater variation in coherency across the region of interest, which is an effect of the simulated fault rupture. This rupture simulation considers several discrete point sources to simulate striking along a planar fault. Considering several ground motions entering the ROI at once increases the scattering effects observed within the region of interest, resulting in the variation and striping pattern observed at station separation distances of 200 meters.
Coherency results for the magnitude 6.5 earthquake event at 600 meters, shown in Figure 4-19, are like those seen at a station separation distance of 200 meters. The striping pattern observed for 200-meter results is also apparent and well defined in results for a station separation distance of 600 meters. The influence scattering has on coherency results is clearly seen for the magnitude 6.5 event and is extremely important to consider for the soft soils seen in the San Francisco Bay Area. A station separation distance of 1000 meters for the simulated fault rupture is different from results seen at 200 and 600 meters. The striping pattern, seen at 200 and 600 meters, is no longer clear, rather rapid coherency loss overcomes the entirety of the ROI which is shown in Figure 4-20. Coherency values observed in this simulation are the lowest values seen from all events and respective distances. The results make sense since ground motions are entering the region of interest at several locations and propagating in different directions leading to high levels of scattering.

Figure 4-18: Coherency results from event M6.5 with station separation distance equal to 200 meters.
Figure 4-19: Coherency results from event M6.5 with station separation distance equal to 600 meters.

Figure 4-20: Coherency results from event M6.5 with station separation distance equal to 1000 meters.
Figure 4-21 summarizes the changes in coherency by looking at the regional average coherency value at each frequency. These results are compared to those from Abrahamson’s 1993 empirical coherency equation discussed in Chapter 2. Interestingly, these ROI averages follow closely with empirical values developed by Abrahamson 1993. Figure 4-21 is only used to show the relationship between the four events and the trending pattern of coherency decay with respect to frequency.

Plotting coherency functions at specific stations provides more understanding of how site, path, and source effects influence coherency decay through the region of interest. Coherency functions were developed for Ev1, Ev2, Ev3, EvM6.5, and compared to the 1993 empirical equation developed by Abrahamson. While this equation is used for comparison to the results captured from ground motion simulations, one discrepancy should be addressed. The empirical coherency function is derived from ground motions occurring up to distances of 80km away from the SMART-1 array. This differs from coherency data sets used in this research because the ground motion simulations are much closer in proximity when compared to empirical ground motions used for derivation of Abrahamson’s equation. While this equation does give a general sense of the shortcomings that arise from using a general empirical equation, it is likely that this equation
is not suitable for near-field coherency analysis. While some near-field events were used to generate this empirical equation (~5km), the inclusion of far-field events likely weakens the accuracy of the equation for modeling coherency in the near-field.

As mentioned, and seen earlier, Ev1 is represented in blue, Ev2 is represented in red, Ev3 is represented in green, EvM6.5 is represented in cyan, and the 1993 Abrahamson empirical equation is represented in orange. Eight stations were chosen to be presented as a part of Chapter 4 results. Coherency results are shown for each event at station separation distances of 200, 400, 600, 800, and 1000 meters. Four of the stations presented were used for GP training, and four of the stations were not used for GP training. These eight stations will be looked at again in Chapter 5 to see how well modeling was able to capture changes in ground coherency. Appendix C includes additional stations where coherency was evaluated with respect to frequency. Table 4-2 details the distances of each event to the highlighted stations in this chapter.

<table>
<thead>
<tr>
<th>Station</th>
<th>Distance From Event 1 (km)</th>
<th>Distance From Event 2 (km)</th>
<th>Distance From Event 3 (km)</th>
<th>Distance From Event M6.5 (km)</th>
<th>Used for Model Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>S026.074_03</td>
<td>10.38</td>
<td>20.96</td>
<td>15.64</td>
<td>15.45</td>
<td>TRUE</td>
</tr>
<tr>
<td>S026.006_03</td>
<td>17.02</td>
<td>9.11</td>
<td>11.66</td>
<td>10.24</td>
<td>TRUE</td>
</tr>
<tr>
<td>S005.028_03</td>
<td>8.41</td>
<td>5.67</td>
<td>2.72</td>
<td>1.23</td>
<td>TRUE</td>
</tr>
<tr>
<td>S002.006_03</td>
<td>14.82</td>
<td>2.42</td>
<td>7.85</td>
<td>7.07</td>
<td>TRUE</td>
</tr>
<tr>
<td>S002.055_03</td>
<td>1.38</td>
<td>13.27</td>
<td>7.30</td>
<td>7.74</td>
<td>FALSE</td>
</tr>
<tr>
<td>S010.005_03</td>
<td>15.50</td>
<td>4.60</td>
<td>8.92</td>
<td>7.81</td>
<td>FALSE</td>
</tr>
<tr>
<td>S019.053_03</td>
<td>6.50</td>
<td>14.34</td>
<td>9.38</td>
<td>8.96</td>
<td>FALSE</td>
</tr>
<tr>
<td>S026.041_03</td>
<td>9.53</td>
<td>12.69</td>
<td>9.30</td>
<td>8.27</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
Figure 4-22: Coherency function results for station S026.074_03.

The first station highlighted in this chapter, S026.074_03, is in the southwest corner of the region of interest, as shown in Figure 4-22. It is positioned closest to event 1, although it can be regarded as far away from all four events with respect to other stations. The first thing to note about this station is the difference between coherency values of event 1 compared to events 2, 3, and M6.5. Coherency trends towards zero much slower for event 1 when compared to all other events. This makes sense as event 1 ROI contour plots seen in Figures 4-9 through 4-11 show slower decay of coherency in the southern portion of the region of interest.
The second station considered for coherency analysis, ‘S026.006_03’, is in the northwest corner of the region of interest, as shown in Figure 4-23. Coherency appears to decay similarly between all four events at this simulation station which coincides with results collected from ROI contour plots discussed above. Event 3 appears to retain higher values of coherency than all other events at station S026.006_03, which is likely due to ground motion path effects.
Figure 4-24: Coherency function results for station S005.028_03.

Station ‘S005.028_03’ is in the near-field and falls 2.72 km away from Event 3 and 1.23 km away from the midpoint of the simulated 6.5 magnitude fault rupture event. Coherency values seen at this station from Event 3 remain much larger than those of the other events which can be seen in Figure 4-24. Referring to Figure 4-7, the pocket of coherent ground motions developed around the source is seen at this station under event 3 ground motions. While this station also lies close to the midpoint of event M6.5, greater coherency loss is observed compared to event 3. The reasoning for this, as mentioned earlier, is because the nature of event M6.5 is entirely different from events 1, 2, and 3. EvM6.5 simulates multiple point sources along the given plane to recreate the behavior during a fault slip. Since ground motions are entering the region at several different locations, greater scattering effects are observed, leading to lower coherency values.
The fourth station presented in this chapter is ‘S002.006_03’, located 2.42km away from event 2. Unlike the previous near-field station, ‘S005.028_03’, coherency values close to event 2 show extreme variability at all station separation distances, as seen in Figure 4-25. The reasoning for this is likely due to amplitude amplification and trapping effects seen near this station. Looking back at the surface wave velocity profile for event 2, which is seen in Figure 4-6, lingering surface wave velocities can be seen in the northeast region of the ROI near this station between timestamps 9.0 and 10.0 seconds. This leaves the possibility of amplitude amplification occurring at this location and greater variation in the surrounding ground motions when compared to other stations during this event.
Figure 4-26: Coherency function results for station S002.055_03.

Similar to ‘S005.028_03’, seen in Figure 4-24, coherency function results for ‘S002.055_03’ reveal coherent motion in the near-field, as seen in Figure 4-26. Located 1.38 km away from event 1, station ‘S002.055_3’ retains high values of coherency as frequency increases. Looking back at ROI contour plots, the pocket developed around event 1 shows a minor loss of coherency at this station. As mentioned earlier, this pocket develops because path and site effects are not influencing the ground motions as much as other locations during event 1. Events 2, 3, and M6.5 display similar loss of coherency, although event 2 shows more fluctuation at this location. This station clearly highlights the limitations of using the Abrahamson 1993 empirical coherency equation as there is high variability in levels of coherency that are dependent on site, path, and source effects.
Station ‘S010.005_03’, seen in Figure 4-27, shows similar characteristics between events. Unlike several of the events looked at so far, loss of coherency at this station trends towards 0 around the same pace. Differences in fluctuation are observed from each event, although they all show similar decay with respect to frequency. Unlike station ‘S002.055_03’, where the events were drastically different from the Abrahamson 1993 empirical coherency equation, there is a better fit of this location with respect to the simulated events. Particularly, event 2 follows the empirical equation accurately at all station separation distances.
Figure 4-28: Coherency function results for station S019.053_03.

Station ‘S019.053_03’ shows interesting similarities and differences between the four events. Events 1 and 3 retain higher levels of coherency at this station, except between 1.0 and 2.0Hz, where low troughs can be observed. Events 2 and M6.5 experience greater coherency loss as frequency increases and converge to similar values at 3.5Hz across all station separation distances as seen in Figure 4-28. The observations from events 2 and M6.5 provide evidence of greater influence from scattering effects, as seen in the previous ROI contour plots discussed.
The final station looked at for coherency analysis is station ‘S026.041_03’. The point source events appear to decay similarly, although this is not the case when looking at the simulated fault rupture event, as seen in Figure 4-29. The dramatic loss of coherency seen from EvM6.5 at this station provides further evidence of greater scattering and amplitude amplification effects throughout the region of interest from the fault rupture when compared to the point source events. Particularly, stations further away from the events have been shown to be more susceptible to wave passage effects.
5 Gaussian Process Regression Model Development for Simulation-based Coherency Functions

As formerly discussed, this research aims to better the understanding of physical parameters that influence strong ground motions and work towards a physics-based coherency model to improve the accuracy of generating spatially varying ground motions. The following chapter provides results from performing GPR on the coherency dataset developed from SW4 simulations. Section 5.1 will give an overview of software packages used to perform GPR and best practices for using these packages in future projects. Section 5.2 will dive into forming appropriate data sets and Gaussian training using 4, 5, and 6-feature models. This section will also discuss the differences in model accuracy that depends on the number of input features, event location, and training data. Section 5.3 provides results and discussion on 4-feature, 5-feature, and 6-feature model predictions. Results provided in section 5.3 coincide with the stations introduced in Chapter 4.

5.1 Software Packages

The models developed to perform GPR on coherency simulations are written using Python. For further information regarding the modules, refer to Appendix B, where informative Python functions are available. The Anaconda environment has been used to run the Jupyter Notebook, an open-source software capable of interpreting Python programming. There are many advantages to using Jupyter, including its interactive user interface, organizational tools, and presentation capabilities. Jupyter worked nicely with results obtained from MATLAB because it provided access to several Gaussian Process modules and an easy interface for interpreting and graphically representing coherency analysis results.
Two packages were potential candidates for performing GPR on the coherency dataset generated from SW4 simulations, Scikit-learn and GPflow (Pedregosa et al., 2011, & Matthews et al., 2017). Initial modeling was performed with Scikit-learn, which features many regression algorithms, including GPR functions for modeling, training, and forming predictions across datasets. This module successfully modeled on smaller datasets but struggled with the size and complexity of the data presented in this thesis. Notably, the performance of Scikit-learn was troublesome because it is unable to allocate processes to a GPU but instead performs all operations on a CPU. Since training a model requires memory on the order of $O(n^3)$ to train a model, CPUs struggle since they cannot perform several computations at once. On the other hand, GPUs can take advantage of parallel computing. Since Gaussian Process Regression requires several matrix computations, finding a Python module capable of allocating computations to a GPU was critical. Unlike Scikit-learn, GPflow could take advantage of a machine’s GPU to perform matrix calculations, allowing models to train faster than traditional CPU operations.

GPflow is a powerful Python module that builds on top of preexisting Tensorflow operations and provides users with a plethora of Gaussian algorithms. Tensorflow is another module that is known for its deep-learning capabilities. To use the Gaussian process algorithms from the GPflow module, it is required to have Tensorflow installed and imported. Great success was found in modeling with GPflow algorithms and has been used for all predictions reported in this chapter. Additional regression results can be found in Appendix D.
As discussed in Chapter 3, the SE-ARD kernel, the RBF kernel for multi-dimensional functions is used for regression analysis in GPflow and displayed again in Equation 5-1.

\[
\text{SE-ARD}(x, x') = \prod_{d=1}^{D} \sigma_d^2 \exp \left( -\frac{1}{2} \frac{(x_d - x'_d)^2}{l_d^2} \right) = \sigma^2 \exp \left( -\frac{1}{2} \sum_{d=1}^{D} \frac{(x_d - x'_d)^2}{l_d^2} \right) \quad \text{Eq. 5-1}
\]

where \( \sigma^2 \) is our variance, \((x_d - x'_d)^2\) is the Euclidean distance between two inputs, and \( l \) is the specified input feature’s length scale. Input features for coherency datasets include longitude, latitude, distance, frequency, \( V_{s30} \), and radius to the epicenter. These will be discussed more in Section 5.2.

GPflow takes advantage of SciPy's L-BFGS-B minimizing method for optimizing kernel hyperparameters. This optimizer searches for a Cauchy point, or the minimum along the steepest gradient descent from initialization. The optimizer then minimizes the quadratic form of the gradient and updates hyperparameters within the kernel. This continues until the optimizer reaches either a local or global minimum. Locating a local minimum would lead to less accurate predictions when compared to the global minimum. To ensure the optimizer used for regression is not reaching a local minimum, having multiple runs of a model with different initializations is recommended. After tuning hyperparameters using the SciPy minimizer, predictions are formed along the entirety of the data set. GPflow populates a covariance matrix to determine each prediction's respective variance and standard deviation. All plotting of predictions has been performed using the Python module Matplotlib. Matplotlib offers a wide variety of plotting capabilities and is a favorite in the Python community. Predictions from each trained model follow in Section 5.3, but first, an overlook of model performance and hyperparameter tuning follows in Section 5.2.
5.2 Model Development

5.2.1 Data Set, Model Setup, and Input Requirements

One of the challenges of this research was understanding what makes an appropriate data set for GP regression. Having four data sets containing nearly 8.5 million unique observations, some parsing would need to be considered due to computing power limitations. While a couple of different approaches were investigated towards the beginning of this research, the following explanation discusses how these data sets were decidedly restricted.

First, coherency from SW4 simulations was recorded for frequency values 0.00-3.50Hz with a step size of 0.01Hz. To condense the data sets, the frequency step size was increased to 0.25Hz, resulting in each data set being minimized to 360,000 observations. To further condense the data sets, only even values of station separation distances were considered, i.e., 200m, 400m, 600m, 800m, and 1000m which left 180,000 unique observations in each data set. While these steps condensed the original data sets to just ~2% of their original size, it was still a lot of data for model training. If a typical Gaussian regression approach were taken for this research, approximately 70% of a data set would be used as model input. Even with the condensed data set, this would require a covariance matrix with size [126,000x126,000]. This number of computations was still too large due to computational limitations. It was decided that data from every fourth simulated station would be used as model input to combat memory limits. In other words, 25% of each condensed data set, or 45,000 unique observations, would be used for training Gaussian regression models. This controlled approach allowed each model to see data from 600 of the 2400 stations within the ROI. While this made the training input coarser, it gave the model a good understanding of coherency loss across individual stations. A similar approach was taken prior to this, where the model was exposed to 45,000 unique observations across all
2400 stations. However, it struggled with tuning hyperparameters, which led to poor predictions at unseen observations. Forcing each model to be exposed to several observations at specific stations rather than sparse observations at every station improved model accuracy and hyperparameter tuning.

Another question for Gaussian regression was which input requirements should be used for model training. Latitude, longitude, and frequency were initially used as input features on a smaller data set. Once transitioning to the larger data set consisting of 2400 stations, station separation distance was also included as an input feature for modeling. Modeling these four features provided good results, so other input features were explored. A fifth input feature, $V_{S30}$, was also found to improve model accuracy from 4-feature modeling. This will be shown below in section 5.2.2. These five input features are all connected to a specific location within the ROI. Latitude, longitude, station separation distance, frequency, and $V_{S30}$ remain constant no matter the event. Since these input features do not change with respect to the event, 4 and 5-feature GPR modeling is limited to non-ergodic site-specific events. To extend the capabilities of using GPR for coherency modeling, the radius to the event was considered since the value of this input feature changes depending on the event location. For constructing 6-feature models, a training set considering an equal amount of data from each of the three point-source events was used. Greater levels of uncertainty were expected since the model consisted of a small fraction of data from each of the three events, although it allowed GPR to be applied outside of an event-specific case. Considering the application of a physics-based coherency model, variability in events needs to be considered since coherency changes with respect to its source.
Results from GPR are included for the following cases: Event 1 trained on 4-feature inputs (longitude, latitude, frequency, station separation distance), Events 1, 2, 3, and M6.5 trained on 5-feature inputs (longitude, latitude, frequency, station separation distance, $V_{S30}$), and Events 1, 2, and 3 trained on 6-feature inputs (longitude, latitude, frequency, station separation distance, $V_{S30}$, and radius). Events 2, 3, and M6.5 were not trained on 4-feature models because event 1 results proved to be better in 5-feature modeling when compared to 4-feature modeling results. Event M6.5 was also not considered for 6-feature modeling since the nature of this event was different from the first three events.

### 5.2.2 Model Training

The following section details kernel development and prediction fitting from the eight Gaussian models used to generate coherency predictions. Hyperparameter convergence for the eight models is detailed in Figures 5-1 through 5-9 and summarized in Table 5-1. After discussing model hyperparameters, the model fits for seen and unseen stations during training are detailed, and the mean-squared error from each fit is presented in Figures 5-10 through 5-32.
The kernel variance provides insight into the spread of the data observed during training. As variance increases, so does the spread of the data. Observations of kernel variance from the five models show that the simulated fault rupture, EvM6.5, has a greater spread within the mean function used to form predictions. As observed in the ROI plots in Chapter 4, EvM6.5 displayed higher fluctuation in ground motion coherency when compared to the point source events. This is likely why the SE-ARD kernel for EvM6.5 has a significantly higher variance when compared to the point sources, as seen in Figure 5-1. Table 5-1 summarizes the converged hyperparameters for each event and shows the variance of EvM6.5 converges at a value of 0.778. In comparison, variance for the point source events all converged around a value of 0.5, proving the spread of data from the point source events is less than what is observed from the simulated fault rupture.
The first length scale inside the SE-ARD kernel, seen in Figure 5-2, forms a relationship between longitude and coherency. This value was initialized at 5.0 and trends towards 1 for all six kernels. Interestingly, the 6-feature point source model formed the weakest dependency between longitude and coherency. This weak relationship is likely because of the variability in coherency loss with respect to each event. This is not captured in the 4 and 5-feature models which shows the limitations of event-specific coherency models. Event 1 converged to nearly the same length scale value for the 4-feature and 5-feature models, although it is observed that their path to convergence is not the same. This is a good sanity check as it ensures the relationship holds when new features are introduced to the SE-ARD kernel. Events 2 and 3 also showed similar values for length scale 1, converging to values of 0.86 and 0.85, respectively.
Kernel length scale 2 follows similarly to length scale 1 as it describes the relationship each model has with the other positional feature, latitude, seen in Figure 5-3. The largest of which again comes from the 6-feature point source model, proving the variability in coherency loss with respect to the source location. Also like the longitudinal length scale, the latitudinal length scale hyperparameter for the 4-feature and 5-feature event 1 models converge to the same value. This provides robustness on the dependency these positional arguments have for predicting coherency on event-specific models. Again, it is noticed that the path to convergence between the 4-feature and 5-feature models for event 1 is different, even though the length scale converges to the same value. This difference in path to convergence is related to the additional length scale included in the 5-feature event.
Wave scattering effects are dominantly present in the relationships formed in the station separation distance length scale. The magnitude 6.5 fault simulation is an outlier event, holding a value of just 0.002 for this hyperparameter. The value converging to nearly zero shows there is a strong dependency between station separation distance and coherency from the simulated fault rupture. Supplemental to this length scale Figure 5-5 shows coherency loss with respect to distance for each SW4 simulation. It is observed that station separation distance has a greater influence on the fault rupture simulation, providing further evidence of intensified scattering during this event. The dependency of station separation distance and coherency coincides with the importance seen from empirical and semi-empirical coherency functions.
Figure 5-5: Coherency decay as a function of distance at 'S020.058_03'.

Figure 5-6: Hyperparameter optimization for SE-ARD length scale, frequency.
Kernel length scale 4, which describes the relationship between frequency and coherency is the only length scale where values converge to nearly the same value, as seen in Figure 5-6. Across all six of the kernels, an average length scale value of 0.58 was determined, showing a strong and consistent relationship with frequency. Like kernel length scale 3, the relationships formed between frequency and coherency coincide with empirical and semi-empirical models introduced in Chapter 2.

![Kernel Length Scale 5 - Vs30](image)

Figure 5-7: Hyperparameter optimization for SE-ARD length scale, Vs30.

The fifth length scale excludes the 4-feature model of event 1 since it was not trained on Vs30. Interestingly, this length scale shows the largest spread between events, with event 1 converging to a value of 0.96, and event 3 converging to a value of just 0.13, as seen in Figure 5-7. The larger spread shows the different influences Vs30 has across each event. With this information, it can be concluded that the relationship between Vs30 and ground motion coherency is dependent on source location and probably influenced by path effects in addition to site effects. Length
scale convergence is also different compared to the other four length scales as the values drop dramatically in later kernel iterations.

Figure 5-8: Hyperparameter optimization for SE-ARD kernel length scale, radius to the epicenter.

The sixth length scale introduced for modeling across the point source events shows an interesting path to convergence. It appears to have struggled learning the influence radius to epicenter has on coherency since its converged value was almost identical to its initialized value of 5.0. Part of this may be due to a lack of training data and is likely a sign of model underfitting. Considering the 6-feature model was trained on 45,000 of 540,000 true observations across the point source events (~8% of data) it would have likely found a more accurate model fit if computational restraints were not an issue for GPR.
The kernel likelihood is directly correlated to the variance of the kernel. As discussed in Chapter 3, Gaussian process regression minimizes variance by maximizing the log marginal likelihood. Maximizing the likelihood of the weights of kernel length scales provides the best distribution fit of the training data set and improves the predictions of testing data. It is not appropriate to compare the likelihood estimates between events because of the difference in coherency observations. Different coherency observations change the spread of data, which directly impacts recorded likelihoods.
Table 5-1: Summary of converged kernel hyperparameters from each model.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Optimized Hyperparameters</th>
<th>Lengthscales</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variances</td>
<td>Longitude</td>
<td>Latitude</td>
</tr>
<tr>
<td>Ev1 4 Feature</td>
<td>0.479</td>
<td>0.604</td>
<td>0.553</td>
</tr>
<tr>
<td>Ev1 5 Feature</td>
<td>0.480</td>
<td>0.612</td>
<td>0.559</td>
</tr>
<tr>
<td>Ev2 5 Feature</td>
<td>0.514</td>
<td>0.860</td>
<td>0.821</td>
</tr>
<tr>
<td>Ev3 5 Feature</td>
<td>0.541</td>
<td>0.852</td>
<td>0.685</td>
</tr>
<tr>
<td>EvM6.5 5 Feature</td>
<td>0.778</td>
<td>1.133</td>
<td>0.947</td>
</tr>
<tr>
<td>Point Source 6 Feature</td>
<td>0.636</td>
<td>1.522</td>
<td>1.665</td>
</tr>
</tbody>
</table>

Table 5-1 summarizes converged kernel hyperparameters from each event. While each of the six kernels converged, their length scale values shed light on how source, path, and site effects influence each ground motion differently.

The following figures provide an understanding of model fitting and how optimized hyperparameters influence the predictions of each model. Predictions were formed using initialized hyperparameter values when working with the Ev1 4-feature model to verify that optimized kernels were forming meaningful relationships to the coherency output. Variance and likelihood were initialized to 1.0, and each of the four length scales was initialized inside the kernel with a value of 5.0. The training set consisting of 45,000 of the 180,000 unique observations from the event 1 data set was used as model input for the generic model, and predictions were formed across the entire data set. Figure 5-10 shows the fit between training set predictions from the 4-feature generic kernel and true observations. It is clear to see there is a poor fit between generic kernel predictions and true coherency values. The same was done for
the testing data set, which consisted of the additional 135,000 observations that were not used for model training and can be seen in Figure 5-11. The fit using the generic kernel is even poorer than the predictions seen on the training set, which can be attributed to the fact that this data was never introduced to the model. As discussed previously, the training set consists of 45,000 observations from 600 of the 2400 stations. The testing set not introduced for hyperparameter tuning consists of the other 135,000 observations at the 1,800 unseen stations in the ROI. The poor model fit from the generic kernel works as a sanity check to ensure that hyperparameter tuning is improving model accuracy.

After checking results from generic predictions, hyperparameter tuning was done for the 4-feature event 1 model using the SciPy minimizer described earlier in Chapter 5. After tuning hyperparameters inside the kernel, predictions were formed again using the newly formed optimized kernel detailed in Table 5-1. Figure 5-12 shows the fit for observations introduced during model training, and Figure 5-13 shows the fit for the unseen data. Visually, it is clear to see the model performs much better after tuning the hyperparameters inside of the model’s SE-ARD kernel. To measure the performance of the optimized kernel versus the generic kernel, the mean squared error was calculated at each frequency for each set of observations and plotted in Figure 5-14.
Figure 5-10: Model fit of training data (25% of data set) for event 1 4-feature model using a generic kernel.
Figure 5-11: Model fit of testing data (75% of data set) for event 1 4-feature model using a generic kernel.
Figure 5-12: Model fit of training data (25% of data set) for event 1 4-feature model using an optimized kernel.
Figure 5-13: Model fit of testing data (75% of data set) for event 1 4-feature model using an optimized kernel.
The tracking of mean squared error (MSE) provides good insight into how hyperparameter tuning increases the accuracy of GPR predictions. The dashed lines show how MSE evolved through the generic kernel predictions, while the solid lines show how MSE evolved through the optimized kernel predictions. Figure 5-14 also shows how predictions from the model input, highlighted in green, performed better than the unseen data, highlighted in red. Further regression results at the stations introduced in Chapter 4 and ROI comparisons will be provided for generic and optimized kernel predictions later in this chapter.

Following the 4-feature model input results, additional SW4 simulations were provided for events 2, 3, and M6.5. It was decided to include $V_{S30}$ as an additional input feature and run simulations for all four events. Comparisons were made between MSE of Ev1 4-feature and MSE of Ev1 5-feature to ensure improved accuracy from the inclusion of this additional input feature. Interestingly, the MSE of model input predictions remained the same between each
event, although the MSE of predictions formed on unseen data decreased after $V_{S30}$ was incorporated as an input feature. This result informed a reasonable decision to include $V_{S30}$ for modeling events 2, 3, and M6.5. Following, Figures 5-15 through 5-22 show MSE results from 5-feature predictions against seen and unseen data for events 1, 2, 3, and M6.5, respectively. Figure 5-23 and Figure 5-24 summarize the findings and conclude these results. Generic predictions were not formed for the 5-feature predictions because they have proven to fit poorly against observed coherency from the event 1 SW4 simulation.
Figure 5-15: Model fit of training data (25% of data set) for event 1 5-feature model using an optimized kernel.
Figure 5-16: Model fit of testing data (75% of data set) for event 1 5-feature model using an optimized kernel.
Ev2: 5-Feature Optimized Kernel Predictions for Seen Data

Figure 5-17: Model fit of training data (25% of data set) for event 2 5-feature model using an optimized kernel.
Figure 5-18: Model fit of testing data (75% of data set) for event 2 5-feature model using an optimized kernel.
Figure 5-19: Model fit of training data (25% of data set) for event 3 5-feature model using an optimized kernel.
Figure 5-20: Model fit of testing data (75% of data set) for event 3 5-feature model using an optimized kernel.
Figure 5-21: Model fit of training data (25% of data set) for event M6.5 5-feature model using an optimized kernel.
Figure 5-22: Model fit of testing data (75% of data set) for event M6.5 5-feature model using an optimized kernel.
Figure 5-23: Mean-squared error comparison from the five model training sets.

Figure 5-23 shows the change in MSE between training set predictions and true observations with respect to frequency. Notice, the value of MSE remains the same for the 4 and 5-feature modeling of event 1. This shows there is no improvement in model accuracy for seen stations with the inclusion of $V_{s30}$ in model training.

Figure 5-24: Mean-squared error comparison from the five model testing sets.
Figure 5-24 shows the change in MSE between testing set predictions and true observations with respect to frequency. In this case, improved model accuracy is observed between the 4 and 5-feature predictions for event 1. This observation led to the inclusion of $V_{S30}$ in model training for events 2, 3, and M6.5. It is observed that MSE of the 4-feature model is largest at all frequencies, showing that the inclusion of $V_{S30}$ improved accuracy for predictions at unseen stations.

After understanding how well predictions were formed on event-specific models, 6-feature modeling was performed for each point source event. As discussed previously, the 6-feature model differs from the other five models because it considers multiple events. With more than one event in consideration, greater variability is introduced. Figures 5-25 through 5-32 show the fit each set of point source data had against the 6-feature model developed for events 1, 2, and 3.
Figure 5-25: Model fit of testing data (25% of data set) for event 1 6-feature model using an optimized kernel.
Figure 5-26: Model fit of Training data (75% of data set) for event 1 6-feature model using an optimized kernel.
Figure 5-27: Model fit of testing data (25% of data set) for event 2 6-feature model using an optimized kernel.
Figure 5-28: Model fit of training data (75% of data set) for event 2 6-feature model using an optimized kernel.
Figure 5-29: Model fit of testing data (25% of data set) for event 3 6-feature model using an optimized kernel.
Figure 5-30: Model fit of training data (75% of data set) for event 3 6-feature model using an optimized kernel.
Figure 5-31 summarizes how the MSE evolved as a function of frequency using the 6-feature SE-ARD kernel for predictions against each point source training set. Results show MSE is greater for Ev2 almost exclusively, likely attributed to greater wave scattering effects from this point source event. These higher levels of MSE are also observed in the 5-feature models for Ev2. Ev1 and Ev3 follow closely at lower levels of frequency but deviate at frequencies greater than 2.0Hz.
Figure 5-32: Mean-squared error comparison from the 6-feature model testing sets.

Figure 5-32 summarizes the change in MSE as a function of frequency for the testing sets of each point source event. Like the training predictions, Ev2 shows higher values of MSE at each frequency, adding to the scattering effects observed during this simulation. Overall, MSE values remain low for the predictions using 6-feature modeling providing good evidence towards capturing coherency trends when trained against multiple events.

5.3 Model Predictions

Section 5.3 presents model predictions for each event, separated into subsections for model predictions at training stations, model predictions at testing stations, and ROI contour plot comparisons for each of the regression models. ROI contour plots showing the difference between true coherency values and predictions are also provided. A gradient contour from blue to red is used for these additional ROI plots, where shades of blue detail predictions larger than the observations, and shades of red detail predictions smaller than the observations. Model predictions are provided at stations introduced in Chapter 4, along with ROI contour plots at
station separation distances of 200, 600, and 1000 meters. Additional model predictions can be found in Appendix D, along with ROI contour plots for station separation distances of 400 and 800 meters.

Coherency function analysis results are presented for the stations introduced in Chapter 4. These coherency function analysis results are divided into six subplots. The first subplot describes the station and event location, while subplots two through six describe coherency predictions at station separation distances of 200, 400, 600, 800, and 1000 meters, respectively. For each subplot, four curves are used to describe the relationship between true values, model predictions, and empirical function results. True coherency values from the SW4 simulations are plotted with the same color used to define the event: blue for event 1, red for event 2, green for event 3, and cyan for the magnitude 6.5 fault rupture simulation. The Abrahamson 1993 empirical coherency equation is plotted for comparison against predictions and is represented by an orange curve. The black dashed curve on each station subplot represents predictions formed using the optimized SE-ARD kernel with tuned hyperparameters described in the previous section. A confidence interval is also plotted with green fill to represent plus/minus two standard deviations away from the optimized SE-ARD kernel predictions. For the 4-feature figures, an additional red dashed line is presented to show predictions formed using the generic SE-ARD kernel. Plotting the predictions from the generic kernel provides evidence for the models learning and forming significant relationships with the provided input features through hyperparameter optimization.

ROI contour plots are presented similarly to those seen in Chapter 4. Each ROI contour plot provides a comparison between true coherency values and predictions formed using the
optimized SE-ARD kernel at each frequency step. These plots present each model's overall understanding of coherency within the region at given station separation distances and frequencies. Prediction results will be discussed in Section 5.4.

5.3.1 4-Feature Model Predictions for Event 1
5.3.1.1 4-Feature Model Predictions at Known Stations for Event 1

Figure 5-33: 4-feature model predictions for event 1 at station 'S026.074_03'.
Figure 5-34: 4-feature model predictions for event 1 at station 'S026.006_03'.

Figure 5-35: 4-feature model predictions for event 1 at station 'S002.006_03'.
5.3.1.2 4-Feature Model Predictions at Unknown Stations for Event 1

Figure 5-36: 4-feature model predictions for event 1 at station 'S005.028_03'.

Figure 5-37: 4-feature model predictions for event 1 at station 'S002.055_03'.
Figure 5-38: 4-feature model predictions for event 1 at station 'S010.005_03'.

Figure 5-39: 4-feature model predictions for event 1 at station 'S019.053_03'.

Figure 5-40: 4-feature model predictions for event 1 at station 'S026.041_03'.
Figure 5-41: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 200 meters.
Figure 5-42: Difference between 4-feature predictions and true coherency values for event 1 with a station separation distance of 200 meters.
Figure 5-43: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 600 meters.
Figure 5-44: Difference between 4-feature predictions and true coherency values for event 1 with a station separation distance of 600 meters.
Figure 5-45: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 1000 meters.
Figure 5-46: Difference between 4-feature predictions and true coherency values for event 1 with a station separation distance of 1000 meters.
5.3.2 5-Feature Model Predictions for Event 1

5.3.2.1 5-Feature Model Predictions at Known Stations for Event 1

Figure 5-47: 5-feature model predictions for event 1 at station 'S026.074_03'.

Figure 5-48: 5-feature model predictions for event 1 at station 'S026.006_03'.
Figure 5-49: 5-feature model predictions for event 1 at station 'S002.006_03'.

Figure 5-50: 5-feature model predictions for event 1 at station 'S005.028_03'.
5.3.2.2 5-Feature Model Predictions at Unknown Stations for Event 1

Figure 5-51: 5-feature model predictions for event 1 at station 'S002.055_03'.

Figure 5-52: 5-feature model predictions for event 1 at station 'S010.005_03'.

Figure 5-53: 5-feature model predictions for event 1 at station 'S019.053_03'.

Figure 5-54: 5-feature model predictions for event 1 at station 'S026.041_03'.
Figure 5-55: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 200 meters.
Figure 5-56: Difference between 5-feature predictions and true coherency values for event 1 with a station separation distance of 200 meters.
Figure 5-57: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 600 meters.
Figure 5-58: Difference between 5-feature predictions and true coherency values for event 1 with a station separation distance of 600 meters.
Figure 5-59: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 1000 meters.
Figure 5-60: Difference between 5-feature predictions and true coherency values for event 1 with a station separation distance of 1000 meters.
5.3.3 5-Feature Model Predictions for Event 2

5.3.3.1 5-Feature Model Predictions at Known Stations for Event 2

Figure 5-61: 5-feature model predictions for event 2 at station 'S026.074_03'.

Figure 5-62: 5-feature model predictions for event 2 at station 'S026.006_03'.
Figure 5-63: 5-feature model predictions for event 2 at station 'S002.006_03'.

Figure 5-64: 5-feature model predictions for event 2 at station 'S005.028_03'.
5.3.3.2 5-Feature Model Predictions at Unknown Stations for Event 2

Figure 5-65: 5-feature model predictions for event 2 at station 'S002.055_03'.

Figure 5-66: 5-feature model predictions for event 2 at station 'S010.005_03'.
Figure 5-67: 5-feature model predictions for event 2 at station 'S019.053_03'.

Figure 5-68: 5-feature model predictions for event 2 at station 'S026.041_03'.
5.3.3.3 5-Feature Region of Interest Contour Plot Comparison for Event 2

Figure 5-69: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 200 meters.
Figure 5-70: Difference between 5-feature predictions and true coherency values for event 2 with a station separation distance of 200 meters.
Figure 5-71: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 600 meters.
Figure 5-72: Difference between 5-feature predictions and true coherency values for event 2 with a station separation distance of 600 meters.
Figure 5-73: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 1000 meters.
Figure 5-74: Difference between 5-feature predictions and true coherency values for event 2 with a station separation distance of 1000 meters.
5.3.4 5-Feature Model Predictions for Event 3

5.3.4.1 5-Feature Model Predictions at Known Stations for Event 3

Figure 5-75: 5-feature model predictions for event 3 at station 'S026.074_03'.

Figure 5-76: 5-feature model predictions for event 3 at station 'S026.006_03'.
Figure 5-77: 5-feature model predictions for event 3 at station 'S002.006_03'.

Figure 5-78: 5-feature model predictions for event 3 at station 'S005.028_03'.
5.3.4.2 5-Feature Model Predictions at Unknown Stations for Event 3

Figure 5-79: 5-feature model predictions for event 3 at station 'S002.055_03'.

Figure 5-80: 5-feature model predictions for event 3 at station 'S010.005_03'.
Figure 5-81: 5-feature model predictions for event 3 at station 'S019.053_03'.

Figure 5-82: 5-feature model predictions for event 3 at station 'S026.041_03'.

5.3.4.3 5-Feature Region of Interest Contour Plot Comparison for Event 3

Figure 5-83: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 200 meters.
Figure 5-84: Difference between 5-feature predictions and true coherency values for event 3 with a station separation distance of 200 meters.
Figure 5-85: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 600 meters.
Figure 5-86: Difference between 5-feature predictions and true coherency values for event 3 with a station separation distance of 600 meters.
Figure 5-87: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 1000 meters.
Figure 5-88: Difference between 5-feature predictions and true coherency values for event 3 with a station separation distance of 1000 meters.
5.3.5 5-Feature Model Predictions for Event M6.5

5.3.5.1 5-Feature Model Predictions at Known Stations for Event M6.5

Figure 5-89: 5-feature model predictions for event M6.5 at station 'S026.074_03'.

Figure 5-90: 5-feature model predictions for event M6.5 at station 'S026.006_03'.
Figure 5-91: 5-feature model predictions for event M6.5 at station 'S002.006_03'.

Figure 5-92: 5-feature model predictions for event M6.5 at station 'S005.028_03'.
5.3.5.2 5-Feature Model Predictions at Unknown Stations for Event M6.5

Figure 5-93: 5-feature model predictions for event M6.5 at station 'S002.055_03'.

Figure 5-94: 5-feature model predictions for event M6.5 at station 'S010.005_03'.
Figure 5-95: 5-feature model predictions for event M6.5 at station 'S019.053_03'.

Figure 5-96: 5-feature model predictions for event M6.5 at station 'S026.041_03'.
5.3.5.3 5-Feature Region of Interest Contour Plot Comparison for Event M6.5

Figure 5-97: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 200 meters.
Figure 5-98: Difference between 5-feature predictions and true coherency values for event M6.5 with a station separation distance of 200 meters.
Figure 5-99: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 600 meters.
Figure 5-100: Difference between 5-feature predictions and true coherency values for event M6.5 with a station separation distance of 600 meters.
Figure 5-101: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 1000 meters.
Figure 5-102: Difference between 5-feature predictions and true coherency values for event M6.5 with a station separation distance of 1000 meters.
5.3.6 6-Feature Model Predictions for Event 1

5.3.6.1 6-Feature Model Predictions at Known Stations for Event 1

Figure 5-103: 6-feature model predictions for event 1 at station ‘S026.074_03’.

Figure 5-104: 6-feature model predictions for event 1 at station ‘S026.006_03’.
Figure 5-105: 6-feature model predictions for event 1 at station ‘S002.006_03’.

Figure 5-106: 6-feature model predictions for event 1 at station ‘S005.028_03’.
5.3.6.2 6-Feature Model Predictions at Unknown Stations for Event 1

Figure 5-107: 6-feature model predictions for event 1 at station ‘S002.055_03’.

Figure 5-108: 6-feature model predictions for event 1 at station ‘S010.005_03’.
Figure 5-109: 6-feature model predictions for event 1 at station 'S019.053_03'.

Figure 5-110: 6-feature model predictions for event 1 at station 'S026.041_03'.
5.3.6.3 6-Feature Region of Interest Contour Plot Comparison for Event 1

Figure 5-111: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 200 meters.
Figure 5-112: Difference between 6-feature predictions and true coherency values for event 1 with a station separation distance of 200 meters.
Figure 5-113: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 600 meters.
Figure 5-114: Difference between 6-feature predictions and true coherency values for event 1 with a station separation distance of 600 meters.
Figure 5-115: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 1000 meters.
Figure 5-116: Difference between 6-feature predictions and true coherency values for event 1 with a station separation distance of 1000 meters.
5.3.7 6-Feature Model Predictions for Event 2

5.3.7.1 6-Feature Model Predictions at Known Stations for Event 2

Figure 5-117: 6-feature model predictions for event 2 at station ‘S026.074_03’.

Figure 5-118: 6-feature model predictions for event 2 at station ‘S026.006_03’.
Figure 5-119: 6-feature model predictions for event 2 at station ‘S002.006_03’.

Figure 5-120: 6-feature model predictions for event 2 at station ‘S005.028_03’.
5.3.7.2 6-Feature Model Predictions at Unknown Stations for Event 2

Figure 5-121: 6-feature model predictions for event 2 at station ‘S002.055_03’.

Figure 5-122: 6-feature model predictions for event 2 at station ‘S010.005_03’.
Figure 5-123: 6-feature model predictions for event 2 at station ‘S019.053_03’.

Figure 5-124: 6-feature model predictions for event 2 at station ‘S026.041_03’.
Figure 5-125: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 200 meters.
Figure 5-126: Difference between 6-feature predictions and true coherency values for event 2 with a station separation distance of 200 meters.
Figure 5-127: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 600 meters.
Figure 5-128: Difference between 6-feature predictions and true coherency values for event 2 with a station separation distance of 600 meters.
Figure 5-129: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 1000 meters.
Figure 5-130: Difference between 6-feature predictions and true coherency values for event 2 with a station separation distance of 1000 meters.
5.3.8  6-Feature Model Predictions for Event 3

5.3.8.1  6-Feature Model Predictions at Known Stations for Event 3

Figure 5-131: 6-feature model predictions for event 3 at station ‘S026.074.03’.

Figure 5-132: 6-feature model predictions for event 3 at station ‘S026.006.03’.
Figure 5-133: 6-feature model predictions for event 3 at station ‘S002.006_03’.

Figure 5-134: 6-feature model predictions for event 3 at station ‘S005.028_03’.
5.3.8.2 6-Feature Model Predictions at Unknown Stations for Event 3

Figure 5-135: 6-feature model predictions for event 3 at station ‘S002.055_03’.

Figure 5-136: 6-feature model predictions for event 3 at station ‘S010.005_03’.
Figure 5.137: 6-feature model predictions for event 3 at station ‘S019.053_03’.

Figure 5.138: 6-feature model predictions for event 3 at station ‘S026.041_03’.
Figure 5-139: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 200 meters.
Figure 5-140: Difference between 6-feature predictions and true coherency values for event 3 with a station separation distance of 200 meters.
Figure 5-141: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 600 meters.
Figure 5-142: Difference between 6-feature predictions and true coherency values for event 3 with a station separation distance of 600 meters.
Figure 5-143: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 1000 meters.
Figure 5.144: Difference between 6-feature predictions and true coherency values for event 3 with a station separation distance of 1000 meters.
5.4 Discussion of Model Predictions

Coherency predictions from each of the eight models show that the SE-ARD kernel was able to form good relationships with the input features used to predict coherency values within the region of interest. Overall, the accuracy of each model gave consistent results and proved to be strongly related to the true values of coherency. It is also shown that the Abrahamson 1993 empirical equation poorly fits coherency values observed from the coherency analysis results of the SW4 simulations, which may have to do with discrepancies between empirical and simulated ground motion records. Although Gaussian modeling was not perfect, and each event showed similar performance issues, GPR was successfully applied to the data sets and formed accurate predictions on coherency within the ROI.

Forming predictions on training sets provided a good understanding of how well the model interpreted known data. These predictions provide an understanding of training errors within the data sets. Since the coherency data for each model is naturally noisy, the optimized SE-ARD kernels used for predictions were required to make certain generalizations for the model inputs, inducing a level of bias into its predictions. Even though the training data was seen for hyperparameter optimization of the SE-ARD kernel, predicting across training sets showed a level of error because of generalizations formed to relate each coherency observation. If there were a perfect correlation between each true coherency observation in the training set, i.e., zero noise within the relationship between input features and coherency observations, predictions on the training set would be identical to their actual values with a standard deviation of zero. Since this is not a practical occurrence for real-world data sets, it is expected that some level of bias is introduced inside of each model. In general, higher levels of bias seen in regression models
provide a looser fit to the model data, indicative of underfitting. Lower levels of bias seen in regression models provide a stronger fit to the model, although this is indicative of overfitting. Model bias is not inherently bad as it highlights the noise within the observed data.

The generic kernel predictions presented for the 4-feature model input show a lack of correlation between true and predicted values. This is expected since no relationship is formed between the input feature and coherency output. The generic kernel predictions also explain how well relationships were formed from the optimized kernel predictions. It is clearly shown that strong relationships are formed between the input features and coherency for the 4-feature model when predictions are compared between the generic and optimized kernels. For this reason, predictions were not generated for generic kernels in the 5-feature or 6-feature models.

Looking at the results from the Abrahamson 1993 empirical coherency equation, it is clearly seen that this traditional empirical equation generalizes the trend of coherency loss as frequency increases and does not capture the true variability observed from SW4 simulations. This is likely due to the equation only taking into consideration a general relationship between frequency, station separation distance, and coherency, while the SE-ARD kernel can form non-ergodic, site-specific relationships for coherency output in the San Francisco Bay Area. The kernel results show a strong relationship between coherency and station separation distance for the magnitude 6.5 fault rupture event. The station separation distance is extremely relevant to expected coherency, especially in more realistic fault-rupture scenarios which coincide with current empirical and semi-empirical functions.
Forming predictions at station locations not seen during kernel optimization show a weaker fit to the true values of coherency when compared to predictions formed on the training set. This is expected because the relationships between input features and coherency for hyperparameter optimization were formed with no prior knowledge of the testing sets. While the models do not predict on the testing sets as well as they do on training sets, testing set predictions do show good improvement compared to predictions formed by the generic kernel.

The ROI contour plots conclude the predictions formed across the defined region. The prediction contour plots emit a level of generalization, showing the training data was appropriate enough to capture the differences in coherency with respect to station location, separation distance, and frequency, but not at a granular level. These contour plots show how model bias influenced the predictions across each data set. When compared to the contour plots displaying true coherency values, the predictions are cloudy, although this is expected and highly influenced by the level of noise seen within the data set. Training on a larger portion of the data set is one option to combat the bias seen in each model. Unfortunately, this was not possible with the available computational powers, which is why only 25% of the data set was used for training on 4 and 5-feature models, and only ~8% for 6-feature models. Another possible way of improving the model predictions would be to introduce other input parameters which influence ground motion coherency. Later modeling will investigate further parameters such as earthquake magnitude.
6 Summary and Conclusions

6.1 Summary

Accurate modeling of ground motions is critical for the appropriate design of long-span structures. Due to the influence of site, path, and source effects on ground motion coherency, variability in ground motions is expected. The performance of long-span structures can dramatically suffer if ground motions are not modeled accurately. Long-span structures are particularly susceptible to incoherent ground motions because of their large footprints. Bridges, pipelines, and tunnels are just a few examples of structures susceptible to performance issues due to incoherent ground motions. Current engineering practice does not properly consider the effects of coherency. In many cases, engineers use empirical equations developed on little observational data. A better understanding of the physical parameters that influence ground motion coherency will help improve engineering practices and how structures are designed. As of now, there are currently no physics-based models capable of accounting for the actual effects of ground motion coherency.

The research performed in this thesis addresses why current models cannot capture the true effects of incoherent ground motions and presents further work towards accurately modeling coherency within the San Francisco Bay Area. GPR has been shown to model ground motion coherency accurately by considering input features not captured from traditional coherency equations. It has been shown that machine learning techniques significantly improve the accuracy of ground motion coherency and provide confidence levels from the results. Accounting for station location, frequency, station separation distance, average shear wave
velocity, and epicentral radius to determine coherency has provided a better understanding of the importance these physical parameters have on ground motion coherency.

6.2 Conclusions

There is still much to learn on how incoherent ground motions affect the performance of long-span structures. However, the research presented in this thesis provides forward progress towards the development of a physics-based coherency function. The investigation of simulated ground motions presented in this thesis has provided a better understanding of the relationship between ground motion coherency and the physical input features used for modeling.

Future work in this field of research will consider other physical parameters that may influence ground motion coherency and improve upon model accuracy. As mentioned previously, other physical parameters that may influence ground motion coherency include site characteristics, the magnitude of an event, fault mechanism, and source location. Different GPR kernels or kernel combinations, and extensions of the existing empirical or semi-empirical models to consider spatially varying coefficients may also provide better model predictions.

Different machine learning techniques may also provide better accuracy for modeling ground motion coherency. As discussed in Chapter 5, only 25% of the selected data sets were used for training because of computational limitations. While this amount of data captured the influence of different input features, modeling would improve if more data could be used for training. One way to increase the amount of data used for coherency modeling is through Sparse Gaussian Processes. Sparse GPs are approximate forms of GPR that train on several smaller data sets,
allowing for more data to be seen during training at faster computation times. Training across a smaller frequency step size could also be explored if other forms of modeling overcome computational limits. Another powerful machine learning technique that may also be used for coherency modeling is artificial neural networks.

Eventually, while the coherency models developed in the San Francisco Bay Area show accurate modeling of coherency, they do not apply to other locations such as Los Angeles or Seattle. To formulate a physics-based coherency model that can be applied at different locations outside the scope of this research, additional simulations at sites outside of the San Francisco Bay Area will need to be considered.
7 References


Appendix A: Informative MATLAB Scripts

Appendix A details the MATLAB scripts used for development of coherency data in Chapter 4. The ‘MainCoherency.m’ script is the parent script, while ‘coherencyanalysis.m’, ‘plotsac.m’, and ‘readsac.m’ are functions used to pull data from SW4 files and calculate lagged coherency at each station in the ROI.

MainCoherency.m:

```matlab
clc;clear;close all;
p3D50R1 = '3D_SPBA_cross.2_50m_R.sw4output/';
myf = 'MainCoherencyData'; mkdir(myf);
 fid = fopen('cross2.50.B.dat','r');
data = textscan(fid,'rec sta=%s file=%s lon=%f lat=%f depth=0.0 nsew=1');
lat = data{1,4}; lon = data{1,3}; STA = data{1,1};
sta = reshape(STA,14,2400);sta =sta'; sta2 = reshape(sta,2400*14,1);
[x,y,~]=ll2utm(lat,lon);
x = reshape(x,14,2400)/1000; x = x';
y = reshape(y,14,2400)/1000; y = y';
latr = reshape(lat,14,2400);
lonr = reshape(lon,14,2400);
geoplot(latr(3,:),lonr(3,:),'k*')

% fid = fopen('lon-latitude-elevation.in','w');
% for i = 1:length(lon)
% fprintf(fid,'%6f %6f %6f \n',lon(i),lat(i),0);
% end
% fclose(fid);

for i = 1:2400
    disp(i)
    dx = x-x(i,3);
    dy = y-y(i,3);
    d  = sqrt(dx.^2+dy.^2);
    d100 = find(abs(d-0.10)<0.01);
    d200 = find(abs(d-0.20)<0.01);
    d300 = find(abs(d-0.30)<0.01);
    d400 = find(abs(d-0.40)<0.01);
    d500 = find(abs(d-0.50)<0.01);
    d600 = find(abs(d-0.60)<0.01);
    d700 = find(abs(d-0.70)<0.01);
    d800 = find(abs(d-0.80)<0.01);
    d900 = find(abs(d-0.90)<0.01);
    d1000 = find(abs(d-1.0)<0.01);

    %master
    bm = [p3D50R1 sta{i,3}];
    [vx,vy,vz,ti] = plotsac(bm);
    dt = ti(2)-ti(1);
    fs = 1/dt; freq = 0:0.01:3.5;
```

twin = tukeywin(length(ti), 0.05);
axm = vx.*twin;
aym = vy.*twin;
azm = vz.*twin;

coherencyanalysis(d100, p3D50R1, 'd100.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d200, p3D50R1, 'd200.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d300, p3D50R1, 'd300.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d400, p3D50R1, 'd400.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d500, p3D50R1, 'd500.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d600, p3D50R1, 'd600.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d700, p3D50R1, 'd700.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d800, p3D50R1, 'd800.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d900, p3D50R1, 'd900.txt', axm, aym, freq, fs, sta2, bm, twin);
coherencyanalysis(d1000, p3D50R1, 'd1000.txt', axm, aym, freq, fs, sta2, bm, twin);
end
coherencyanalysis.m:

```matlab
function coherencyanalysis(d, path, filename, axm, aym, freq, fs, sta2, bm, twin)
    gx = zeros(length(freq), length(d));
    gy = gx;
    % gxang = gx; gyang = gy;
    for j = 1:length(d)
        bi = [path sta2{d(j)}];
        [vx, vy, ~, ~] = plotsac(bi);
        ax = vx.*twin;
        ay = vy.*twin;
        [cx, ~] = mscohere(axm, ax, [], [], freq, fs);
        [cy, f2] = mscohere(aym, ay, [], [], freq, fs);
        gx(:, j) = abs(cx);
        gy(:, j) = abs(cy);
        % gxang(:, j) = angle(cx);
        % gyang(:, j) = angle(cy);
    end
    g = [gx gy];
    % gang = [gxang gyang];
    gmean = mean(g, 2);
    cohdir = ['MainCoherencyData/','bm','/'];
    mkdir(cohdir);
    % csvwrite([cohdir, filename], [f2' g gang]);
    csvwrite([cohdir, filename], [f2' gmean]);
end

plotsac.m:

% PLOTSAC
% Read receiever data in format specified by USGS for the Hayward fault earthquake scenarios
% and plot it in 3 subwindows
% plotsac(basename, colorstring, erasefirst, timeshift)
% Input: basename - Name of receiever data file, basename.e, basename.n, basename.u
% colorstring: string passed to plot, like 'r' for red lines
% erasefirst: 0 does a 'hold on' for the current plot, otherwise erases the current figure
% timeshift: change independent variable to be t+timeshift
% function [ux, uy, uz, t] = plotsac(basename, colorstring, erase, tshift )

if nargin < 4
    tshift = 0;
end

if nargin < 3
    erase = 1;
end

if nargin < 2
    colorstring='b';
end;

[ux, dt, lat, lon] = readsac(sprintf('%s%s', basename, 'e'));
```
[uy] = readsac(sprintf('%s.%s', basename, 'n'));
[uz] = readsac(sprintf('%s.%s', basename, 'u'));

nt = length(ux);
t = dt*[0:nt-1];

% if (erase ~= 0)
% clf;
% end
% % east component
% subplot(3,1,1)
% if (erase == 0)
% hold on;
% end
% h=plot(t+tshift,ux,colorstring);
% set(h,'LineWidth',2.0)
% set(gca,'FontSize',20)
% legend('East','location','southeast');
% axis tight;
%
% % north component
% subplot(3,1,2)
% if (erase == 0)
% hold on;
% end
% h=plot(t+tshift,uy,colorstring);
% set(h,'LineWidth',2.0)
% set(gca,'FontSize',20)
% legend('North','location','southeast');
% axis tight;
%
% % up component
% subplot(3,1,3)
% if (erase == 0)
% hold on;
% end
% h=plot(t+tshift,uz,colorstring);
% set(h,'LineWidth',2.0)
% set(gca,'FontSize',20)
% legend('Up','location','southeast');
% axis tight;

readsac.m:
% READSAC
% Read SAC receiver data.
% [u, dt, lat, lon, b, e, npts, year, jday, hour, min, sec, msec, cmpaz, cmpinc, idep, stnam] = readsac( fname, format )
% Input: fname - Name of SAC file
% format - Little endian ('l') or big endian ('b')
% byte order for binary data. Default is 'l'.
% Output: u - The data component on SAC file
% dt - Uniform time step for u
% stalat, stalon - Latitude and longitude position of the SAC station.
% b - begin time relative reference datum
% e - end time relative reference datum
% npts - Number of elements in u
% year - reference datum
% Julian day - reference datum
% minute - reference datum
% second - reference datum
% micro sec - reference datum
% cmpaz - Azimuth angle of component (degrees)
% cmpinc - Inclination angle of component (degrees)
% idep - Code of component stored (6-displacement, 7-
velocity)
% stnam - Name of the station

function [u, dt, lat, lon, b, e, npts, year, jday, hour, min, sec, msec,
cmpaz, cmpinc, idep, stnam ] = readsac( fname, format )
    if nargin < 2
        format = 'l';
    end;
    fid = fopen(fname,'r',format);
    if fid < 0
        disp( ['Error: could not open file ' fname] );
    else
        dt = fread( fid,1,'float32' );
        fseek(fid,4*4,0);
        t0 = fread( fid,1, 'float32' );
        t1 = fread( fid,1, 'float32' );
        fseek(fid,24*4,0);
        lat = fread(fid,1, 'float32' );
        lon = fread(fid,1, 'float32' );
        fseek(fid,2*4,0);
        evlat = fread(fid,1, 'float32' );
        evlon = fread(fid,1, 'float32' );
        fseek(fid,4,0);
        evdepth = fread(fid,1, 'float32' );
        fseek(fid,4*18,0);
        cmpaz = fread(fid,1, 'float32' );
        cmpinc = fread(fid,1, 'float32' );
        fseek(fid,4*11,0);
        % fseek(fid,4*31,0);
        % integers from offset 70
        year = fread(fid,1,'int32');
        jday = fread(fid,1, 'int32');
        hour = fread(fid,1, 'int32');
        min = fread(fid,1, 'int32');
        sec = fread(fid,1, 'int32');
        msec = fread(fid,1, 'int32');
        nvhdr = fread(fid,1, 'int32');
        fseek(fid,4*2,0);
        npts=fread(fid,1, 'int32');
        fseek(fid,6*4,0);
        idep = fread(fid,1, 'int32');
        fseek(fid,23*4,0);
    end;
stnam = fread(fid,8,'char');
stnam = stnam';
fseek(fid,46*4,0);

% output reference time stamp
% disp(['Year = ' num2str(year) ' Julian Day = ' num2str(jday) ' Hour = ' num2str(hour) ' Min = ' num2str(min) ' Sec = ' num2str(sec) ' Micro Sec = ' num2str(msec)]);
% output required header data
% disp(['Begin time (B) = ' num2str(t0) ' End time (E) = ' num2str(t1) ' Station lat lon = ' num2str(lat) ' ' num2str(lon) ' nvhdr = ' num2str(nvhdr) ' npts = ' num2str(npts)]);
% read time series
u=fread(fid,npts,'float32');
% copy begin and end times
b = t0;
e = t1;
% disp(['cmpaz = ' num2str(cmpaz)]);
% disp(['cmpinc = ' num2str(cmpinc)]);
% disp(['idep = ' num2str(idep)]);
% printf('stnam   =  %s
', stnam);

fclose(fid);
end
Appendix B: Informative Python Syntax

For convenience, the syntax of pertinent GPR Python functions used for this research are provided. Snips from GPflow, Scikit-learn, and SciPy documentation follow in Appendix B for reference to concepts covered in Chapter 3 and Chapter 5.

```python
class gpflow.kernels.SquaredExponential(variance=1.0, lengthscale=1.0, **kwargs)

Bases: gpflow.kernels.stationary.IsotropicStationary

The radial basis function (RBF) or squared exponential kernel. The kernel equation is

\[ k(r) = \sigma^2 \exp\left(-\frac{1}{2} r^2\right) \]

where: \( r \) is the Euclidean distance between the input points, scaled by the lengthscale parameter \( \ell \). \( \sigma^2 \) is the variance parameter.

Functions drawn from a GP with this kernel are infinitely differentiable!

Figure B-1: Squared exponential kernel used for GPR models (from Matthews et al., 2017).
```
**minimize**(*closure*, *variables*, *method='L-BFGS-B*', *step_callback=None*, *compile=True*, *allow_unused_variables=False*, **scipy_kwargs*)

Minimize is a wrapper around the `scipy.optimize.minimize` function handling the packing and unpacking of a list of shaped variables on the TensorFlow side vs. the flat numpy array required on the Scipy side.

**Args:**

- **closure**: A closure that re-evaluates the model, returning the loss to be minimized.
- **variables**: The list (tuple) of variables to be optimized (typically `model.trainable_variables`)
- **method**: The type of solver to use in SciPy. Defaults to "L-BFGS-B". *step_callback*: If not None, a callable that gets called once after each optimisation step. The callable is passed the arguments *step*, *variables*, and *values*. *step* is the optimisation step counter, *variables* is the list of trainable variables as above, and *values* is the corresponding list of tensors of matching shape that contains their value at this optimisation step.

---

Figure B-2: Minimize wrapper function for SciPy L-BFGS-B minimizer function (*from Matthews et al., 2017*).
Figure B-3: Defining GPR model for coherency predictions (from Matthews et al., 2017).
Figure B-4: Callout for SciPy minimizing method (from Matthews et al., 2017).

Figure B-5: Callout for GP predictions using mean function (from Matthews et al., 2017).

Figure B-6: Callout for trainable variables (from Matthews et al., 2017).

Figure B-7: Model training loss callout (from Matthews et al., 2017).
**Figure B-8:** Display option for kernel hyperparameters (from Matthews et al., 2017).

**Figure B-9:** Training and testing split of data sets (from Pedregosa et al., 2011).

**Figure B-10:** Function used for calculating MSE (from Pedregosa et al., 2011).

**Figure B-11:** L-BFGS-B minimizer called in GPflow (from Virtanen et al., 2020).
Appendix C: Supplemental Coherency Analysis Results

Appendix C provides supplemental information for data generated from MATLAB coherency analysis. The following results follow the same structure as Figures found in Chapter 4.

Figure C-1: Coherency function results for station 'S020.058_03'.
Figure C-2: Coherency function results for station 'S020.034_03'.

Figure C-3: Coherency function results for station 'S011.044_03'.
Figure C-4: Coherency function results for station 'S011.020_03'.

Figure C-5: Coherency function results for station 'S004.074_03'.
Figure C-6: Coherency function results for station 'S004.006_03'.

Figure C-7: Coherency function results for station 'S002.026_03'.
Figure C-8: Coherency function results for station 'S005.008_03'.

Figure C-9: Coherency function results for station 'S002.027_03'.
Figure C-10: Coherency function results for station 'S009.077_03'.

Figure C-11: Coherency function results for station 'S011.035_03'.

Figure C-12: Coherency function results for station 'S014.077_03'.

Figure C-13: Coherency function results for station 'S017.023_03'.
Figure C-14: Coherency function results for station 'S021.058_03'.

Figure C-15: Coherency function results for station 'S022.008_03'.
Figure C-16: Coherency function results for station 'S026.079_03'.
Appendix D: Supplemental GPR Results

Figures in Appendix D are provided to supplement Chapter 5. The following results give more insight on accuracy of model training. Stations not introduced in the body of this thesis are shown below, along with ROI contour plots for station separation distances of 400 and 800 meters.

4-feature Model Predictions at Known Stations for Event 1:

Figure D-1: 4-feature model predictions for event 1 at station 'S020.058_03'.
Figure D-2: 4-feature model predictions for event 1 at station 'S020.034_03'.

Figure D-3: 4-feature model predictions for event 1 at station 'S011.044_03'.

Figure D-4: 4-feature model predictions for event 1 at station 'S011.020_03'.

Figure D-5: 4-feature model predictions for event 1 at station 'S004.074_03'.
Figure D-6: 4-feature model predictions for event 1 at station 'S004.006_03'.

Figure D-7: 4-feature model predictions for event 1 at station 'S002.026_03'.
Figure D-8: 4-feature model predictions for event 1 at station 'S005.008_03'.

4-feature Model Predictions at Unknown Stations for Event 1:

Figure D-9: 4-feature model predictions for event 1 at station 'S002.027_03'.
Figure D-10: 4-feature model predictions for event 1 at station 'S009.077_03'.

Figure D-11: 4-feature model predictions for event 1 at station 'S011.035_03'.
Figure D-12: 4-feature model predictions for event 1 at station 'S014.077_03'.

Figure D-13: 4-feature model predictions for event 1 at station 'S017.023_03'.

Figure D-14: 4-feature model predictions for event 1 at station 'S021.058_03'.

Figure D-15: 4-feature model predictions for event 1 at station 'S022.008_03'.
Figure D-16: 4-feature model predictions for event 1 at station 'S026.079_03'.
4-feature Region of Interest Contour Plot Comparison for Event 1:

Figure D-17: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 400 meters.
Figure D-18: 4-feature ROI prediction comparisons for event 1 with a station separation distance of 800 meters.
5-feature Model Predictions at Known Stations for Event 1:

Figure D-19: 5-feature model predictions for event 1 at station 'S020.058_03'.

Figure D-20: 5-feature model predictions for event 1 at station 'S020.034_03'.

Figure D-21: 5-feature model predictions for event 1 at station 'S011.044_03'.

Figure D-22: 5-feature model predictions for event 1 at station 'S011.020_03'.
Figure D-23: 5-feature model predictions for event 1 at station 'S004.074_03'.

Figure D-24: 5-feature model predictions for event 1 at station 'S004.006_03'.
Figure D-25: 5-feature model predictions for event 1 at station 'S002.026_03'.

Figure D-26: 5-feature model predictions for event 1 at station 'S005.008_03'.
5-feature Model Predictions at Unknown Stations for Event 1:

Figure D-27: 5-feature model predictions for event 1 at station 'S002.027_03'.

Figure D-28: 5-feature model predictions for event 1 at station 'S009.077_03'.
Figure D-29: 5-feature model predictions for event 1 at station 'S011.035_03'.

Figure D-30: 5-feature model predictions for event 1 at station 'S014.077_03'.
Figure D-31: 5-feature model predictions for event 1 at station 'S017.023_03'.

Figure D-32: 5-feature model predictions for event 1 at station 'S021.058_03'.
Figure D-33: 5-feature model predictions for event 1 at station 'S022.008_03'.

Figure D-34: 5-feature model predictions for event 1 at station 'S026.079_03'.
5-feature Region of Interest Contour Plot Comparison for Event 1:

Figure D-35: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 400 meters.
Figure D-36: 5-feature ROI prediction comparisons for event 1 with a station separation distance of 800 meters.
5-feature Model Predictions at Known Stations for Event 2:

Figure D-37: 5-feature model predictions for event 2 at station 'S020.058_03'.

Figure D-38: 5-feature model predictions for event 2 at station 'S020.034_03'.
Figure D-39: 5-feature model predictions for event 2 at station 'S011.044_03'.

Figure D-40: 5-feature model predictions for event 2 at station 'S011.020_03'.
Figure D-41: 5-feature model predictions for event 2 at station 'S004.074_03'.

Figure D-42: 5-feature model predictions for event 2 at station 'S004.006_03'.

Figure D-43: 5-feature model predictions for event 2 at station 'S002.026_03'.

Figure D-44: 5-feature model predictions for event 2 at station 'S005.008_03'.
5-feature Model Predictions at Unknown Stations for Event 2:

Figure D-45: 5-feature model predictions for event 2 at station 'S002.027_03'.

Figure D-46: 5-feature model predictions for event 2 at station 'S009.077_03'.
Figure D-47: 5-feature model predictions for event 2 at station 'S011.035_03'.

Figure D-48: 5-feature model predictions for event 2 at station 'S014.077_03'. 
Figure D-49: 5-feature model predictions for event 2 at station 'S017.023_03'.

Figure D-50: 5-feature model predictions for event 2 at station 'S021.058_03'.
Figure D-51: 5-feature model predictions for event 2 at station 'S022.008_03'.

Figure D-52: 5-feature model predictions for event 2 at station 'S026.079_03'.
5-feature Region of Interest Contour Plot Comparison for Event 2:

Figure D-53: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 400 meters.
Figure D-54: 5-feature ROI prediction comparisons for event 2 with a station separation distance of 800 meters.
5-feature Model Predictions at Known Stations for Event 3:

Figure D-55: 5-feature model predictions for event 3 at station 'S020.058_03'.

Figure D-56: 5-feature model predictions for event 3 at station 'S020.034_03'.
Figure D-57: 5-feature model predictions for event 3 at station 'S011.044_03'.

Figure D-58: 5-feature model predictions for event 3 at station 'S011.020_03'.
Figure D-59: 5-feature model predictions for event 3 at station 'S004.074_03'.

Figure D-60: 5-feature model predictions for event 3 at station 'S004.006_03'.
Figure D-61: 5-feature model predictions for event 3 at station 'S002.026_03'.

Figure D-62: 5-feature model predictions for event 3 at station 'S005.008_03'.
5-feature Model Predictions at Unknown Stations for Event 3:

Figure D-63: 5-feature model predictions for event 3 at station 'S002.027_03'.

Figure D-64: 5-feature model predictions for event 3 at station 'S009.077_03'.
Figure D-65: 5-feature model predictions for event 3 at station 'S011.035_03'.

Figure D-66: 5-feature model predictions for event 3 at station 'S014.077_03'.
Figure D-67: 5-feature model predictions for event 3 at station 'S017.023_03'.

Figure D-68: 5-feature model predictions for event 3 at station 'S021.058_03'.
Figure D-69: 5-feature model predictions for event 3 at station 'S022.008_03'.

Figure D-70: 5-feature model predictions for event 3 at station 'S026.079_03'.
Figure D-71: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 400 meters.
Figure D-72: 5-feature ROI prediction comparisons for event 3 with a station separation distance of 800 meters.
5-feature Model Predictions at Unknown Stations for Event M6.5

Figure D-73: 5-feature model predictions for event M6.5 at station 'S020.058_03'.

Figure D-74: 5-feature model predictions for event M6.5 at station 'S020.034_03'.
Figure D-75: 5-feature model predictions for event M6.5 at station 'S011.044_03'.

Figure D-76: 5-feature model predictions for event M6.5 at station 'S011.020_03'.

Figure D-77: 5-feature model predictions for event M6.5 at station 'S004.074_03'.

Figure D-78: 5-feature model predictions for event M6.5 at station 'S004.006_03'.
Figure D-79: 5-feature model predictions for event M6.5 at station 'S002.026_03'.

Figure D-80: 5-feature model predictions for event M6.5 at station 'S005.008_03'.
5-feature Model Predictions at Unknown Stations for Event M6.5:

Figure D-81: 5-feature model predictions for event M6.5 at station 'S002.027_03'.

Figure D-82: 5-feature model predictions for event M6.5 at station 'S009.077_03'.
Figure D-83: 5-feature model predictions for event M6.5 at station 'S011.035_03'.

Figure D-84: 5-feature model predictions for event M6.5 at station 'S014.077_03'.
Figure D-85: 5-feature model predictions for event M6.5 at station 'S017.023_03'.

Figure D-86: 5-feature model predictions for event M6.5 at station 'S021.058_03'.
Figure D-87: 5-feature model predictions for event M6.5 at station 'S022.008_03'.

Figure D-88: 5-feature model predictions for event M6.5 at station 'S026.079_03'.
5-feature Region of Interest Contour Plot Comparison for Event M6.5:

Figure D-89: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 400 meters.
Figure D-90: 5-feature ROI prediction comparisons for event M6.5 with a station separation distance of 800 meters.
6-feature Model Predictions at Known Stations for Event 1:

Figure D-91: 6-feature model predictions for event 1 at station 'S020.058_03'.

Figure D-92: 6-feature model predictions for event 1 at station 'S020.034_03'.

Figure D-93: 6-feature model predictions for event 1 at station 'S011.044_03'.

Figure D-94: 6-feature model predictions for event 1 at station 'S011.020_03'.

Figure D-95: 6-feature model predictions for event 1 at station 'S004.074_03'.

Figure D-96: 6-feature model predictions for event 1 at station 'S004.006_03'.
Figure D-97: 6-feature model predictions for event 1 at station 'S002.026_03'.

Figure D-98: 6-feature model predictions for event 1 at station 'S005.008_03'.
6-feature Model Predictions at Unknown Stations for Event 1:

Figure D-99: 6-feature model predictions for event 1 at station 'S002.027_03'.

Figure D-100: 6-feature model predictions for event 1 at station 'S009.077_03'.
Figure D-101: 6-feature model predictions for event 1 at station 'S011.035_03'.

Figure D-102: 6-feature model predictions for event 1 at station 'S014.077_03'.
Figure D-103: 6-feature model predictions for event 1 at station 'S017.023_03'.

Figure D-104: 6-feature model predictions for event 1 at station 'S021.058_03'.
Figure D-105: 6-feature model predictions for event 1 at station 'S022.008_03'.

Figure D-106: 6-feature model predictions for event 1 at station 'S026.079_03'.
Figure D-107: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 400 meters.
Figure D-108: 6-feature ROI prediction comparisons for event 1 with a station separation distance of 800 meters.
6-feature Model Predictions at Known Stations for Event 2:

Figure D-109: 6-feature model predictions for event 2 at station 'S020.058_03'.

Figure D-110: 6-feature model predictions for event 2 at station 'S020.034_03'.

Figure D-111: 6-feature model predictions for event 2 at station 'S011.044_03'.

Figure D-112: 6-feature model predictions for event 2 at station 'S011.020_03'.
Figure D-113: 6-feature model predictions for event 2 at station 'S004.074_03'.

Figure D-114: 6-feature model predictions for event 2 at station 'S004.006_03'.
Figure D-115: 6-feature model predictions for event 2 at station 'S002.026_03'.

Figure D-116: 6-feature model predictions for event 2 at station 'S005.008_03'.
6-feature Model Predictions at Unknown Stations for Event 2:

Figure D-117: 6-feature model predictions for event 2 at station 'S002.027_03'.

Figure D-118: 6-feature model predictions for event 2 at station 'S009.077_03'.
Figure D-119: 6-feature model predictions for event 2 at station 'S011.035_03'.

Figure D-120: 6-feature model predictions for event 2 at station 'S014.077_03'.
Figure D-121: 6-feature model predictions for event 2 at station 'S017.023_03'.

Figure D-122: 6-feature model predictions for event 2 at station 'S021.058_03'.
Figure D-123: 6-feature model predictions for event 2 at station 'S022.008_03'.

Figure D-124: 6-feature model predictions for event 2 at station 'S026.079_03'.
6-feature Region of Interest Contour Plot Comparison for Event 2:

Figure D-125: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 400 meters.
Figure D-126: 6-feature ROI prediction comparisons for event 2 with a station separation distance of 800 meters.
6-feature Model Predictions at Known Stations for Event 3:

Figure D-127: 6-feature model predictions for event 3 at station 'S020.058_03'.

Figure D-128: 6-feature model predictions for event 3 at station 'S020.034_03'.
Figure D-129: 6-feature model predictions for event 3 at station 'S011.044_03'.

Figure D-130: 6-feature model predictions for event 3 at station 'S011.020_03'.
Figure D-131: 6-feature model predictions for event 3 at station 'S004.074_03'.

Figure D-132: 6-feature model predictions for event 3 at station 'S004.006_03'.
Figure D-133: 6-feature model predictions for event 3 at station 'S002.026_03'.

Figure D-134: 6-feature model predictions for event 3 at station 'S005.008_03'.
6-feature Model Predictions at Unknown Stations for Event 3:

Figure D-135: 6-feature model predictions for event 3 at station 'S002.027_03'.

Figure D-136: 6-feature model predictions for event 3 at station 'S009.077_03'.
Figure D-137: 6-feature model predictions for event 3 at station 'S011.035_03'.

Figure D-138: 6-feature model predictions for event 3 at station 'S014.077_03'.
Figure D-139: 6-feature model predictions for event 3 at station 'S017.023_03'.

Figure D-140: 6-feature model predictions for event 3 at station 'S021.058_03'. 
Figure D-141: 6-feature model predictions for event 3 at station 'S022.008_03'.

Figure D-142: 6-feature model predictions for event 3 at station 'S026.079_03'.
6-feature Region of Interest Contour Plot Comparison for Event 3:

Figure D-143: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 400 meters.
Figure D-144: 6-feature ROI prediction comparisons for event 3 with a station separation distance of 800 meters.