Cal Poly

Caltech

SCEC
AN NSF+USGS CENTER

UC Irvine

UCLA

UC Santa Barbara

USC

Uncertainty and Sensitivity Analysis of Ground-Motions Simulation at Un-Instrumented Sites Using Gaussian Process Regression

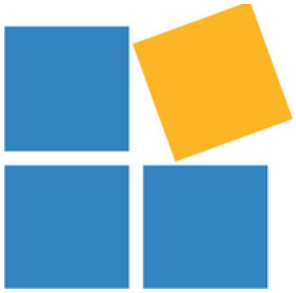
Aidin Tamhidi
Nicolas M. Kuehn, PhD
Yousef Bozorgnia, PhD

Department of Civil and Environmental Engineering
University of California, Los Angeles

Report GIRS-2022-11

DOI: 10.34948/N39G6W

University of California, Los Angeles (headquarters)



Natural Hazards Risk & Resiliency Research Center

B. John Garrick Institute for the Risk Sciences

Uncertainty and Sensitivity Analysis of Ground-Motions Simulation at Un-Instrumented Sites Using Gaussian Process Regression

Aidin Tamhidi
Nicolas M. Kuehn, PhD
Yousef Bozorgnia, PhD

Civil and Environmental Engineering Department
University of California, Los Angeles

A report on research conducted with supports from National Science Foundation, California Department of Transportation, and Pacific Gas & Electric Company.

Report GIRS-2022-11
DOI: 10.34948/N39G6W

Natural Hazards Risk and Resiliency Research Center
B. John Garrick Institute for the Risk Sciences
University of California, Los Angeles (Headquarters)

June 2022

ABSTRACT

Earthquake ground motion time series plays a critical role in the performance assessment of the structures, especially when nonlinear response history analysis for a specific structural system is required. The number of currently available recording instruments is sparse. Therefore, it is necessary to have a reliable methodology to construct the ground motion time series at the desired target un-instrumented sites. Using the Gaussian Process Regression (GPR), we recently presented an approach for generating ground motion time series at target sites where there are no available recording sensors. This model is trained based on physics-based simulated earthquake datasets in northern California and evaluated using the recorded motions during the 2019 **M**7.1 Ridgecrest earthquake sequence and 2020 **M**4.5 South El Monte datasets in Southern California. This GPR method interpolates the observed Discrete Fourier Transform (DFT) coefficients to construct the frequency-content of the ground motion at the target location and generate time series at the site. The optimized hyperparameter of the GPR model depends on the observation density of the training dataset. Thus, in this study, we tuned the hyperparameter of the GPR model based on observation density using the 2019 **M**7.1 Ridgecrest earthquake dataset recorded by the Community Seismic Network (CSN). In addition, we introduce a methodology to generate random realizations of ground motions using the trained GPR model at each target site. We utilize this methodology for the 2019 **M**7.1 Ridgecrest earthquake to conduct uncertainty quantification of the estimated motions at short and long periods. The results illustrate that uncertainty of the generated time series is lower for longer periods than that for shorter periods. In addition, we carried out the sensitivity analysis of both predictions' error and uncertainty with respect to a variety of governing parameters such as density of the observations surrounding the target site and estimated uncertainty of the local site conditions. It is shown that the observation density plays a key role in both reducing the prediction error as well as the uncertainty of the estimation. Moreover, we studied the improvement of the performance of the GPR model in the prediction of ground motions for 2019 **M**7.1 Ridgecrest as well as the 2020 **M**4.5 South El Monte earthquakes recorded by the California Integrated Seismic Network (CISN) through feeding more observed motions from CSN sites to the model. The results illustrate that the prediction error decreases, especially for those target sites located inside the added observed network (CSN) boundary. However, the prediction uncertainty is not changed considerably, especially at short periods.

ACKNOWLEDGMENTS

This study is partially supported by the University of California, Los Angeles (UCLA) Graduate Fellowship to the first author, which is gratefully appreciated. In addition, partial supports from the National Science Foundation (Award number 2025310), California Department of Transportation, and Pacific Gas & Electric Company are gratefully appreciated. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the supporting organizations.

The authors would like to thank Tadahiro Kishida and Omar Issa for their efforts in processing raw ground motions of earthquakes. Also, Sean Ahdi, Pengfei Wang, and Chukwuebuka C. Nweke have kindly assisted in estimating Vs_{30}, $Z_{1.0}$, and $Z_{2.5}$ values at the recording stations of the Community Seismic Network sites.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

The ground motion time series play an important role in earthquake engineering, including post-earthquake performance assessment and analysis of the site-specific structures. The current number of ground-level recording instruments in California is approximately 2000 over multiple networks: Southern California Seismic Network, Northern California Seismic Network, and California Strong Motion Instrumentation Program (Southern California Earthquake Data Center, 2021). Thus, for post-earthquake performance assessment for site-specific structures, the input ground motion time series must be estimated for locations with no recording instruments. Currently, "ShakeCast" and "ShakeMap" developed by U.S. Geological Survey (Fraser et al., 2008; Wald et al., 2008; Lin et al., 2018; Worden et al., 2018) provide Ground Motion Intensity Measures (GMIM) after an event in a near-real-time manner. These platforms use the neighboring recorded GMIM to interpolate and estimate the target sites' GMIM (Worden et al., 2018; Baker and Chen, 2020; Otake et al., 2020). However, for more detailed information about the seismic response of the structures after an event, nonlinear response history analysis requires the ground motion time series as input. Therefore a reliable generation of the ground motion time series is required (Petrone et al. 2021). Furthermore, these generated ground motions must be compatible with the spatial variation of amplitude, phase, and frequency models over the target region (Zerva and Zervas 2002; Zerva 2009; Chen and Baker 2019; Jayaram and Baker, 2009; Adanur et al., 2016; Tian et al., 2016; Zerva et al., 2018; Tamhidi et al., 2022).

In general, there are two methods for generating ground motion time series: 1) physics-based simulations, which use fine-fault and seismic velocity models that can account for source, path, and site effects (Boore, 2003; Aagaard et al., 2008a; Aagaard et al., 2008b; Atkinson and Assatourians, 2015) as well as the topography of the Earth surface (Rodgers et al., 2019). The high computational costs of such approaches are a concern since they demand precise information on the site attributes and fault patterns. As a result, these procedures are impractical for assessing post-earthquake damages rapidly (Loos et al., 2020; Mangalathu and Jeon, 2020), and 2) Simulations based on the coherency functions using cross-spectral density (CSD) and auto-spectral density (ASD) functions (Kameda and Morikawa, 1992; Konakli and Der Kiureghian, 2012; Zentner, 2013; Rodda and Basu, 2018). Simulated ground motions commonly are generated based on the CSD function, which itself is determined using empirical coherency functions, the coefficients of which are typically set through data-driven methods (Abrahamson et al., 1991). Furthermore, finely detailed information regarding the site properties and wave propagation characteristics might be needed through these methods, which brings the similar issue of being computationally expensive and time-consuming. Therefore, both methods have their critical challenges for a rapid post-earthquake structural damage assessment in a real-time manner.

A recently published model on the conditioned simulation of ground motions interpolates the discrete Fourier transform (DFT) coefficients of the observed motions to construct the ground motion time series at the target locations (Tamhidi et al., 2021). Tamhidi et al. employed Gaussian Process Regression (GPR), also known as Kriging (Rasmussen and Williams, 2006), for the interpolation procedure. The GPR method is shown to be capable of performing conditioned simulations of ground motions at a set of dense target sites with comparatively cheap computational costs (Tamhidi et al., 2021). In addition, Tamhidi et al. (2021) illustrated that the trained GPR method could accurately predict the ground motion content within the frequency range pertinent to most earthquake engineering problems. The previously established GPR method's performance is investigated based on the predicted ground motions constructed with the posterior mean of DFT coefficients. In addition, it is required to know about the uncertainty of these predicted ground motions through the trained GPR model.

The focus of the current study is the quantification of the uncertainty and validity of the predicted motions using the GPR model introduced by Tamhidi et al. (2021). First, we tune the hyper-parameter of the model for various observation density values. This will enable the users to choose the hyper-parameter of the models pertinent to their specific problem based on the available observations. Then, we introduce a methodology to generate random realizations of ground motions using the trained GPR model and an established inter-frequency correlation model (Bayless and Abrahamson, 2019). This random realization methodology provides an ensemble of random samples of motions for the target sites. Thus, the user employs the potential variation of the ground motions as input to the desired structures for analysis. We utilized this methodology for the 2019 **M**7.1 Ridgecrest Earthquake Sequence recorded over the Community Seismic Network (CSN) and investigated the sensitivity of the accuracy and uncertainty of the estimated motions to various parameters, such as observation density surrounding each target site and uncertainty of the estimated local site condition. Moreover, we studied the improvement of the GPR model's performance for the prediction of the ground motions recorded at California Integrated Seismic Network (CISN) stations for both 2019 **M**7.1 Ridgecrest and 2020 **M**4.5 South El Monte earthquake datasets by feeding more observed motions from CSN to the model.

Chapter 2 of this report provides an overview of the theoretical background of the proposed methodology and Gaussian Process Regression. The model performance is evaluated on several earthquake ground motion datasets in Chapter 3, and the optimized hyperparameters are introduced based on the observed dataset's density in Chapter 4. The production of ground motion realizations and the uncertainty quantification of these generated motions are discussed in Chapters 5 and 6, respectively. Eventually, Chapter 7 evaluates the GPR's performance in the presence of more observed ground motions from different seismic networks.

2. Theoretical Background

One can decompose the acceleration time series of the ground motion at the station s, $a_s(t)$, constructed with N discrete data points, $a_s(t_i)$, $i = 1, \dots, N$, at equal time intervals, Δt, into its Discrete Fourier Transform (DFT) coefficients A_k (Oppenheim and Willsky 2010)

$$a_s(t_i) = \sum_{k=0}^{N-1} A_k e^{j\omega_k t_i} \tag{2.1}$$

where

$$A_k = \frac{1}{N}\sum_{i=0}^{N-1} a_s(t_i)[\cos(\omega_k t_i) + j \cdot \sin(\omega_k t_i)] = \mathcal{R}e_k + j \cdot \mathcal{I}m_k. \tag{2.2}$$

In Eqs. (2.1) and (2.2), ω_k, is the k^{th} natural frequency of the DFT and $j = \sqrt{-1}$. $\mathcal{R}e_k$ and $\mathcal{I}m_k$ are also the real and imaginary parts of the DFT coefficient at the k^{th} natural frequency. One can reconstruct the whole ground motion time series having the real and imaginary parts of the DFT coefficients at different frequencies, $k = 0, \dots, N-1$ using Eq. (2.2) and Eq. (2.1).

Tamhidi et al. (2021) simulated the ground motion time series at uninstrumented sites by estimating the real and imaginary parts of the DFT coefficients (at various frequencies) having surrounded neighboring stations' motions observed. They implemented Gaussian Process Regression (GPR) to predict the $\mathcal{R}e_k$ and $\mathcal{I}m_k$ in Eq. (2.2) using the observed $\mathcal{R}e'_k$ and $\mathcal{I}m'_k$ at neighboring stations, s', at each frequency.

Next, we will review the definition of a Gaussian Process (GP) and how GPR is deployed to estimate the value of the GP at the unobserved locations using the observed ground motions.

2.1. Gaussian Process Regression

GPR is a supervised machine learning method that was used in an extensive area of research, such as post-earthquake damage assessment, conditioned simulation of ground motion, and seismic fragility assessment (Tamhidi et al., 2019, 2020, 2021, 2022; Gentile and Galasso 2019). A GP is a collection of random variables as a function such that every finite subset of them follows the multivariate Gaussian distribution (Rasmussen and Williams 2006). The general form of a function as a GP is shown in Eq. (2.3).

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \tag{2.3}$$

In Eq. (2.3), $m(\boldsymbol{x})$ is the mean function value at the input location $\boldsymbol{x}$, and $k(\boldsymbol{x},\boldsymbol{x}')$ is the covariance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ locations.

Suppose $\boldsymbol{f}$ is the observed value of the GP and $\boldsymbol{f}_*$ are the GP values at the target unobserved locations. Also, let's denote the observed locations' input matrix as $\boldsymbol{X}$ each row of which includes one observed location's input feature vector. Similarly, we call $\boldsymbol{X}_*$ as the input matrix of the target (unobserved) locations. The predictive distribution of the $\boldsymbol{f}_*$ then is given by (Rasmussen and Williams 2006)

$$\boldsymbol{f}_* | \boldsymbol{X}_*, \boldsymbol{X}, \boldsymbol{f} \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_{**}) \tag{2.4}$$

where

$$\boldsymbol{\mu}_* = \boldsymbol{\mu} + \boldsymbol{K}_{\boldsymbol{x}_*\boldsymbol{x}} \boldsymbol{K}_{\boldsymbol{x}\boldsymbol{x}}^{-1} (\boldsymbol{f} - \boldsymbol{\mu}) \tag{2.5}$$

$$\boldsymbol{\Sigma}_{**} = \boldsymbol{K}_{x_*x_*} - \boldsymbol{K}_{\boldsymbol{x}_*\boldsymbol{x}} \boldsymbol{K}_{\boldsymbol{x}\boldsymbol{x}}^{-1} \boldsymbol{K}_{\boldsymbol{x}\boldsymbol{x}_*} \tag{2.6}$$

In Eq. (2.4), $\boldsymbol{\mu}_*$ and $\boldsymbol{\Sigma}_{**}$ denotes the posterior mean vector and covariance matrix of the target locations. In addition, $\boldsymbol{\mu}$ denotes the prior mean vector of the observed locations. The GPR's output and smoothness depend on the covariance kernel function, $k(r)$, where r denotes the distance between the input vectors. $k(r)$ is used to construct the covariance matrix, $\boldsymbol{K}$, in Eqs. (2.5) and (2.6). Tamhidi et al. (2021) demonstrated that Matérn with ν = 1.5 is the optimum covariance kernel function for the GPR model to simulate the ground motion time series conditioned on surrounding observed motions. Eqs. (2.7) and (2.8) illustrate the Matérn (ν = 1.5) kernel function and the distance between two input vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$.

$$k_{\nu=1.5}(r) = \sigma_f^2 (1 + \sqrt{3}r) exp(-\sqrt{3}r) \tag{2.7}$$

$$r = \theta \sqrt{\sum_{i=1}^{d} (x_i - x'_i)^2} \tag{2.8}$$

In Eq. (2.7) σ_f is the variance that governs how uncertain the GPR's estimate is for a given input location. In Eq. (2.8), d is the number of features existing in each input vector, x_i stands for the i^{th} feature of the input vector at location $\boldsymbol{x}$, and θ is a positive normalizing factor, also known as the inverse of length-scale, l, where $\theta = 1/l$. In this study, one single θ is used to normalize all

features of the input vectors (cf. Eq. 2.8). Such a covariance function is called an isotropic covariance function (Rasmussen and Williams 2006).

2.2. Conditioned Simulation of Ground Motion

It is required to specify the input vector of each site to be used by the GPR model as the first step of conditioned ground motion simulation. Tamhidi et al. (2021) demonstrated that for the regions with a considerable variation of site condition, a four-element input vector of $\boldsymbol{x} = \{x_1, x_2, x_3, \log(V_{s_{30}})\}$ is adequately informative about the features of the sites to estimate the ground motions. $V_{s_{30}}$ stands for the average shear wave velocity in the uppermost 30 m, and x_1 through x_3 are the Cartesian coordinates of the site on the 3D surface of the Earth.

The GPR model has parameters including distance normalizing factor, θ, the GP prior mean, μ (cf. Eq. 2.5), and variance σ_f (cf. Eq. 2.7) to be optimized. We have implemented the Maximum a Posteriori Estimates (MAPEs) using maximizing the penalized log-likelihood of the observations to optimize the parameters of the GPR model. Denoting the parameters as $\gamma = (\theta, \mu, \sigma_f)$, Eq. (2.9) illustrates the penalized log-likelihood of the observations (either $\mathcal{R}e_k$ or $\mathcal{I}m_k$ at each k^{th} frequency).

$$Q(\gamma) = -\frac{1}{2}(\boldsymbol{f}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{K_{xx}}^{-1}(\boldsymbol{f}-\boldsymbol{\mu}) - \frac{1}{2}\log|\boldsymbol{K_{xx}}| - \frac{n}{2}\log 2\pi - n\,d\,p_\lambda(\theta). \tag{2.9}$$

In Eq. (2.9), $\boldsymbol{f}$ denotes the GP observed values, T stands for the transpose operator, n is the number of observed sites and $p_\lambda(\theta)$ is a non-negative penalty function for normalizing factor θ. There are several established penalty functions such as the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996), Smoothly Clipped Absolute Deviation (SCAD) (Fan and Li 2001), and L2 penalty function. In this study, we used the L2 penalty function. The L2 penalty function in Eq. (2.9) is defined as

$$p_\lambda(\theta) = \lambda\theta^2. \tag{2.10}$$

The GPR model is completely defined once the optimum parameters, $\hat{\gamma}$, having penalized log-likelihood function (cf. Eq. 2.9) maximized.

Tamhidi et al. (2021) demonstrated how one could find the optimum regularization factor, $\hat{\lambda}$, through the 1906 San Francisco earthquake physics-based simulated ground motions (Aagaard et al., 2008b) using the SCAD penalty function for a fixed observation density value. They illustrated how to use Normalized Root Mean Square Error (NRMSE) of the response spectra between the estimated and exact ground motion as a criterion to find out the best $\hat{\lambda}$ implementing a five-fold Cross-Validation (CV) procedure. Tamhidi et al. (2021) state that a smaller number of observations

within a region (smaller density of observed motions) leads to a higher required regularization factor, which is consistent with Li and Sudjianto (2005). The fine-tuning of the hyper-parameter, λ, is discussed in more detail in Chapter 4.

3. Model Evaluation

We discussed that the GPR model has parameters $\gamma = (\theta, \mu, \sigma_f)$ dependent on the chosen regularization factor (hyperparameter) of the model, λ. One common way to find out the optimum hyper-parameter of the models is the Cross-Validation (CV) method (Shao 1993). Tamhidi et al. (2021) implemented the physics-based simulated ground motions for the 1906 M7.9 San Francisco earthquake (Aagaard et al., 2008b) to show how the λ could be optimized. They conducted a five-fold CV procedure to find out $\hat{\lambda}$ as the best regularization factor over the training set, which was a randomly chosen subset of the simulated ground motions. Tamhidi et al (2021) found that $\hat{\lambda} = 0.7$ is the optimum regularization factor using the SCAD penalty function and Matérn ($\nu = 1.5$) covariance function for the observation density corresponding to the training set used for the optimization procedure. It is worth noting that the derived $\hat{\lambda} = 0.7$ was obtained for both horizontal component directions, Fault Normal (FN) and Fault Parallel (FP). Interested readers are referred to Tamhidi et al. (2021) to have more detail on the training procedure and hyperparameter optimization of the GPR model.

Next, we aim to illustrate the model's performance in predicting the ground motions using the physics-based simulated motions.

3.1. The 1906 M7.9 San Francisco Simulated Motions

The 1906 **M**7.9 San Francisco physics-based simulated ground motions are generated at 40,700 sites on a uniform 1.5 km × 1.5 km grid. Tamhidi et al. (2021) chose two regions called Palo Alto and South Napa to train as well as test the GPR model. These regions are shown in Figure 3.1. The Palo Alto and South Napa regions have 104 and 111 sites, respectively. Around 80% of them are randomly chosen as training (observed) set, and the remaining 20% are set aside as test sites to predict their ground motions. The training set was already used to find out the optimum $\hat{\lambda} = 0.7$ for the GPR model using the SCAD penalty function and Matérn with $\nu = 1.5$. Figure 3.2 illustrates the distribution of the chosen training and test sites by Tamhidi et al. (2021).

The accuracy criterion used by Tamhidi et al. (2021) is the normalized root mean square error (NRMSE) between estimated (conditioned simulated using GPR) and exact (physics-based) motions' 5%-damped pseudo-spectral acceleration (PSA). The NRMSE is given by

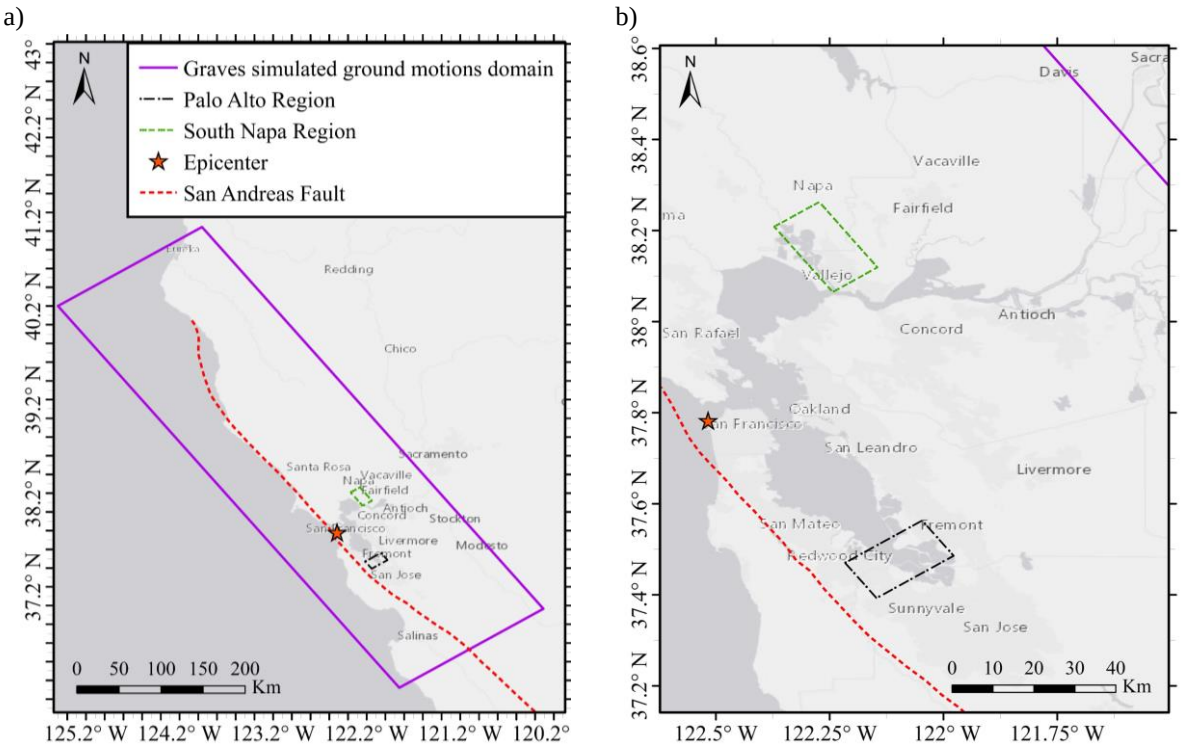


Figure 3.1. a) Aagaard et al. (2008b) 1906 **M**7.9 San Francisco earthquake simulated ground motions domain and b) Palo Alto and South Napa study regions for training and testing of the GPR model

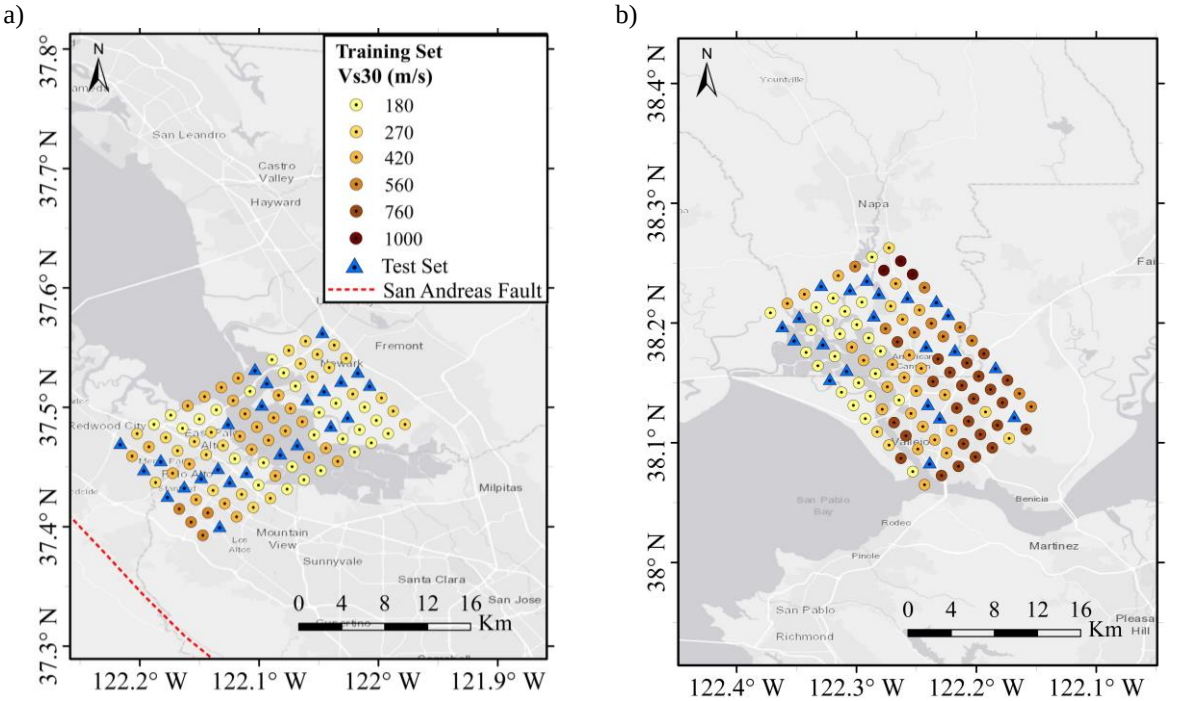


Figure 3.2. Distribution of the training and test sets for the a) Palo Alto and b) South Napa regions in 1906 **M**7.9 San Francisco Physics based sumulated earthquake

$$NRMSE = \sqrt{\frac{1}{\tau}\sum_{i=1}^{N_{period}} \frac{(PSA_i - \widehat{PSA}_i)^2}{\widehat{PSA}_i{}^2}} \tag{3.1}$$

where, N_{period} equals the number of periods included in the response spectra. The GPR model described in Section 2 using the $\hat{\lambda}$ = 0.7 is implemented to predict the ground motion time series at each test site for the Palo Alto and South Napa regions (Figure 3.2) using the corresponding observed sites. The following steps are taken to estimate the ground motion time series at each test site:

1. Given the observed ground motions (training set), the model parameters, $\hat{\gamma}$= $(\hat{\theta}, \hat{\mu}, \widehat{\sigma_f})$ are obtained at each frequency for the real and imaginary parts of the DFT coefficients using the corresponding $\hat{\lambda}$.
2. The posterior means (Equation 2.5) for the DFT coefficients at the test sites are obtained for each frequency using the values of $\hat{\gamma}$ from step 1.
3. The entire ground motion time series is constructed using Eqs. (2.1) and (2.2).

Figure 3.3 demonstrates the distribution of the NRMSE for the RotD50 spectrum between the estimated and exact ground motions at each test site location for both Palo Alto and South Napa study regions. In Figure 3.3, there are five chosen test sites within each of the Palo Alto and South Napa study regions to illustrate their estimated motions. Figures 3.4 and 3.5 demonstrate the predicted motions' RotD50 spectra, velocity time series, and Fourier Amplitude Spectrum (FAS) and the corresponding exact ones for the five chosen test sites within Palo Alto and South Napa study regions, respectively.

a) b)

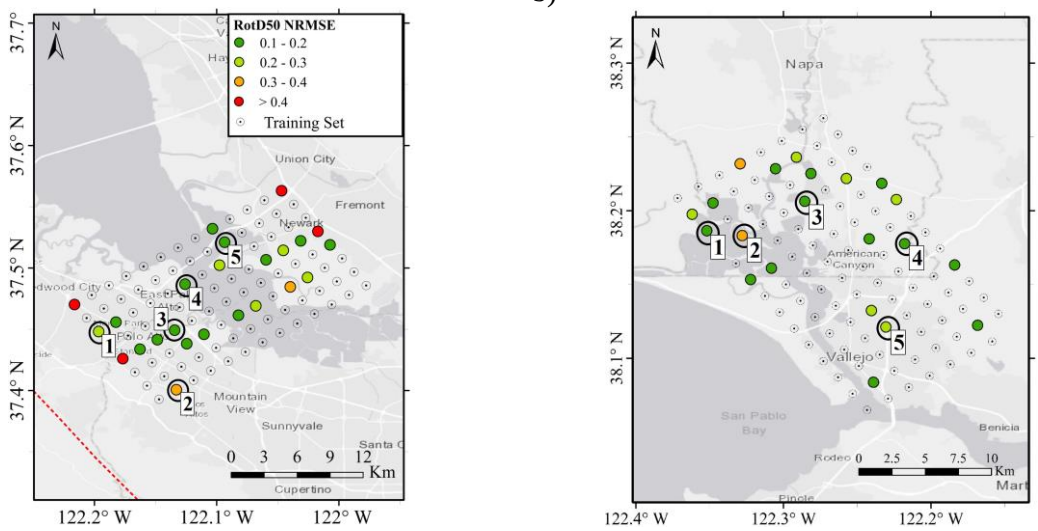


Figure 3.3. The distribution of the test set's NRMSE for the RotD50 spectrum at a) Palo Alto and b) South Napa study regions in 1906 **M**7.9 San Francisco Physics based sumulated earthquake

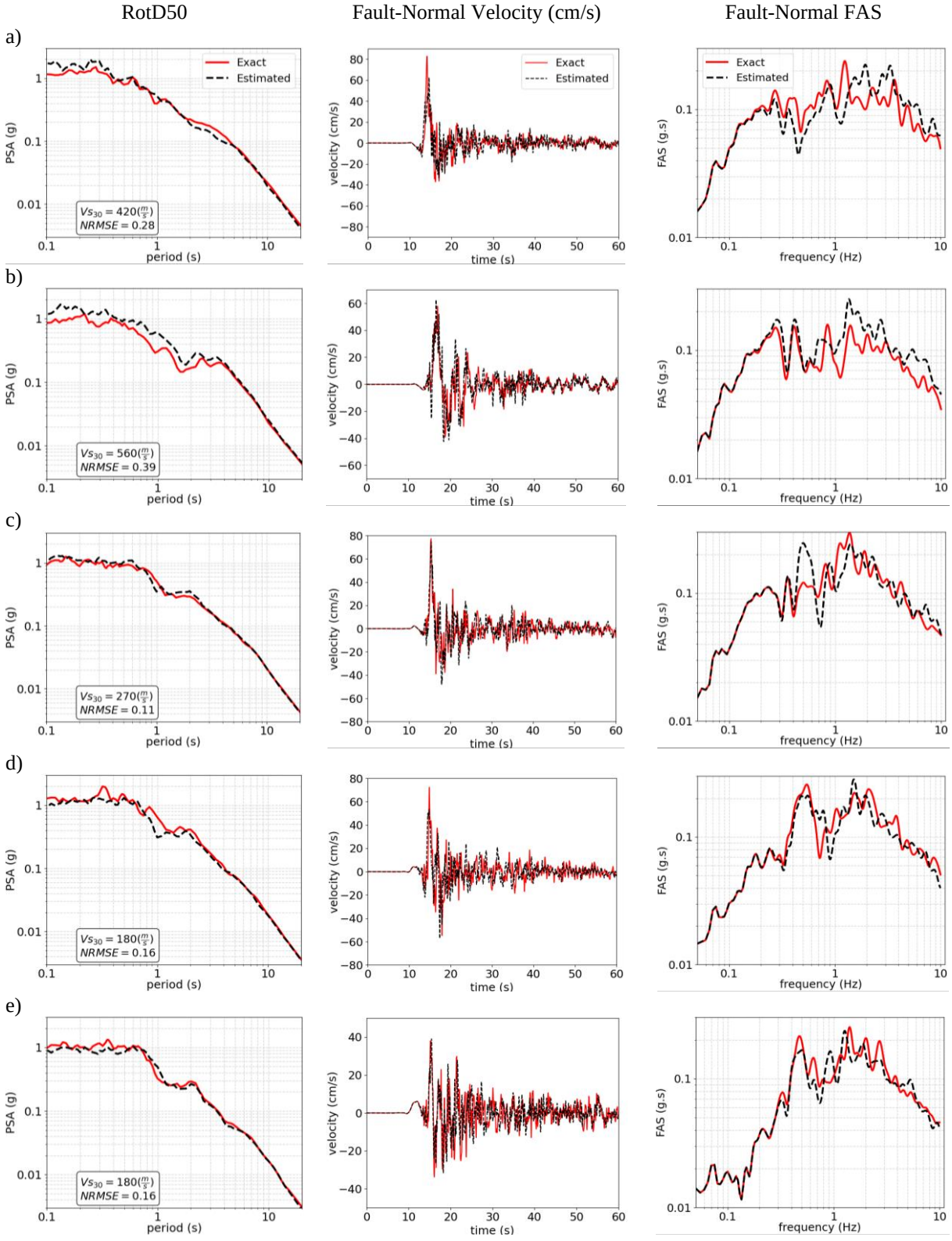


Figure 3.4. The RotD50, velocity time series, and FAS of the predicted as well as the exact motions along Fault-Normal direction for the chosen test sites: a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within Palo Alto study region in 1906 **M**7.9 San Francisco Physics based simulated earthquake

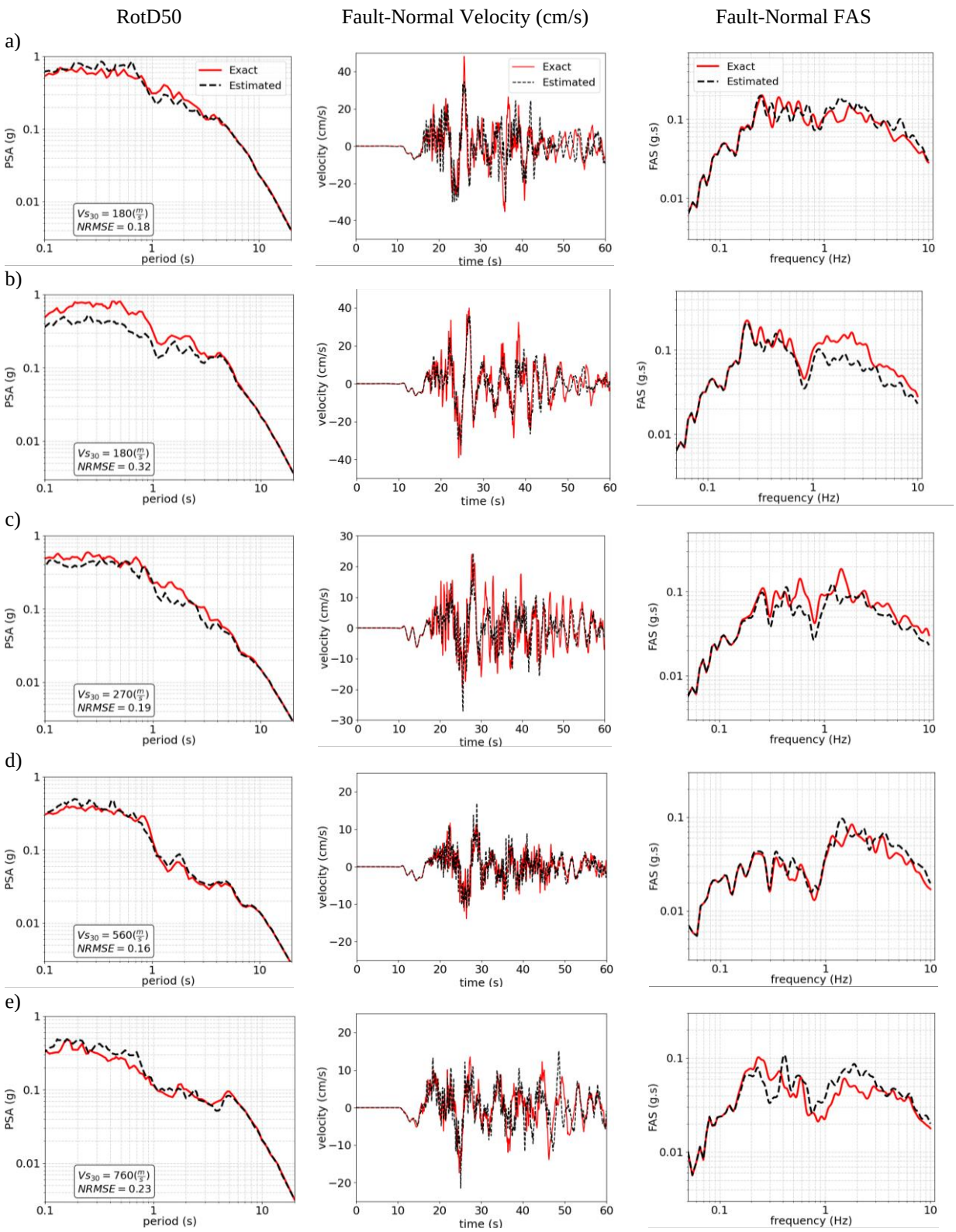


Figure 3.5. The RotD50, velocity time series, and FAS of the predicted as well as the exact motions along Fault-Normal direction for the chosen test sites: a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within South Napa study region in 1906 **M**7.9 San Francisco Physics based simulated earthquake

In addition, Table 3.1 summarizes the average test set's NRMSE for response spectra along FN and FP directions and the RotD50 spectrum for both study regions.

Table 3.1. 1906 **M**7.9 San Francisco test set's NRMSE for response spectra along FN and FP directions as well as the RotD50 spectrum (Tamhidi et al., 2021)

Study Region	FN		FP		RotD50	
	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
Palo Alto	0.34	0.29	0.38	0.38	0.31	0.36
South Napa	0.23	0.06	0.26	0.1	0.19	0.05

It is observable from Figures 3.4 and 3.5 that the trained GPR was able to estimate the ground motion time series decently and accurately compared to the exact ones at most of the test sites. There are four sites in total that their prediction was not accurate, and those are constrained to the boundaries of the observations' network where there is a less uniform distribution of the observed sites. It is also shown that the GPR model can predict the long period pulses due to the directivity effect reasonably well (Figure 3.4). By comparing the results of Table 3.1, it is demonstrated that the prediction accuracy of the GPR model for the regions far away from the fault might be higher than those very close to the fault as the GPR employs an isotropic covariance function which allocates a uniform correlation to the neighboring observations around the target site. The prediction accuracy of the GPR could be improved by deploying an anisotropic covariance model for the regions closer to the fault (Tamhidi et al., 2021; Rasmussen and Williams 2006).

3.2. M7.0 Hayward Fault Scenario Earthquake Simulated Motions

Tamhidi et al. (2021) tested the performance of the GPR model on another physics-based simulated dataset that was not present during the training procedure. They used **M**7.0 Hayward fault scenario earthquake simulated ground motions (Rodgers et al., 2019), considering the 3D topographic features of the Earth's surface. These motions are simulated at 2301 stations on a uniform 2 km × 2km grid. These motions are generated using $V_{S_{min}} = 500$ m/s; therefore, they chose 326 sites the $V_{S_{30}}$ of which are equal or greater than 500 m/s based on the USGS 2018 model (USGS, 2018). The $V_{S_{30}}$ of these selected 326 sites are between 500 m/s and 520 m/s. About 80% of these 326 stations are randomly chosen as training (observed) set, and the rest 20% are considered as the test set. Figure 3.6 demonstrates the distribution of the training set as well as NRMSE between the predicted (conditioned simulated) and exact (physics-based simulated) motions' 5%-damped linear RotD50 response spectra at the test set locations. In Figure 3.6, there are five chosen test sites that we illustrated the predicted as well as exact motions' velocity time series and FAS along FN and the RotD50 spectra in Figure 3.7.

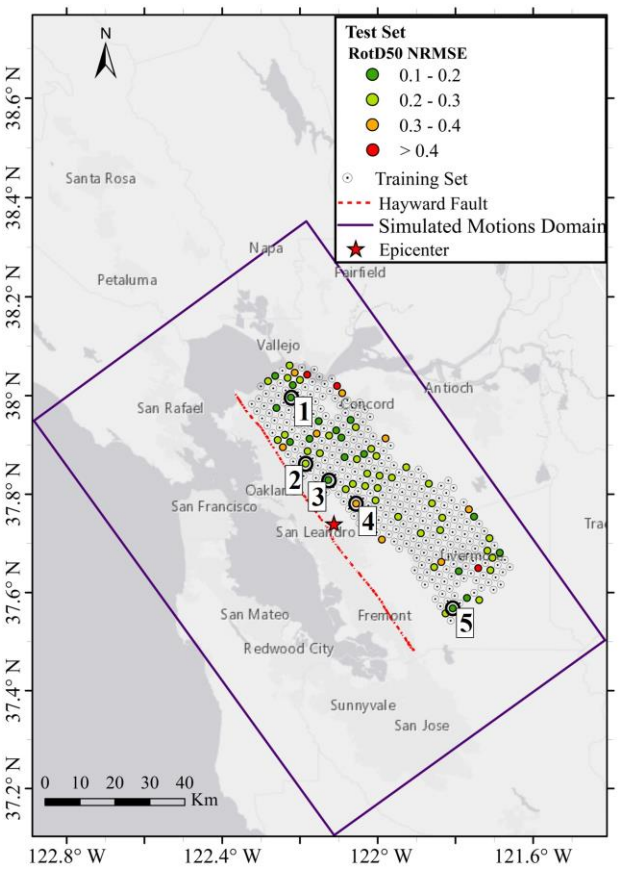


Figure 3.6. Distribution of the training and test set's NRMSE for the 5% damped RotD50 spectrum for the **M**7.0 Hayward fault scenario earthquake simulated motions

Figure 3.7 depicts the prediction results for the five test sites selected in Figure 3.6. As shown in Figure 3.7, the RotD50 spectra of the predicted motions are quite close to those of the physics-based simulated motions. Furthermore, as shown in Figures 3.7a through Figure 3.7c, the long period pulses of the ground motion (or Peak Ground Velocity, PGV) are reasonably accurately estimated. Moreover, Figure 3.6 depicts that most of the test sites' ground motions are correctly predicted, while a few sites' predictions were a little more erroneous, indicating that most of them are limited to the network's edge, where there is not a uniform observation surrounding them. The average NRMSE among all stations for predicted motions' response spectra along FN, FP, and RotD50 is 0.28, 0.31, and 0.25, respectively. The test results from both 1906 **M**7.9 San Francisco and **M**7.0 Hayward fault earthquakes demonstrated the applicability of the GPR model to simulate ground motions. However, this training was done just for one specific observation density, and thus, there is a requirement to fine-tune the hyperparameter of the model, λ, for various observation densities as Li and Sudjianto (2005) states the dependency of the penalty function regularization factor on the density of the observation. In the following chapter, we used the ground motions recorded by the Community Seismic Network (CSN) during the 2019 M7.1 Ridgecrest earthquake to fine-tune the regularization factor for various observation densities.

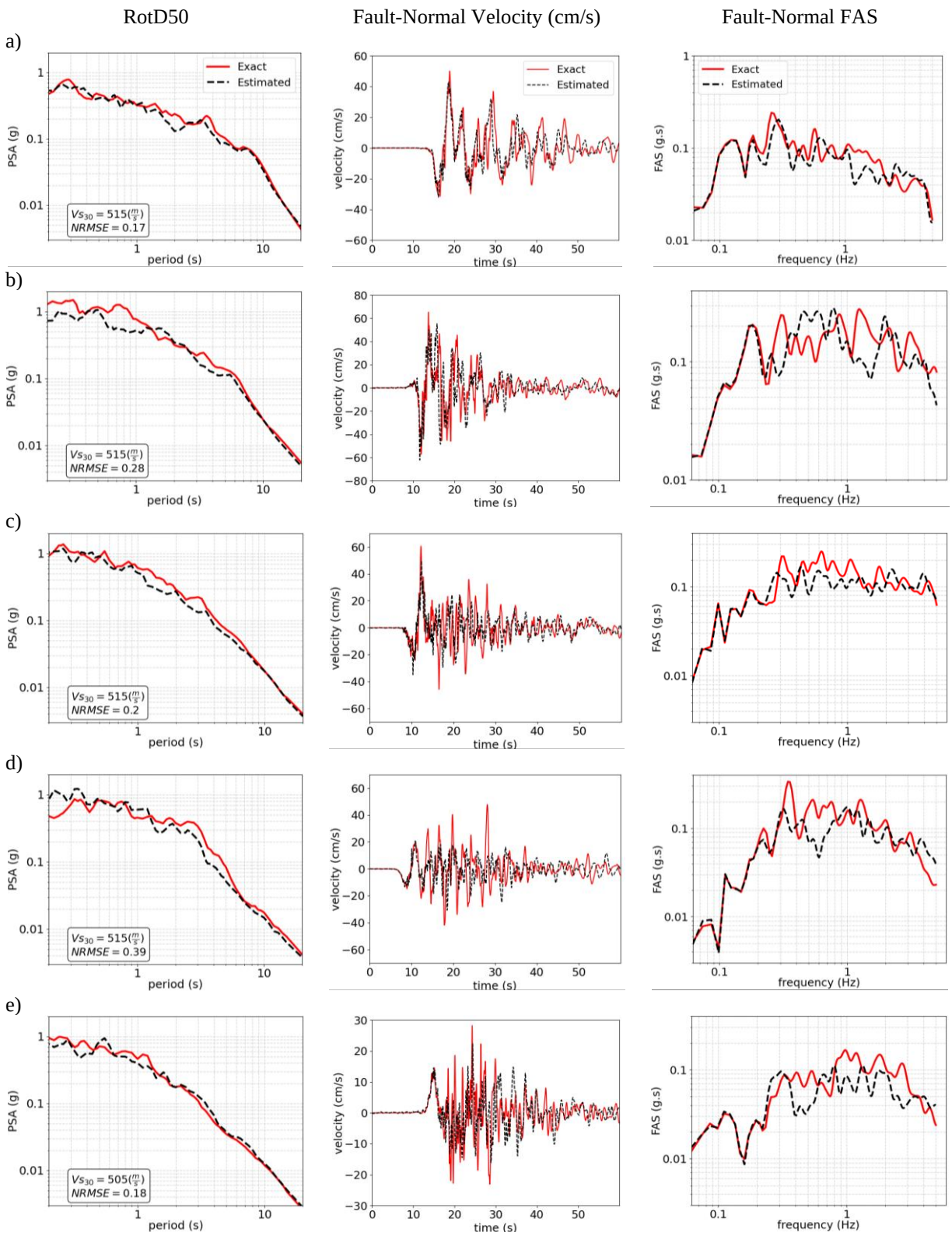


Figure 3.7. The RotD50, velocity time series, and FAS of the predicted as well as the exact motions along Fault-Normal direction for the chosen test sites: a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within **M**7.0 Hayward fault scenario earthquake simulated motion dataset

4. Optimum Regularization Factor Based on Density of Observations

The derived optimum parameters of the GPR, $\hat{\gamma} = (\hat{\theta}, \hat{\mu}, \hat{\sigma_f})$, depend on the assumed penalty function and the regularization factor (cf. Eq. 2.9). In other words, the regularization factor, λ, is the hyperparameter of the model that controls how observations lead to the optimum $\hat{\gamma}$. As Li and Sudjianto (2005) stated, the sparsity of the observations leads to a higher variance in the vicinity of the optimum solution in the log-likelihood function of the observations, which needs an introduced bias (penalty) to reduce the variance. Therefore, it is expected to need a higher penalty (or λ) for the log-likelihood function of the observations when there is a smaller number of observed sites within a region (Tamhidi et al., 2021). As a result, there would be different optimum regularization factor, $\hat{\lambda}$, for various observation densities. Therefore, it is required to tune the $\hat{\lambda}$ as a function of observation density.

We used CSN stations in Los Angeles (Clayton et al. 2020) to investigate the effect of observation density on $\hat{\lambda}$. We used the recorded earthquake ground motions during the 2019 **M**7.1 Ridgecrest earthquake for that purpose.

4.1. M7.1 Ridgecrest Earthquake

The 2019 **M**7.1 Ridgecrest earthquake happened on July 6th, 2019, in Searles Valley, 17.9 km Northeast of Ridgecrest, California. We implemented the 252 ground-level recording sites motions of the CSN in this study. Table 4.1 summarizes the 2019 **M**7.1 Ridgecrest earthquake features.

Table 4.1. The 2019 **M**7.1 Ridgecrest earthquake features (USGS, 2019) recorded by CSN

Date	UTC time	M_w	Epicenter	Depth	No. of stations	Network Area (km^2)
July 6th, 2019	03:19:53	7.1	Searls Valley	8.0 km	252	464

The observation density of these recorded motions is calculated by dividing the number of observed ground motions over the area of the enclosing network. Figure 4.1 shows the 252 CSN sites that recorded the **M**7.1 Ridgecrest earthquake and the bounding region of the network with an area of 464 km^2. This means that the observation density within Los Angeles is approximately 0.54 sites/ km^2.

The L2 penalty function was utilized for the penalized log-likelihood function (cf. Eq. 2.9). We created several datasets with various observation density by randomly selecting 252 CSN sites to evaluate the optimum $\hat{\lambda}$ for various observation densities. We made datasets with 252, 201, 151,

100, 50, and 25 recorded ground motions, respectively. The distribution of the randomly chosen sites for each dataset is shown in Figure 4.2. It is worth noting that the site condition, $V_{s_{30}}$, of the recording stations is predicted using a proxy-based model, as explained in Ahdi et al. (2020).

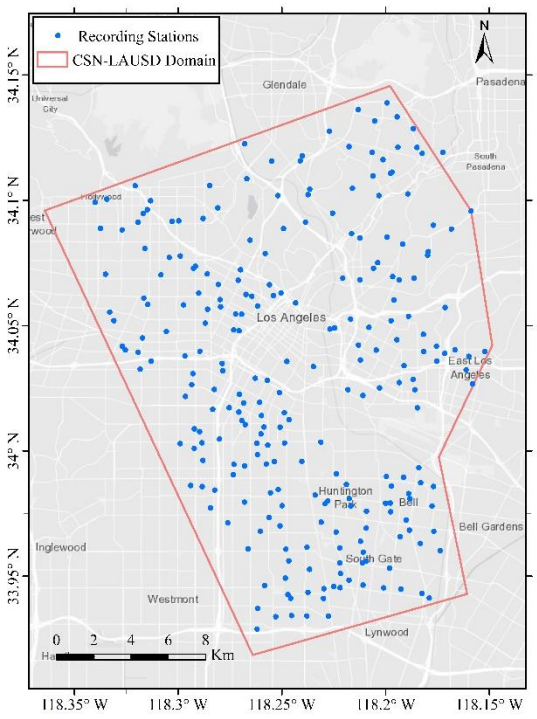


Figure 4.1. Distribution of the 252 CSN-LAUSD sites that recorded 2019 **M**7.1 Ridgecrest earthquake

The selection criterion for finding $\hat{\lambda}$ for each dataset is the normalized root mean square error (NRMSE) between the exact (recorded) and the conditioned simulated (estimated) ground motions' 5%-damped RotD50 (Boore, 2010) response spectra.

It is worth noting that the usable period bandwidth of the predicted motions is the mutual usable bandwidth (Ancheta et al., 2014) among all observed motions, which is the reliable period range after the noise removal of the observed motions.

In order to find the $\hat{\lambda}$ for each dataset shown in Figure 4.2, we implemented Leave One Out (LOO) cross-validation method (Vehtari et al., 2017). The following steps are taken for each λ_{test} to be evaluated for each dataset (N_{sites} = number of recording sites):

1. For each individual site, *s*, within the dataset; $s = 1, \ldots, N_{sites}$
   1.1. Obtain the optimum parameters $\hat{\gamma}$ for the observed motions, which are all the recorded ground motions except the motion recorded at site *s*, using λ_{test} and maximizing $Q(\boldsymbol{\gamma})$ in Eq. (2.9).

1.2. Estimate the ground motion time series at the station *s* using posterior mean (Eq. 2.5) for both real and imaginary parts of the DFT coefficients, employing $\hat{\gamma}$ derived in step 1.1.

1.3. Obtain the RotD50 spectrum of the estimated and exact (recorded) ground motions at station *s* and calculate the NRMSE between them within the usable bandwidth. Store this NRMSE as $Error_s$.

2. Take the average of $Error_s$ among all sites within the dataset ($s = 1, \ldots, N_{sites}$), i.e., $Error_{avg}$ and store it as the corresponding error to the λ_{test}.

Eventually, we chose the λ_{test} with the lowest $Error_{avg}$ as the $\hat{\lambda}$ for the corresponding dataset. Table 4.1 illustrates the obtained $\hat{\lambda}$ for each of the datasets in Figure 4.2.

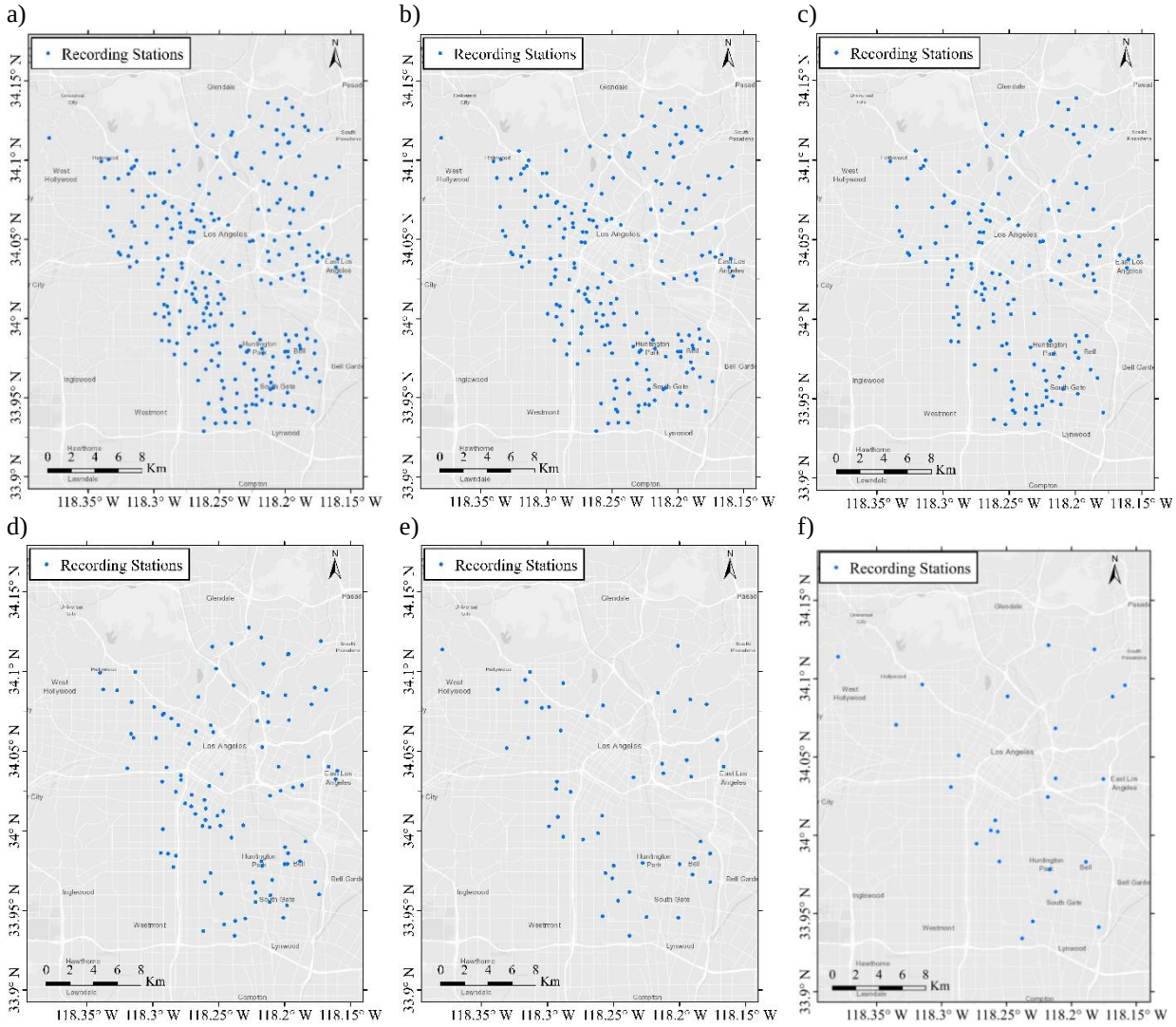


Figure 4.2. Distribution of the randomly chosen datasets of CSN recorded **M**7.1 Ridgecrest earthquake motions with a) 252, b) 201, c) 151, d) 100, e) 50, and f) 25 number of sites

Table 4.2. The $\hat{\lambda}$ obtained for different observation density and the $Error_{avg}$ corresponding to that

No. of Observation	Density (station/km^2)	$\hat{\lambda}$	Average RotD50 NRMSE ($Error_{avg}$)
251	0.54	0.05	0.27
200	0.43	0.1	0.28
150	0.32	0.1	0.27
99	0.21	0.1	0.31
49	0.10	0.2	0.30
24	0.05	0.4	0.40

We have employed the corresponding $\hat{\lambda}$ for each of the datasets shown in Figure 4.2 to estimate the ground motion time series at each station of that dataset, considering all the rest of the stations' motions as observation (i.e., LOO analysis). Figure 4.3 demonstrates the distribution of the NRMSE between the exact (recorded) and estimated motions' RotD50 spectrum.

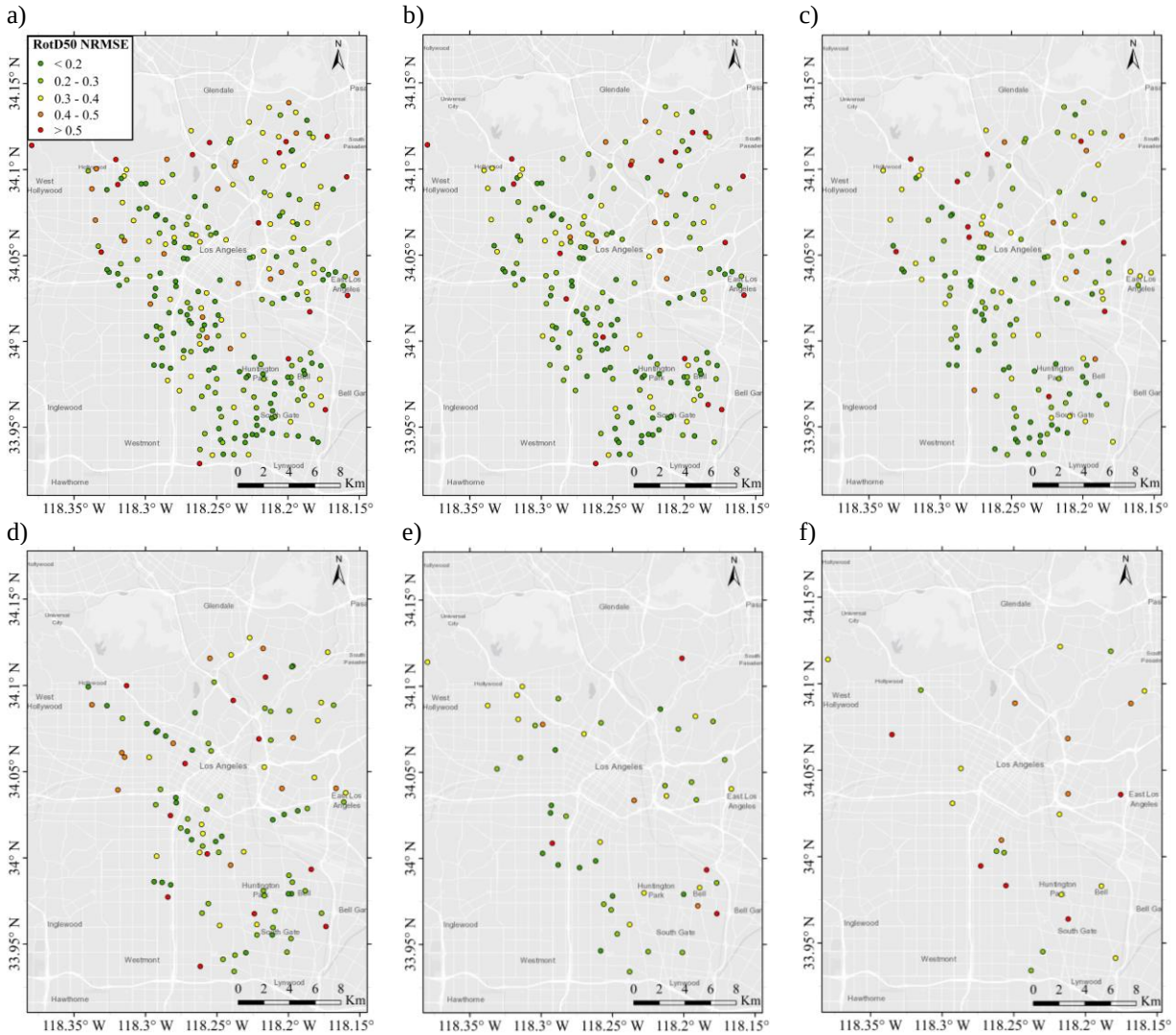


Figure 4.3. Distribution of the RotD50 NRMSE between the recorded and estimated ground motions for stations of CSN which recorded **M**7.1 Ridgecrest earthquake within the datasets having a) 252, b) 201, c) 151, d) 100, e) 50, and f) 25 number of sites

As discussed above, Table 4.2 demonstrates that the required $\hat{\lambda}$ to have a reliable ground motion estimation increased by decreasing the number of observations within a region (lower observation density). Figure 4.3 indicates that the estimated ground motions at each station using the rest stations' motions as observation are reliably accurate for most of the stations in each dataset. However, Figure 4.3 shows that the number of stations with higher RotD50 NRMSE (yellow, orange, and red stations) increases by decreasing the number of observed sites. This is also illustrated in Table 4.2 that the average RotD50 NRMSE among all stations increases for having a lower number of observations.

Table 4.3 illustrates the average and standard deviation of the $Error_{avg}$ between the exact and recorded motions' response spectra along East-West (EW), North-South (NS) directions and RotD50 for each dataset shown in Figure 4.3.

Table 4.3 . Average and Standard Deviation of NRMSE between recorded and estimated motions' response spectra for each dataset shown in Figure 4.3

No. of Sites	$\hat{\lambda}$	EW		NS		RotD50	
		Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
252	0.05	0.31	0.15	0.30	0.14	0.27	0.14
201	0.1	0.33	0.21	0.32	0.18	0.28	0.18
151	0.1	0.32	0.15	0.31	0.16	0.27	0.14
100	0.1	0.36	0.18	0.35	0.18	0.31	0.18
50	0.2	0.34	0.14	0.34	0.15	0.30	0.14
25	0.4	0.44	0.19	0.46	0.23	0.40	0.16

Table 4.3 shows that, in general, the average response spectra NRMSE for both EW and NS directions is increased by having a lower number of observations, which is expected as the higher number of observations provides more input information about the spatial variation of the amplitude within a region.

In summary, it is concluded that the optimized regularization factor, $\hat{\lambda}$, should be chosen based on the observation density within the target network. It is observed that the required $\hat{\lambda}$ and subsequently penalty function value, $p_{\lambda}(\theta)$, in Eq. (2.9) is increased for having a lower density of observations. In addition, the expected average error of prediction for the smaller number of observations is higher.

5. Random Realization of Generated Ground Motions

The ground motions predicted in Chapter 3 were constructed using the mean estimated DFT coefficients' real and imaginary parts, $\mathcal{R}e_k$ and $\mathcal{I}m_k$, at each k^{th} frequency, k = 0, …, N-1. One can have a posterior mean vector, and the posterior covariance matrix of the real and imaginary parts at each frequency for all target sites having the GPR estimated values (cf. Eqs. 2.5 and 2.6). At each prediction step in this study, we create the ground motion time series at only one target site. As a result, Eqs. (2.5) and (2.6) provide posterior mean and posterior standard deviation for the real and imaginary parts at each frequency (cf. Eqs. 5.1 and 5.2).

5.1. Ground Motion Random Sample Generation Methodology

Suppose we aim to estimate the uncertainty of the ground motion constructed using the posterior mean DFT coefficients' real and imaginary parts at site s. Using the derived $\hat{\gamma} = (\hat{\theta}, \hat{\mu}, \hat{\sigma_f})$ at each k^{th} frequency, the posterior mean and standard deviation are given by

$$\mu_* = \hat{\mu} + \boldsymbol{K}_{\boldsymbol{x}_*\boldsymbol{x}}\boldsymbol{K}_{\boldsymbol{xx}}{}^{-1}(\boldsymbol{f} - \hat{\boldsymbol{\mu}}), \text{ and} \tag{5.1}$$

$$\sigma_* = \boldsymbol{K}_{x_*x_*} - \boldsymbol{K}_{\boldsymbol{x}_*\boldsymbol{x}}\boldsymbol{K}_{\boldsymbol{xx}}{}^{-1}\boldsymbol{K}_{\boldsymbol{xx}_*}. \tag{5.2}$$

In Eqs. (5.1) and (5.2), μ_* and σ_* are the posterior mean and standard deviation of estimated $\mathcal{R}e_k$ (or $\mathcal{I}m_k$) and $\boldsymbol{f}$ is a vector of the observed $\mathcal{R}e'_k$ (or $\mathcal{I}m'_k$) at the neighboring stations, s'. The $\hat{\mu}$ is the optimized prior mean of $\mathcal{R}e_k$ (or $\mathcal{I}m_k$). We used the correlation between the real and imaginary parts of the observed ground motions DFT coefficients at the k^{th} frequency to estimate the correlation between the real and imaginary parts at the target site. Then, the 2×2 covariance matrix for pair of ($\mathcal{R}e_k$, $\mathcal{I}m_k$) can be established using the estimated correlation matrix and standard deviation (Eq. 5.2). It is possible to generate random samples of 2×1 ($\mathcal{R}e_k$, $\mathcal{I}m_k$) vectors having the 2×1 mean vector of $\mathcal{R}e_k$ and $\mathcal{I}m_k$ provided by Eq. (5.1) and constructed 2×2 covariance matrix. These generated random samples of 2×1 ($\mathcal{R}e_k$, $\mathcal{I}m_k$) vectors can then be converted to amplitude samples, $|A_k|$ using Eq. (2.2). The logarithmic mean and standard deviation of generated random samples of $|A_k|$ are then calculated. We implemented the inter-frequency correlation model developed by Bayless and Abrahamson (2019) to estimate the correlation between $\log(|A_k|)$ and $\log(|A_j|)$ at the k^{th} and j^{th} frequencies, where $k \neq j$. Eventually, a N×N covariance matrix of $\log(|A_k|)$ (k = 0, …, N-1) using the estimated standard deviation and obtained correlation is constructed. The combination of N×N covariance matrix and N×1 estimated mean vector of $\log(|A_k|)$ enabled us to generate N×1 random vectors of $\log(|A_k|)$, which can be converted to random samples of N×1 Fourier Amplitude Spectra (FAS) vectors. The generated

samples of FAS are then combined with the Phase spectrum constructed using the posterior mean of $\mathcal{R}e_k$ and $\mathcal{I}m_k$ at each frequency to generate the ground motion time series realizations at the target site.

5.2. Ground Motion Realizations for M7.0 Hayward Fault Scenario Earthquake

We conducted the described generation of ground motion realizations method on the **M**7.0 Hayward fault scenario earthquake generated by Rodgers et al. (2019). In addition, we have used the test set by Tamhidi et al. (2021) to evaluate the trained GPR model for the **M**7.0 Hayward fault scenario earthquake dataset. Figure 5.1 illustrates three chosen test sites for each of which we generated 100 ground motion realizations.

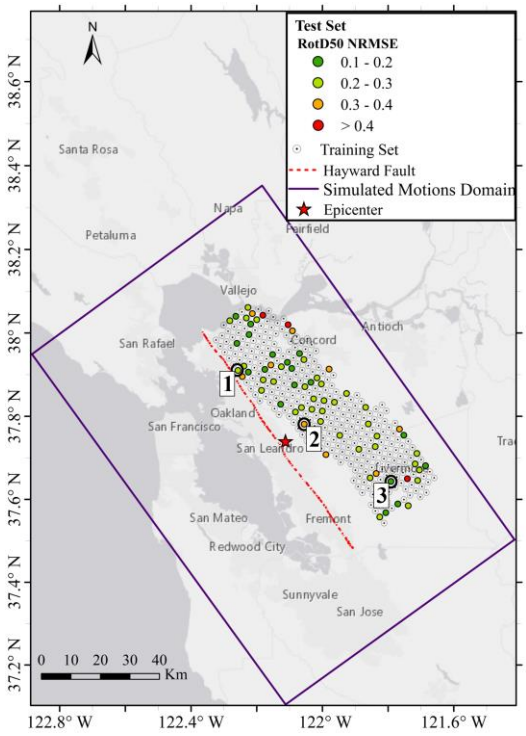


Figure 5.1. Distribution of the training and the chosen test sites for random ground motion realization purpose for the **M**7.0 Hayward fault scenario earthquake simulated motions.

Figure 5.2a illustrates the 5%-damped PSA along the FN direction for the exact (physics-based simulated), mean estimated, and one hundred ground motion realizations for the target site 1. Figure 5.2.b demonstrates the exact, mean estimated, and ground motion realizations' FAS along FN direction at the same target site. The 68 percent confidence interval (mean $\pm$ standard deviation) of PSA and FAS along FN direction at the target site 1 are shown in Figures 5.2c and 5.2d, respectively. Figures 5.2c and 5.2d depict that the uncertainty associated with estimated

motions is higher at short periods (higher frequencies) rather than at longer periods. This phenomenon could be due to having observed motions at distances that is more informative to generate long period (longer wavelengths) content of the motions. On the other hand, the number of observed motions with the current observation density (2 km separation distance) is insufficient to estimate the higher frequency content of the target sites deterministically. Figure 5.3. illustrates five generated ground motion realizations' velocity time series along FN direction for the target site 1. In general, the longer period content of the motions is estimated more confident and more accurate.

Figure 5.4 through Figure 5.7 illustrate the one hundred random ground motion realizations for target sites 2 and 3, respectively. Figure 5.4 through Figure 5.6 demonstrate that the variation of the estimated content at the shorter periods (short wavelengths) is higher with respect to those at longer periods.

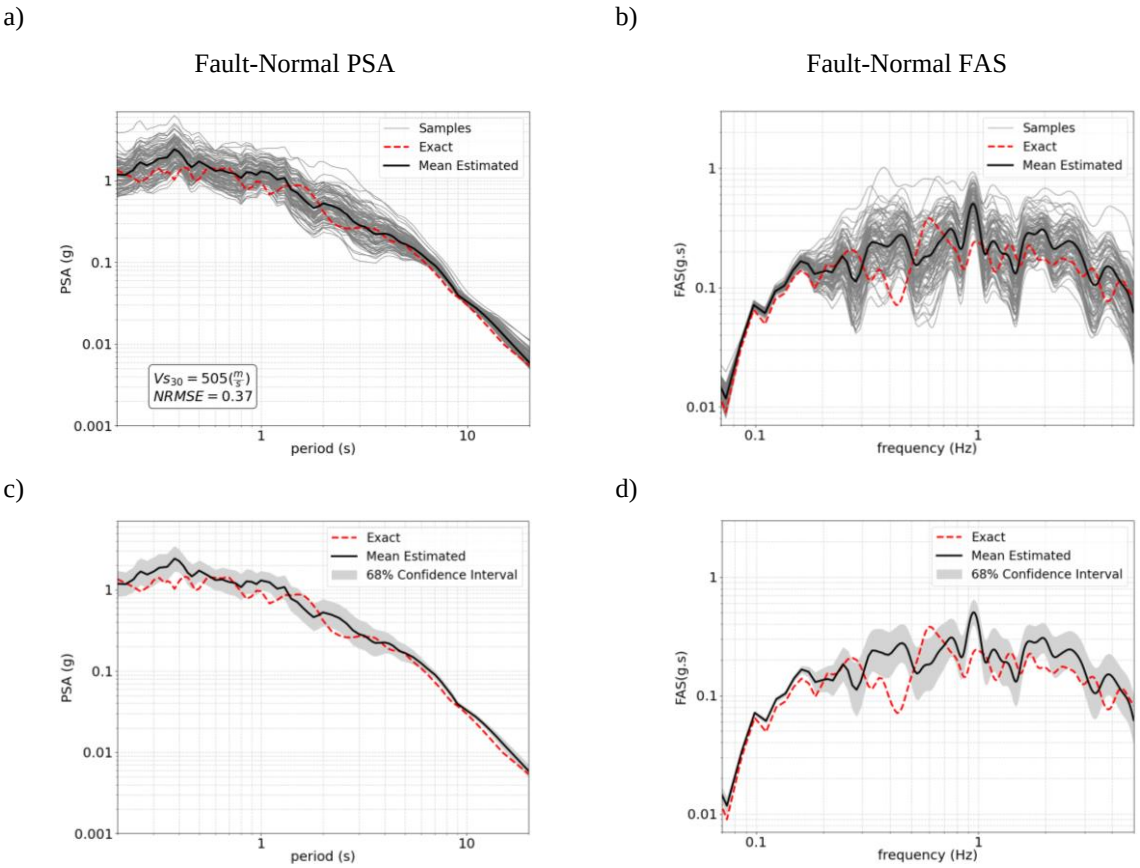


Figure 5.2. a) 5%-damped pseudo spectral acceleration (PSA) samples, b) Fourier Amplitude Spectrum (FAS) samples, c) 68% confidence interval of PSA, and d) 68% confidence interval of FAS along Fault-Normal direction at test site 1 within the **M**7.0 Hayward Fault scenario earthquake simulated motions

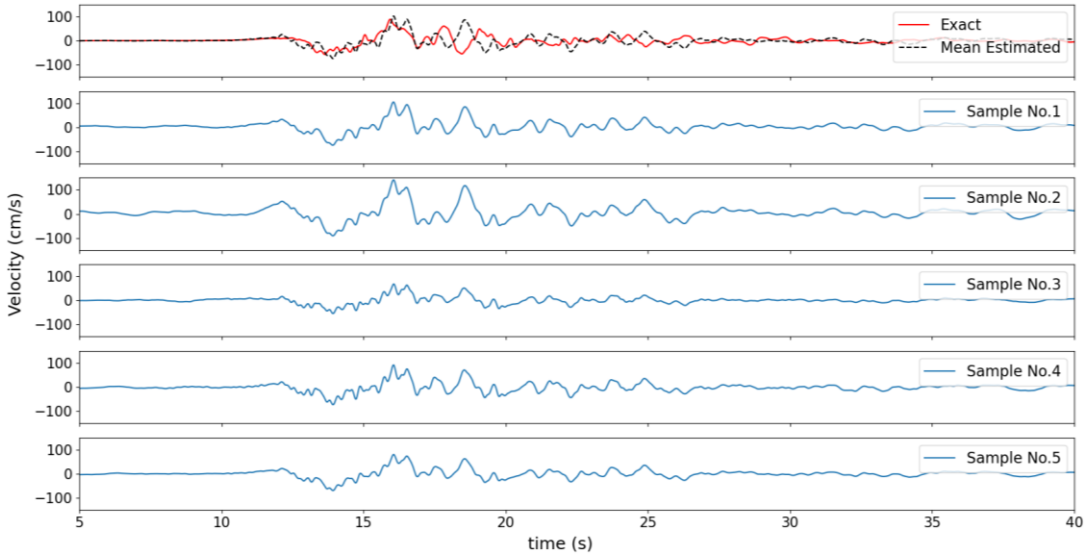
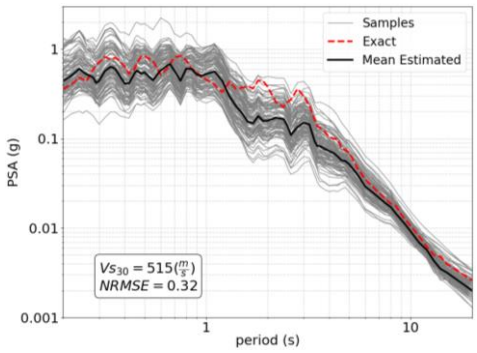


Figure 5.3. Five random generated samples of ground velocity time series along Fault-Normal direction at test site 1 within the **M**7.0 Hayward Fault scenario earthquake simulated motions

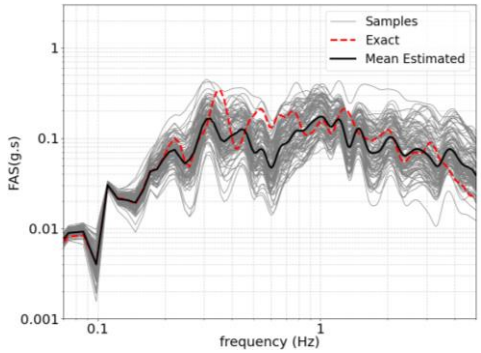
a)

Fault-Normal PSA

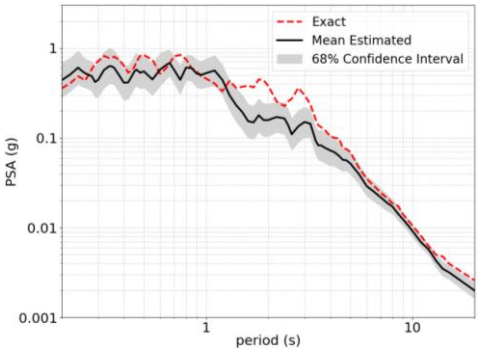


b)

Fault-Normal FAS



c)



d)

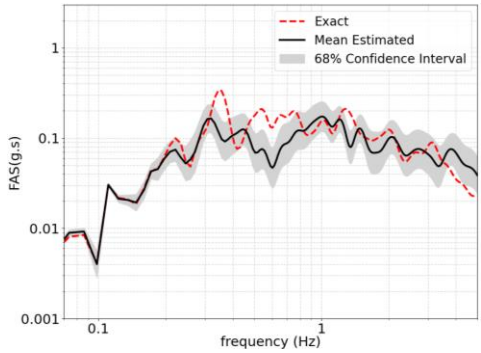


Figure 5.4. a) 5%-damped pseudo spectral acceleration (PSA) samples, b) Fourier Amplitude Spectrum (FAS) samples, c) 68% confidence interval of PSA, and d) 68% confidence interval of FAS along Fault-Normal direction at test site 2 within the **M**7.0 Hayward Fault scenario earthquake simulated motions

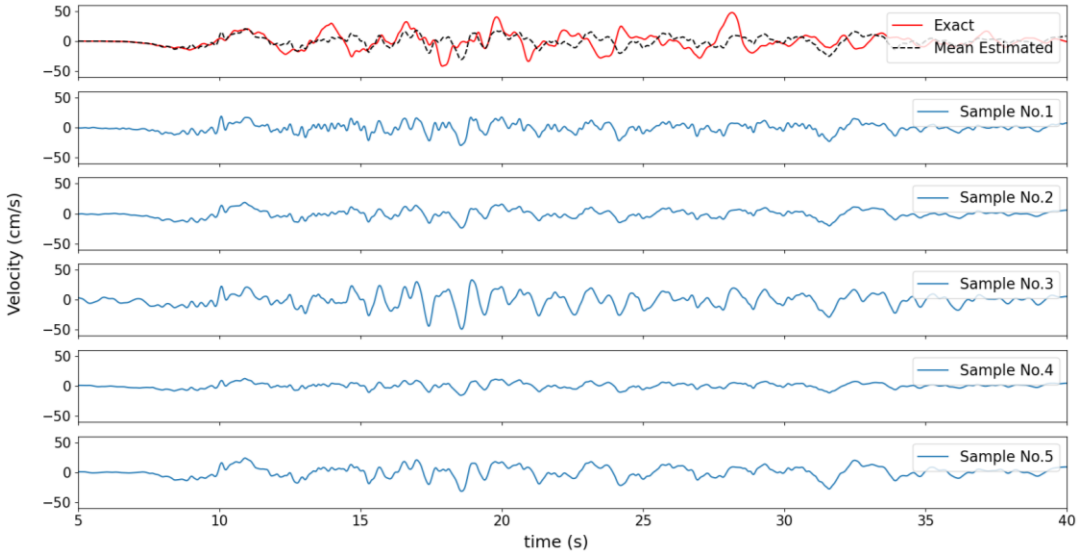


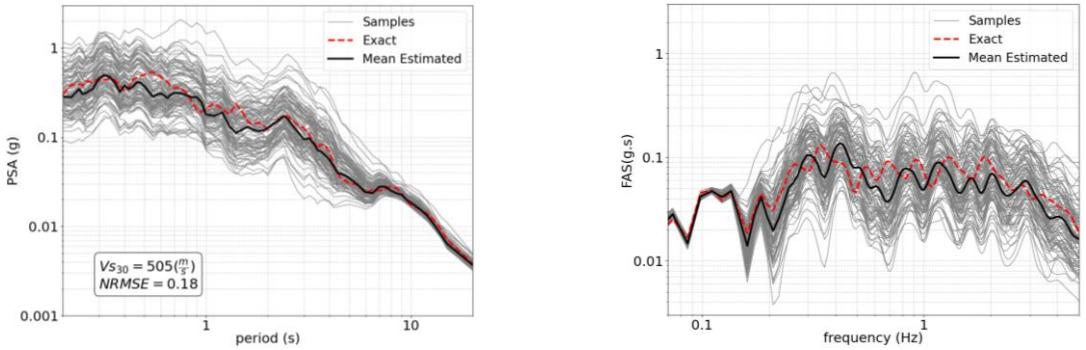
Figure 5.5. Random generated samples of ground velocity time series along Fault-Normal direction at test site 2 within the **M**7.0 Hayward Fault scenario earthquake simulated motions

a)

Fault-Normal PSA

b)

Fault-Normal FAS



c)

d)

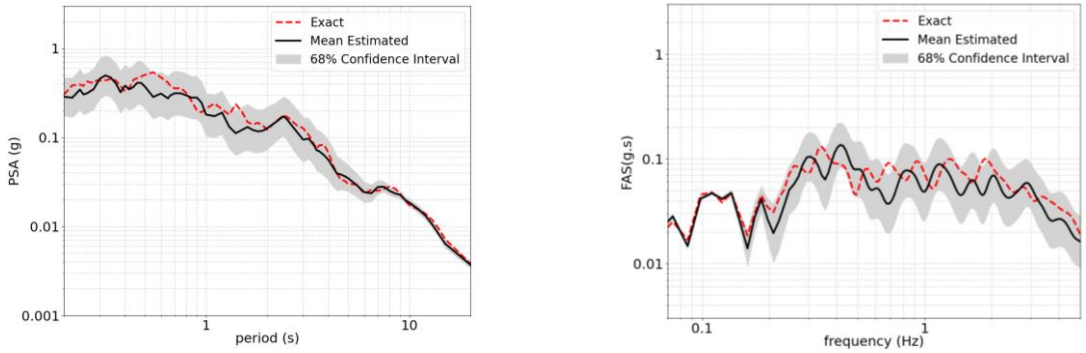


Figure 5.6. a) 5%-damped pseudo spectral acceleration (PSA) samples, b) Fourier Amplitude Spectrum (FAS) samples, c) 68% confidence interval of PSA, and d) 68% confidence interval of FAS along Fault-Normal direction at test site 3 within the **M**7.0 Hayward Fault scenario earthquake simulated motions

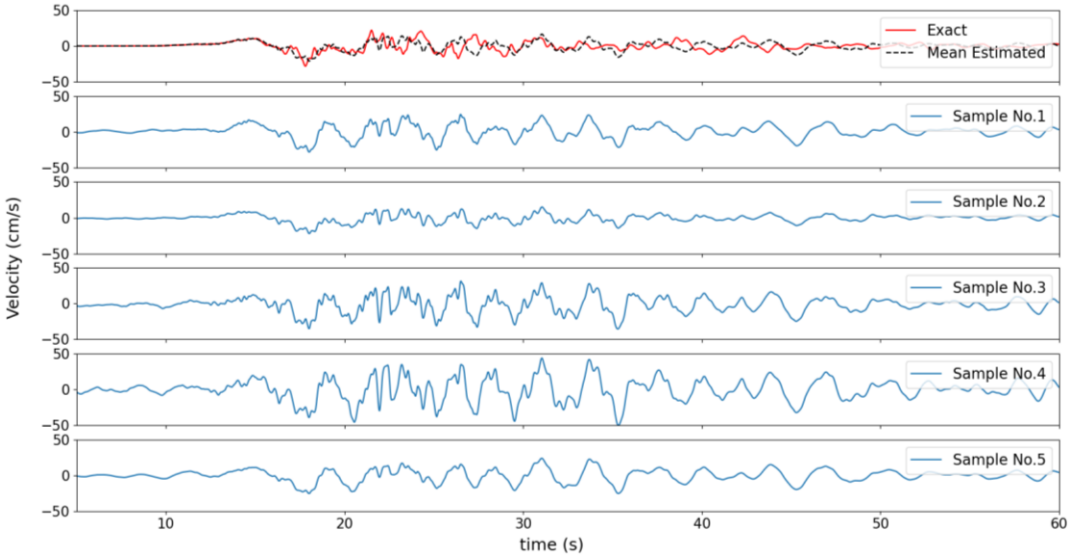


Figure 5.7. Random generated samples of ground velocity time series along Fault-Normal direction at test site 3 within the **M**7.0 Hayward Fault scenario earthquake simulated motions

5.3. Ground Motion Realizations for M7.1 Ridgecrest Earthquake

It is required to evaluate the established ground motion realization method on another real recorded ground motion dataset. Therefore, we conducted the random generation of ground motion method described in Section 5.1 to 2019 **M**7.1 Ridgecrest earthquake ground motions recorded by Community Seismic Network (CSN) in Southern California, Los Angeles (Clayton et al. 2020; Kohler et al. 2020). The ground motion realization generation method is conducted on the two test sites within CSN ground-level recording stations (Figure 5.8) using 149 observed sites.

The two test sites were chosen to generate the mean estimated ground motion and 100 ground motion realizations. Figure 5.8. illustrates the distribution of the observed sites and two selected test sites.

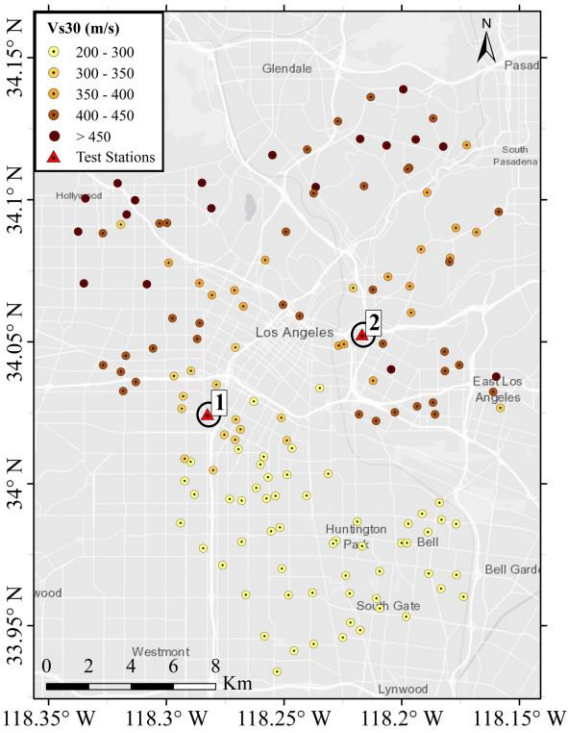


Figure 5.8. Distribution of the training and test set for the CSN stations recorded 2019 **M**7.1 Ridgecrest earthquake for random generation of ground motion realizations

It is worth noting that the optimized regularization factor, $\hat{\lambda}$, specified in Table 4.1 is used to construct the mean estimated ground motion as well as ground motion realizations at the location of two test sites shown in Figure 5.8.

Figure 5.9 illustrates the 5%-damped PSA and FAS along EW direction for the exact (recorded), mean estimated, and one hundred samples of ground motion realizations for the target site 1 within the usable frequency bandwidth. It is observable in Figure 5.9 that estimated motions have more variation (higher uncertainty) at shorter period content (mostly shorter than 1s). In addition, Figure 5.9 illustrates that the recorded motion's PSA and FAS are bounded within the 68% confidence interval of the predicted motions at test site 1 shown in Figure 5.8. Figure 5.10 depict the recorded and mean estimated ground velocity along EW direction and five generated ground motion realizations for the target site 1. Figure 5.10 indicates that the mean estimated ground velocity is appropriately close to the recorded one. In addition, the long period content of the motions for the five shown velocity realizations is almost similar. In contrast, the difference can be seen within the motions' higher frequency (shorter wavelengths) content.

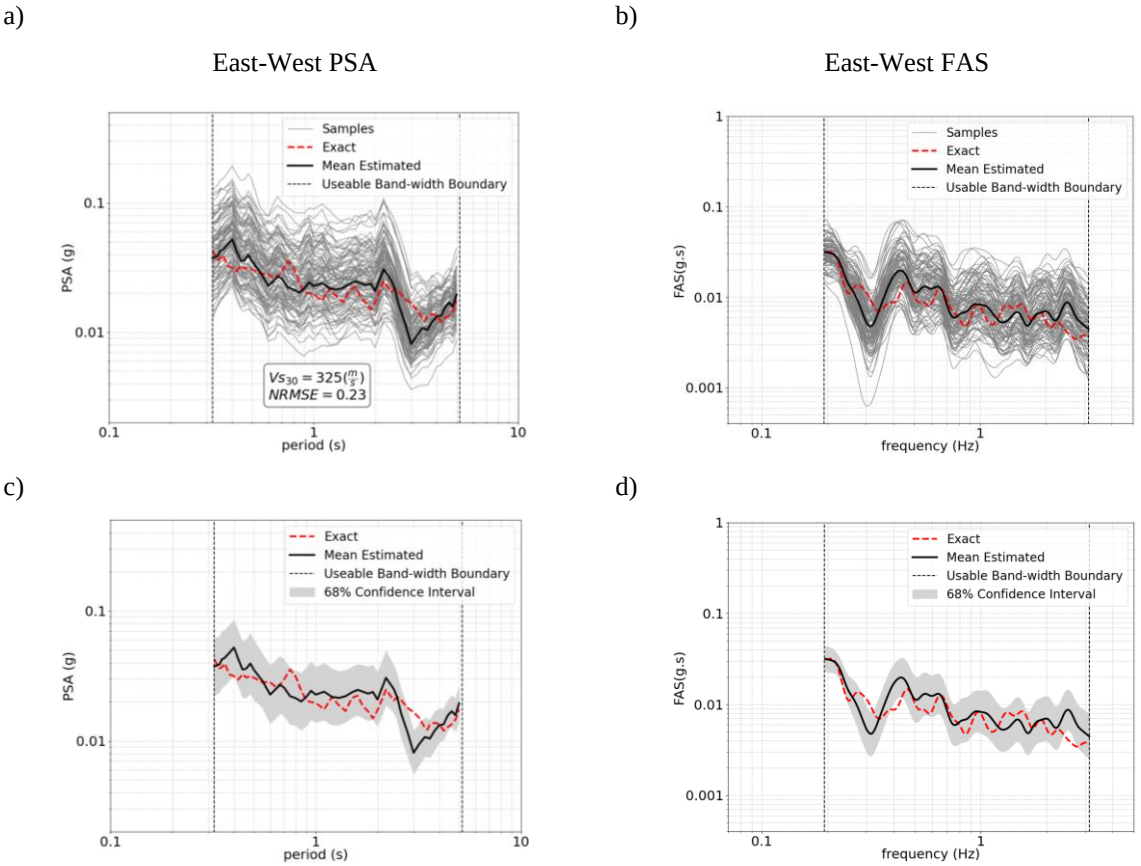


Figure 5.9. a) 5%-damped pseudo spectral acceleration (PSA) samples, b) Fourier Amplitude Spectrum (FAS) samples, c) 68% confidence interval of PSA, and d) 68% confidence interval of FAS along East-West direction at test site 1 within the **M**7.1 Ridgecrest earthquake ground motions dataset recorded by CSN

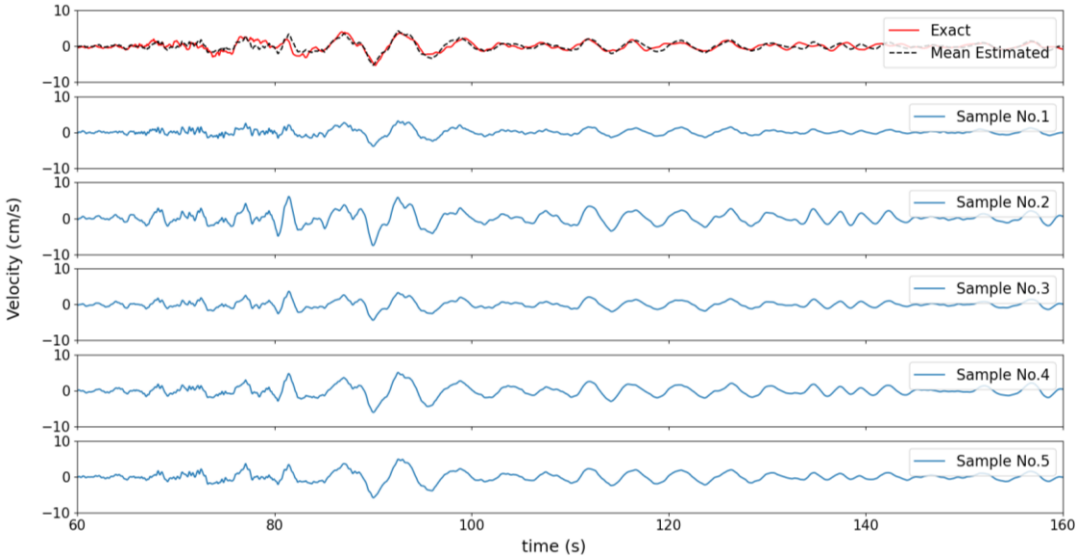


Figure 5.10. Random generated samples of ground velocity time series along East-West direction at test site 1 within the **M**7.1 Ridgecrest earthquake ground motions dataset recorded by CSN

Figure 5.11 illustrates the 5%-damped PSA and FAS along EW direction for the recorded and estimated motions for the target site 2 in Figure 5.8. The similar trend for observing more uncertainty for the higher frequencies is evidenced. It is also shown that the 68% confidence interval of the estimated motions encompasses the recorded motion's PSA and FAS. Figure 5.12 illustrates the mean estimated, recorded, and the five randomly generated ground motion velocity time series at the target site 2. It is shown that the mean estimated ground motion is appropriately close to the recorded one. In addition, the mutual long period waves among all the random realizations as well as the mean estimated ground motions is seen.

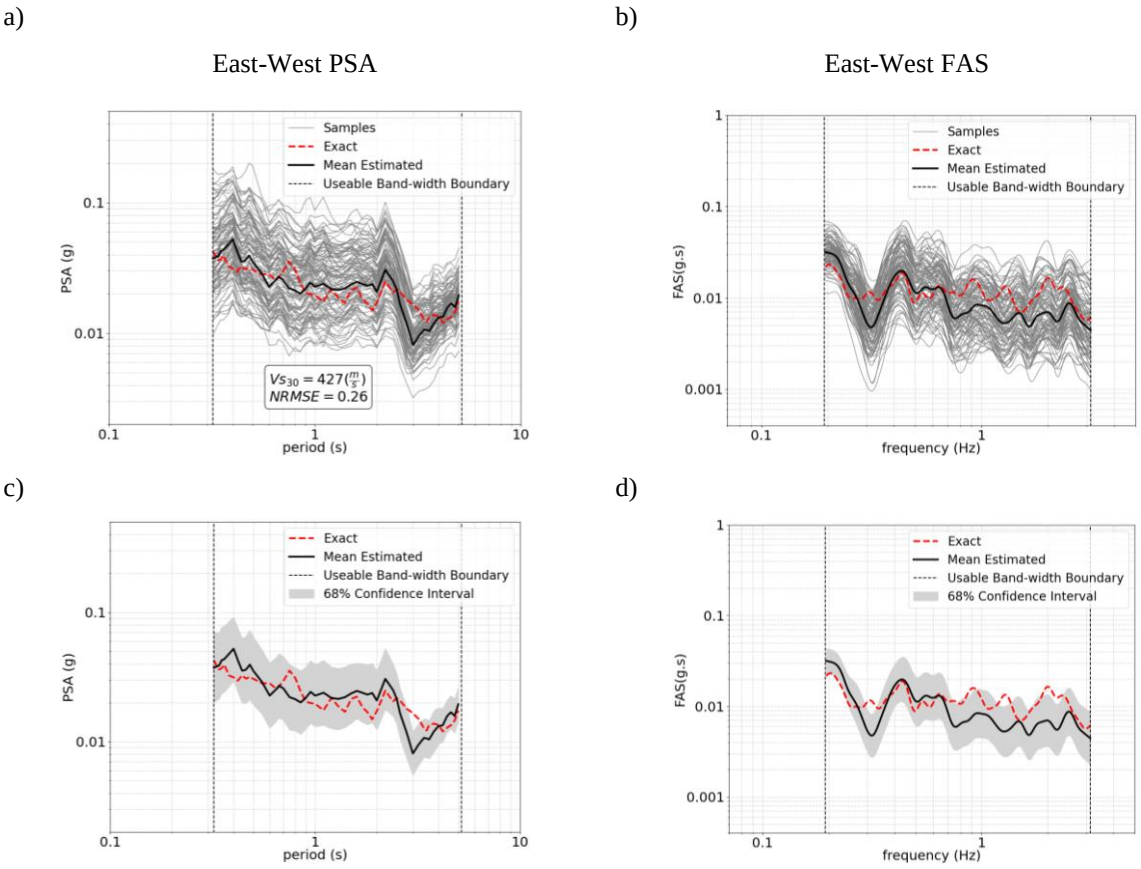


Figure 5.11. a) 5%-damped pseudo spectral acceleration (PSA) samples, b) Fourier Amplitude Spectrum (FAS) samples, c) 68% confidence interval of PSA, and d) 68% confidence interval of FAS along East-West direction at test site 2 within the **M**7.1 Ridgecrest earthquake ground motions dataset recorded by CSN

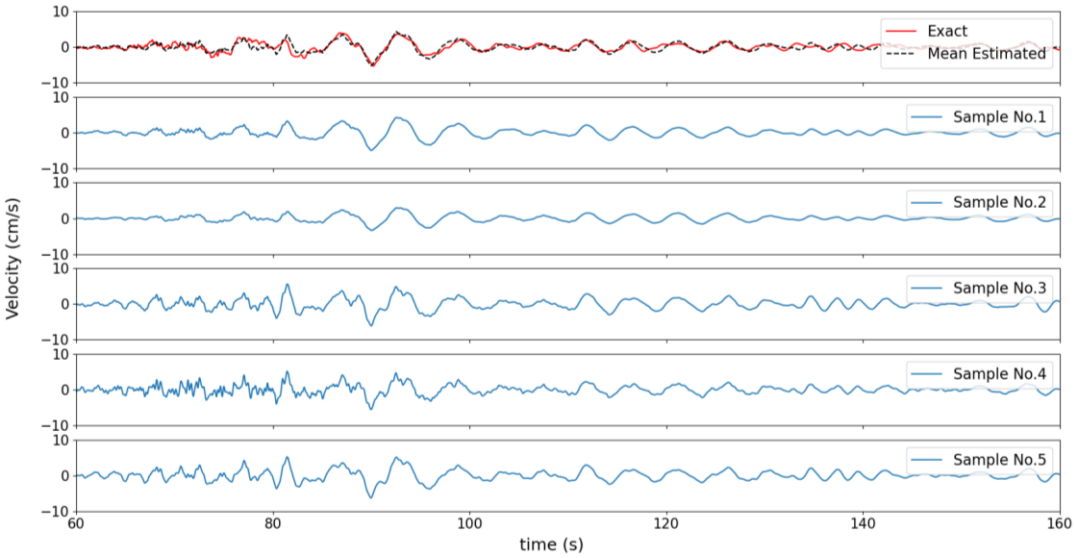


Figure 5.12. Random generated samples of ground velocity time series along East-West direction at test site 2 within the **M**7.1 Ridgecrest earthquake ground motions dataset recorded by CSN

6. Uncertainty Quantification and Sensitivity Analysis of Proposed GPR Method

In this Chapter, we aim to investigate the parameters affecting the GPR output and its uncertainty. The simulator (in this study, GPR), which approximates the complex physical systems (propagation of the waves through geological layers and ground surface movement), has their uncertainty in its outputs (Daub et al., 2021; Sobol, 2001). We implemented the established GPR model over the ground motions recorded by CSN for the 2019 **M**7.1 Ridgecrest earthquake in Los Angeles to quantify this uncertainty. We utilized the GPR model to estimate the ground motions at target sites and quantify the uncertainty of the estimated motions through the generation of ground motion realizations (see Chapter 5).

6.1. M7.1 Ridgecrest Earthquake Recorded by CSN

The various number of sites chosen out of the CSN sites that recorded the M7.1 Ridgecrest earthquake is shown in Figure 4.2. We chose the dataset with 252 sites to quantify the uncertainty of the estimated motions and conduct the sensitivity analysis on them. The LOO estimation of the ground motions at each site (as a target site) is conducted using the $\hat{\lambda}$ equals 0.05 (cf. Table 4.1). We generated the motions constructed with the posterior mean DFT coefficients and one hundred ground motion realizations at each target site. Figure 6.1 illustrates the distribution of the RotD50 NRMSE between exact and mean constructed motions for 252 sites within CSN recorded **M**7.1 Ridgecrest earthquake. There are three chosen sites whose estimated and recorded ground motions' EW PSA are shown in Figure 6.2.

6.1.1. Uncertainty Quantification of Conditioned Simulated Motions

The one hundred generated ground motion realizations provide information regarding the uncertainty of the estimated motion at each period, as is shown in Figure 6.2. In Figure 6.2, the variation of the random samples' PSA is representative of the uncertainty of the estimated motions at each period. Therefore, we computed the logarithmic standard deviation of the PSA at two periods T = 0.4 (s) and T = 2.0 (s), as two representatives for short and long period content of the motions, respectively. Figure 6.3 illustrates the distribution of the logarithmic standard deviation of PSA along the EW direction at periods T = 0.4 (s) and T = 2.0 (s).

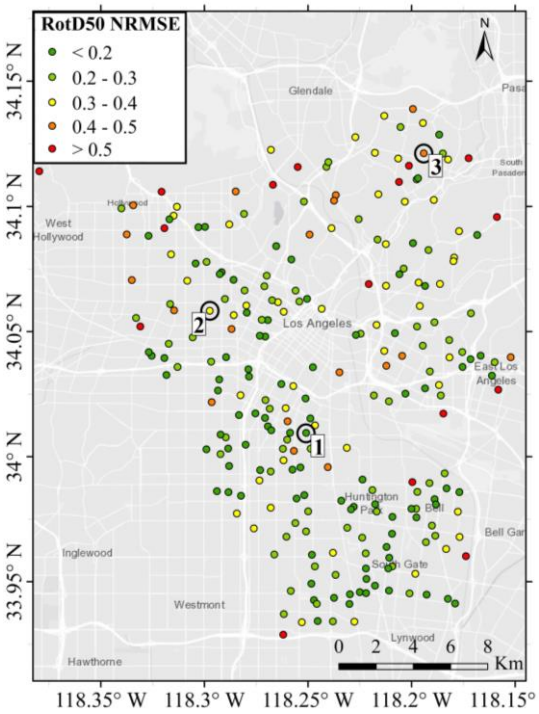


Figure 6.1. Distribution of the RotD50 NRMSE for LOO estimation for 252 sites of CSN recorded 2019 **M**7.1 Ridgecrest earthquake and three chosen test sites

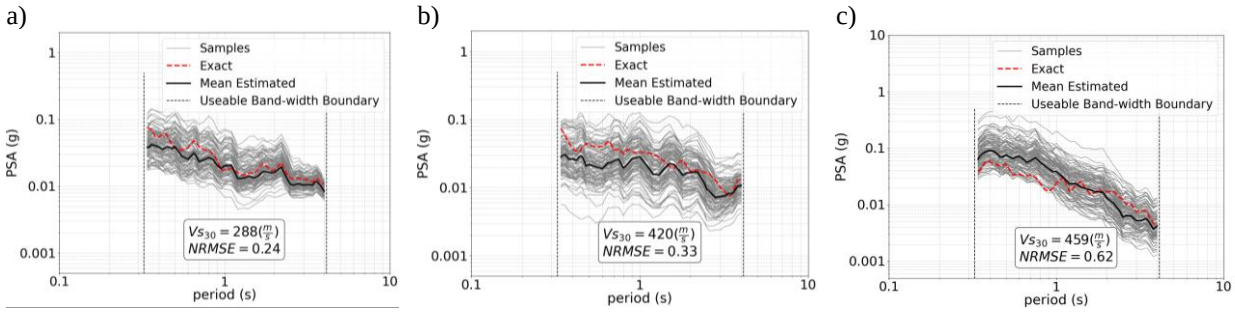


Figure 6.2. The 5% linear response spectrum of the mean estimated motion as well as 100 random ground motion realizations along East-West direction for test sites a) 1, b) 2, and c) 3 for 2019 **M**7.1 Ridgecrest earthquake dataset

Similarly, Figure 6.4 demonstrates the logarithmic standard deviation of PSA along NS directions for each site.

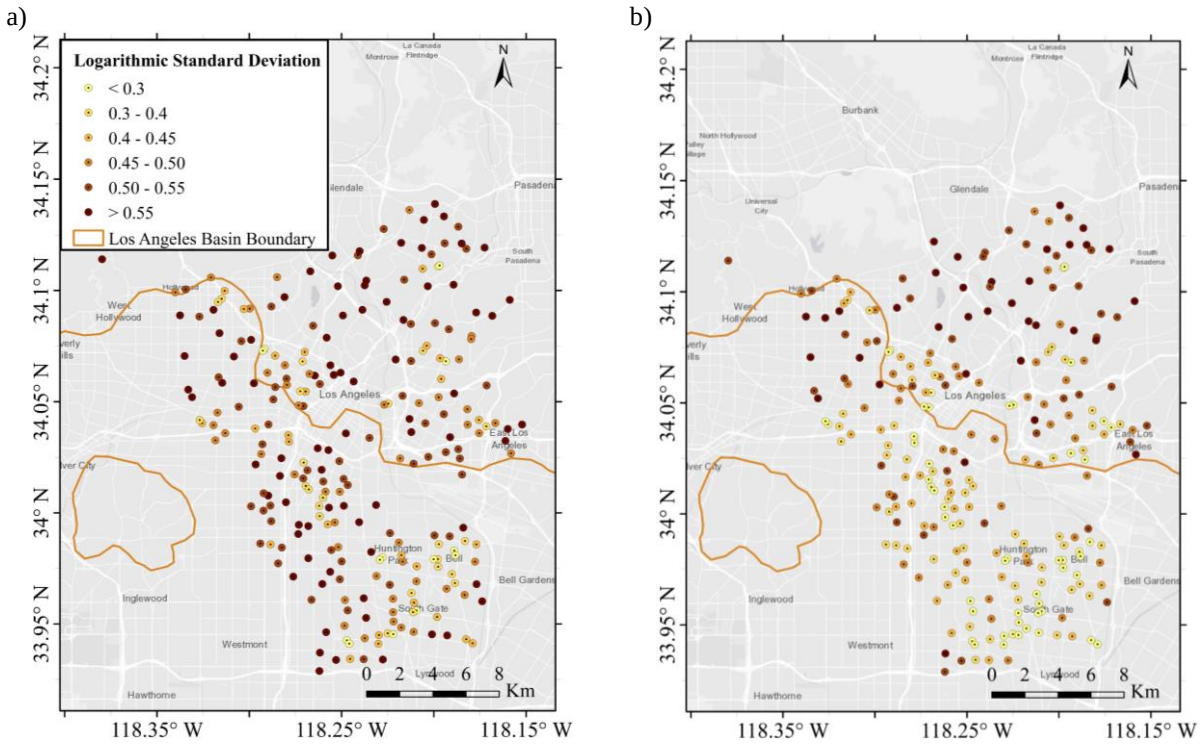


Figure 6.3. The logarithmic standard deviation of the estimated PSA along EW direction at a) 0.4 s and b) 2.0 s for 2019 **M**7.1 Ridgecrest earthquake dataset

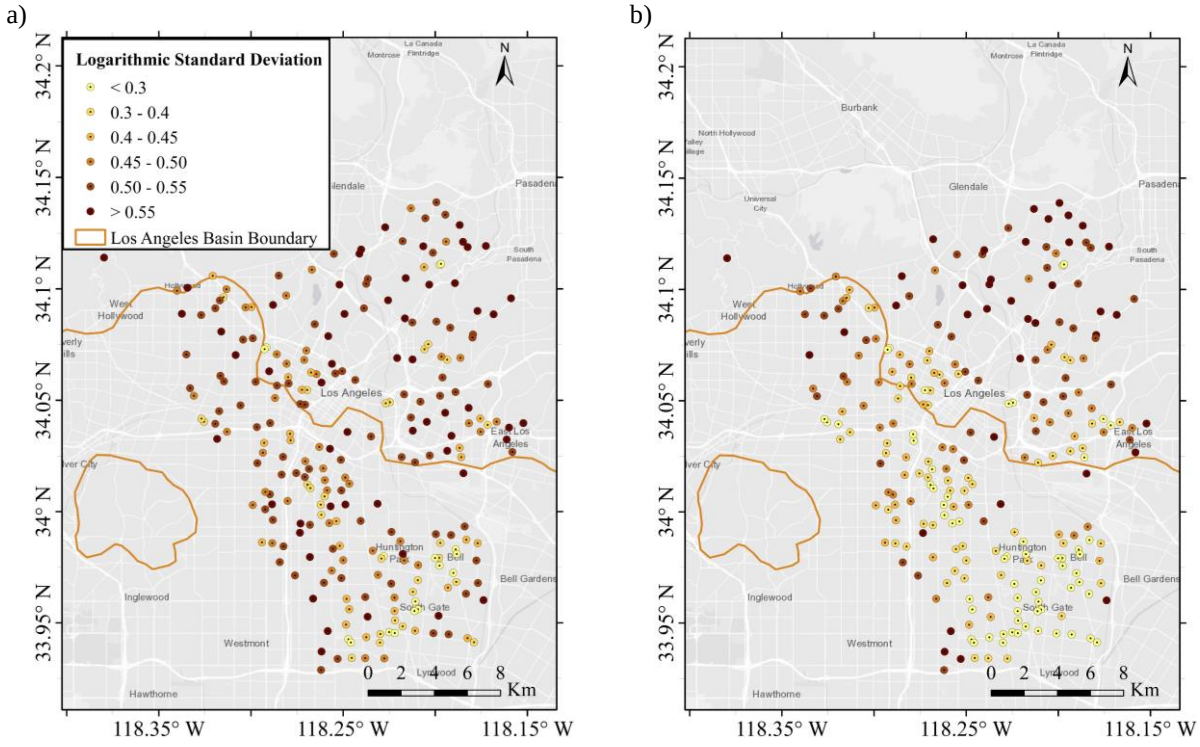


Figure 6.4. The logarithmic standard deviation of the estimated PSA along NS direction at a) 0.4 s and b) 2.0 s for 2019 **M**7.1 Ridgecrest earthquake dataset

Figures 6.3b and 6.4b illustrate that the southern region of the CSN has smaller uncertainty (logarithmic standard deviation) for long period estimation (T = 2.0 s) along both EW and NS directions. However, the uncertainty at shorter period estimations (at T = 0.4 s) does not show much variation within CSN stations; in other words, almost similar uncertainty is obtained for all sites for the short period estimation. It is also observable that the stations located within the Los Angeles basin (shown in Figure 6.3 and Figure 6.4) demonstrate lower uncertainty of the long period estimation compared to those located outside of the basin. This phenomenon is due to the more coherent motions at longer periods for those stations located on top of the Los Angeles basin (Kohler et al., 2020).

The uncertainty variation for the long period content of the motions within CSN (Figure 6.3 and Figure 6.4) is due to two reasons. First, the density of the observations around each target point is different. As a result, the target sites with more nearby observations have more input data to estimate the long period waves of the motion (southern region of the network in Figure 6.3 and Figure 6.4). In contrast, those whose surrounding observations are far from them might show higher uncertainty of the predicted motions. This variation of the uncertainty due to the observation density is investigated in more detail in Section 6.1.2. Second, the southern region of the CSN includes softer site conditions in terms of $V_{S_{30}}$, $Z_{1.0}$ (depth to the $V_S = 1$ km/s) and $Z_{2.5}$ (depth to the $V_S = 2.5$ km/s) due to the Los Angeles basin. Thus, it amplifies the long period content of the motion and makes the long period waves more coherent within southern region of the CSN (LA basin) as is mentioned in Filippitzis et al. (2021). Figure 6.5 illustrates the distribution of the estimated $Z_{1.0}$ and $Z_{2.5}$ using the SCEC CVM-S4 model (Kohler et al., 2003; Nweke et al., 2018). It is recognizable from Figure 6.5 that most of the sites with smaller long period uncertainty (Figure 6.3b and Figure 6.4b) are on top of Los Angeles basin.

Furthermore, we investigated whether the recorded ground motion's PSA at each target site falls within the mean $\pm$ one standard deviation (68% confidence interval) of the estimated motion's response spectrum. Thus, we have generated four maps that indicate whether the estimated motion's 68% confidence interval (CI) (mean $\pm$ one standard deviation) encompasses the recorded motion's PSA for two periods, T = 0.4 (s) and T = 2.0 (s). Figure 6.6 illustrates the location of the recorded motion's PSA relative to the 68% CI of the estimated motion's PSA for each target station for the EW direction. In Figure 6.6, the recorded motion's PSA is located within 68% CI for the 78% and 69% of the stations at T = 0.4 (s) and T = 2.0 (s), respectively. Similarly, Figure 6.7 shows the location of the recorded motion's response spectrum with respect to the 68% CI for the NS direction. In Figure 6.7, the recorded motion's PSA falls within the 68% CI cloud for the 74% and 77% of the stations at T = 0.4 (s) and T = 2.0 (s), respectively.

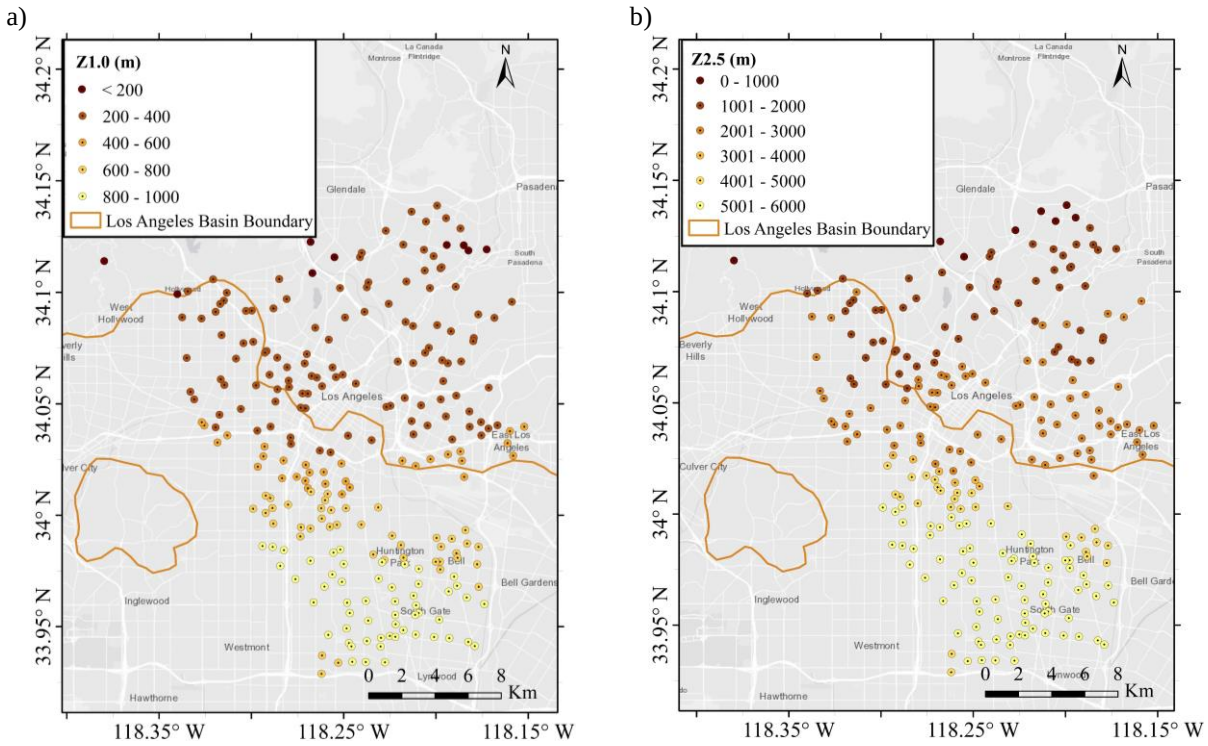


Figure 6.5. The distribution of the estimated a) $Z_{1.0}$ and b) $Z_{2.5}$ for the CSN sites

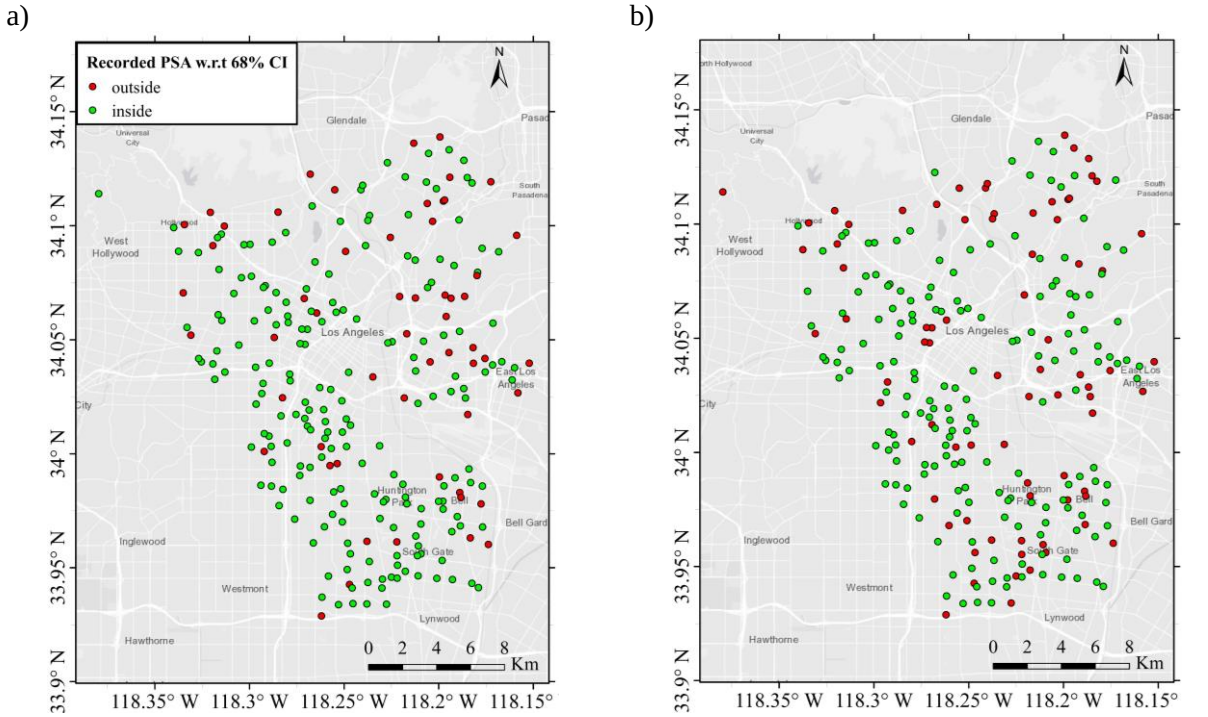


Figure 6.6. Location of the recorded motion's PSA with respect to the 68% Confidence Interval of predicted motion at a) T = 0.4 (s) and b) T = 2.0 (s) along East-West direction for 2019 **M**7.1 Ridgecrest earthquake dataset

a) b)

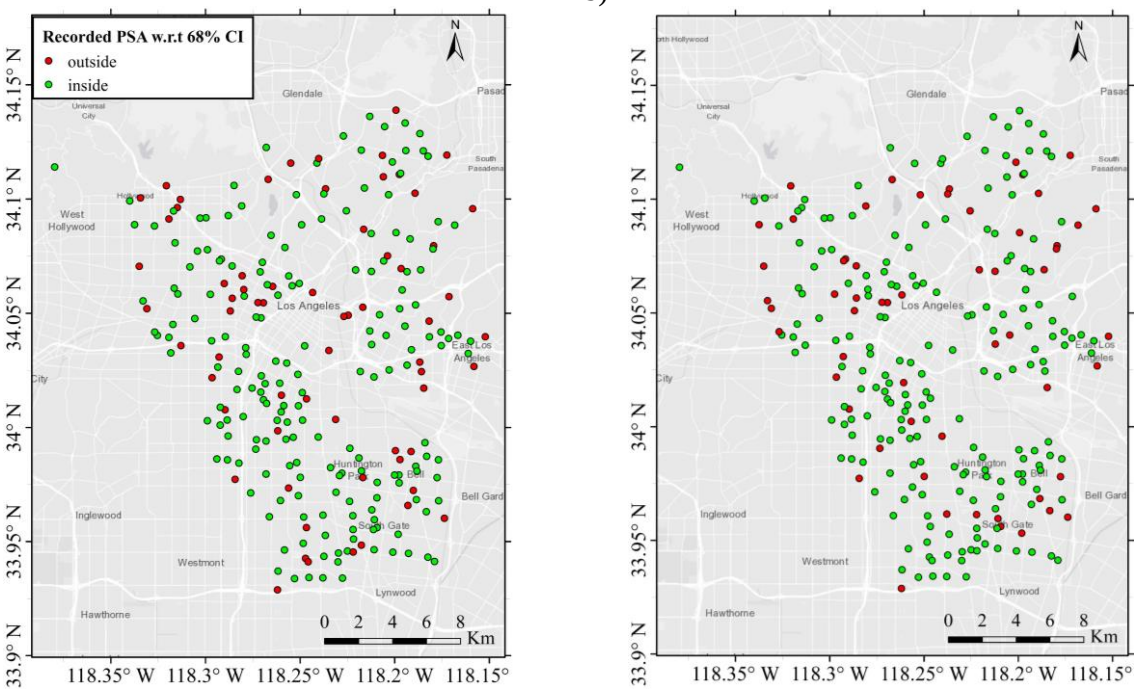


Figure 6.7. Location of the recorded motion's PSA with respect to the 68% Confidence Interval of predicted motion at a) T = 0.4 (s) and b) T = 2.0 (s) along North-South direction for 2019 **M**7.1 Ridgecrest earthquake dataset

6.1.2. Sensitivity Analysis of Conditioned Simulated Motions

We intend to study the effect of various governing characteristics, such as observation density, variation of estimated site conditions' uncertainty, and variation in the surrounding topography, on the precision and uncertainty of the estimated motions. It is essential to have an understanding of the controlling parameters that govern the GPR model output (Kwon and Elnashai 2006).

To start, we evaluated the effect of observation density surrounding each target point on the response spectrum NRMSE. We selected the four nearest observed sites to each target station and computed their average separation distance (using the 4D space stated in Section 2.2). This distance is used to represent the observation density surrounding each target site. In other words, the lower average distance from the four closest sites corresponds to a higher observation density. Figure 6.8 illustrates the scatter plot of the response spectrum NRMSE (within the usable bandwidth) along EW, NS, and the RotD50 versus the average separation distance from the four closest observed sites. The separation distance (x-axis) in Figure 6.8 is unitless as the distance within the 4D space of the GPR model is calculated between normalized vectors with zero mean and unit standard deviation and, thus is unitless.

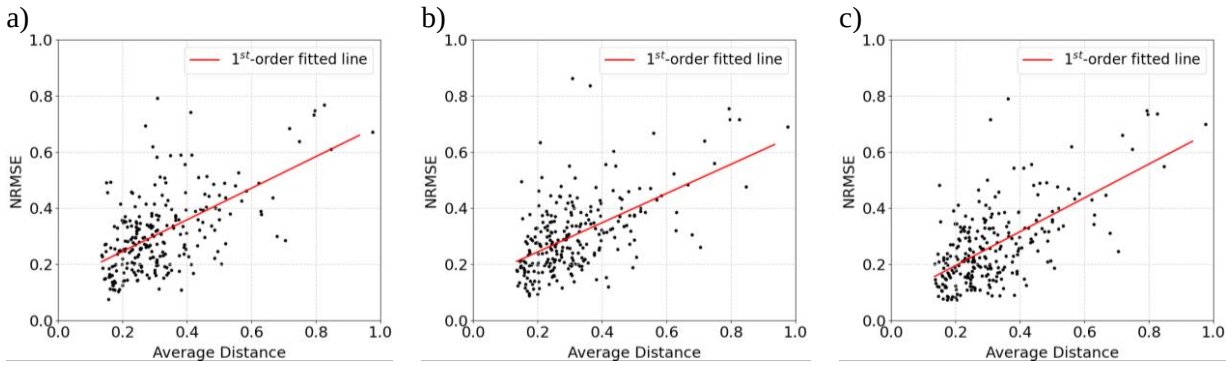


Figure 6.8. Scatter plot of the PSA normalized root mean square error along a) Eas-West and b) North-South, and c) RotD50 spectrum with respect to the average separation distance from four closest observations for 2019 **M**7.1 Ridgecrest earthquake dataset

Figure 6.8 shows that the estimation error increases by increasing the average distance from the closest observations. In other words, there is a direct relationship between the prediction error and the density of the observations surrounding the target site. Figure 6.9 shows the scatter plot of PSA relative error along both EW and NS directions at two periods, T = 0.4 (s) and T = 2.0 (s), concerning the average separation distance from the closest neighbors. The relative error between the estimated and recorded motions' response spectra at the period T, $Error_T$, is calculated by

$$Error_T = \frac{PSA_T - \widehat{PSA}_T}{\widehat{PSA}_T} \tag{6.1}$$

where PSA_T and $\widehat{PSA}_T$ are the predicted motion and recorded motion's response spectrum at period T. It is observable from Figure 6.9 that the effect of the increase in observation density on longer period's error is more than that on short period's error. The latter is interpreted based on the slope of the fitted first-order line to these scatter plots as shown in Figure 6.9. Moreover, it is observable by comparing part a with part d, part b with part e, and part c with part f in Figure 6.9. Therefore, it is recognized that the observation density could improve the ground motion time series prediction, and this improvement is more significant for longer periods based on the current observation density.

Furthermore, we investigated the effect of the observation density on the prediction uncertainty at two periods, T = 0.4 (s) and T = 2.0 (s), along EW and NS directions within Figures 6.10 and 6.11, respectively.

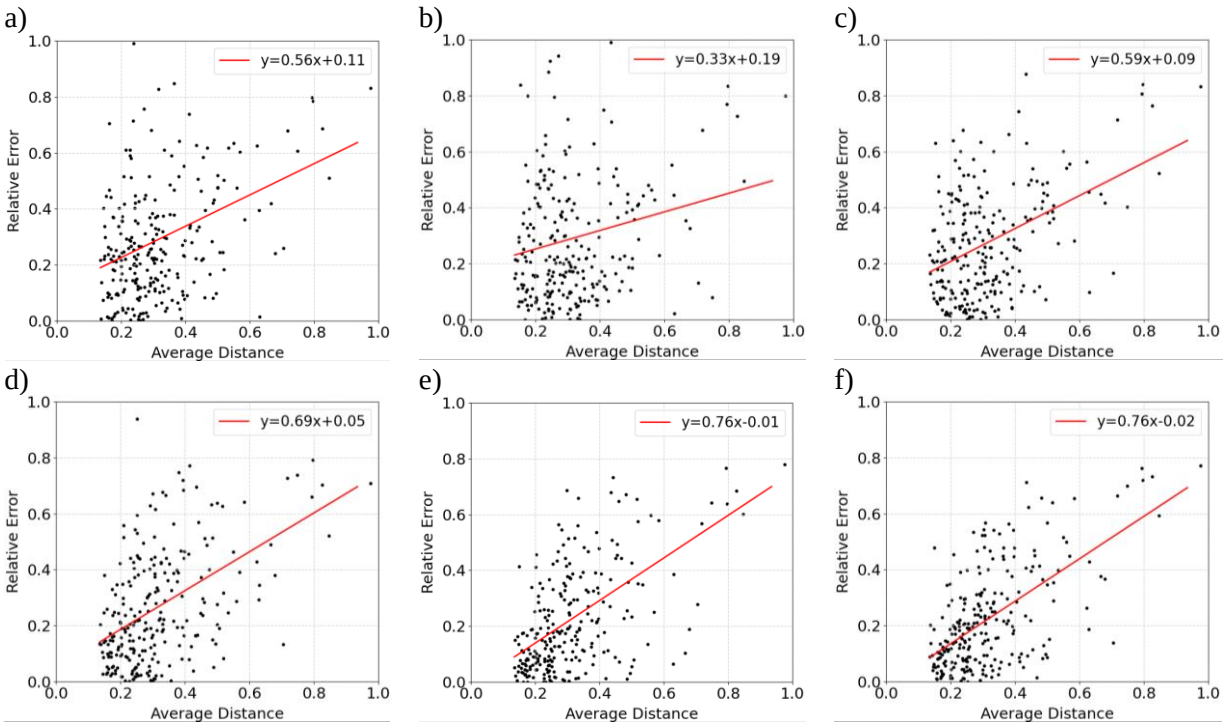


Figure 6.9. Scatter plot of the response spectrum relative error at T = 0.4 (s) along a) East-West, b) North-South, and c) RotD50 spectrum, and at T = 2.0 (s) along d) East-West, e) North-South, and f) RotD50 spectrum for 2019 **M**7.1 Ridgecrest earthquake dataset

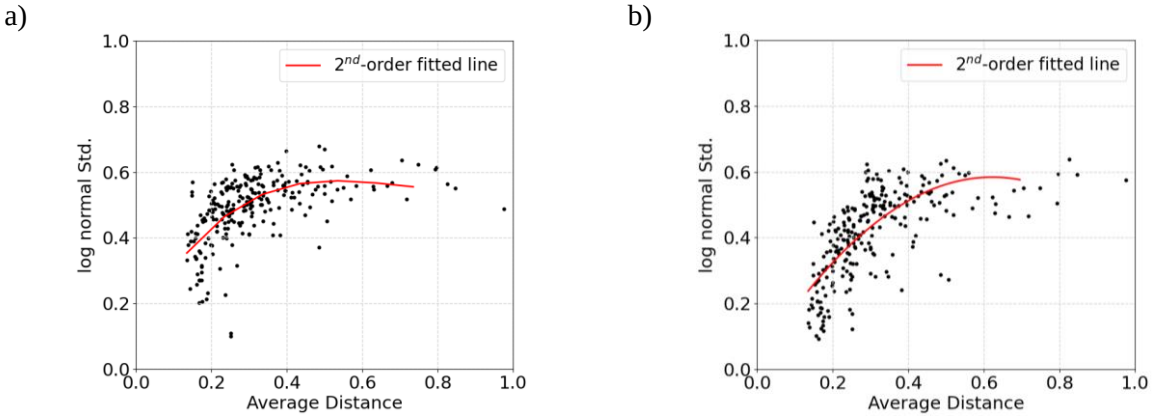


Figure 6.10. Scatter plot of the PSA log normal standard deviation at a) T = 0.4 (s) and b) T = 2.0 (s) along East-West direction for 2019 **M**7.1 Ridgecrest earthquake dataset

a) b)

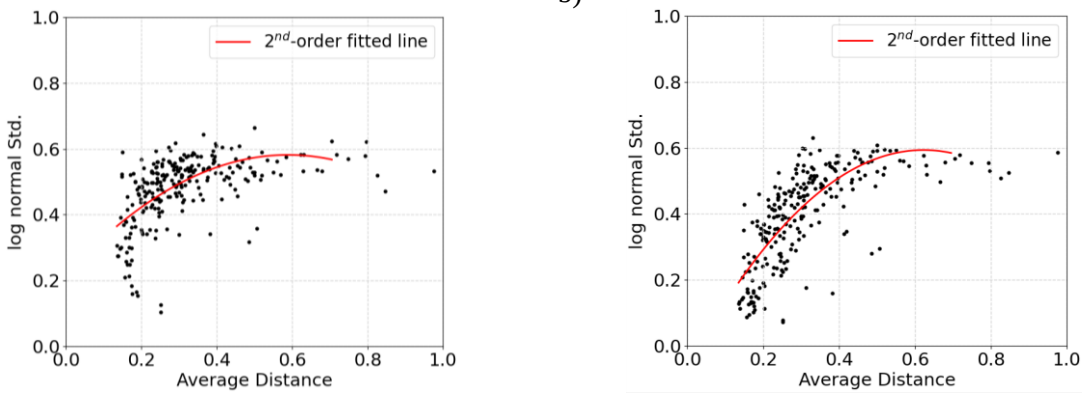


Figure 6.11. Scatter plot of the PSA log normal standard deviation at a) T = 0.4 (s) and b) T = 2.0 (s) along North-South direction for 2019 **M**7.1 Ridgecrest earthquake dataset

Figures 6.10b and 6.11b depict that estimation uncertainty increases by decreasing the observation density (increasing average separation distance from four closest observations) at longer periods, T = 2.0 s. One can recognize that the uncertainty at the longer periods is more sensitive to the separation distance from observations compared to the shorter period one by comparing Figures 6.10a with 6.10b. In other words, the uncertainty of the longer period estimation decreases considerably when more observations exist in the vicinity of the target site. However, this uncertainty does not decrease significantly at the shorter periods. This phenomenon is due to the more complexities and randomness of the short period content of the motion, which makes the added observations less effective for predicting the accurate short period content. Also, it is shown that for longer separation distances, both short and long period estimations' uncertainty is almost similar, and the uncertainty saturates for a long average separation distance.

Figure 6.12 indicates the boxplot of the average separation distance from the four closest observed sites versus whether the recorded motion's PSA falls inside the 68% CI of the predicted one along EW direction at two periods, T = 0.4 (s) and T = 2.0 (s). It is shown in Figure 6.12 that the target stations with lower average separation distance from their closest neighbors (higher observation density) is more probable to include the recorded PSA in their 68% CI rather than those with lower observation density. Similarly, Figure 6.13 demonstrates the boxplot of the average separation distance from the four closest observed stations versus the location of the recorded motion's PSA along NS direction with respect to the 68% CI of predicted motion at two periods, T = 0.4 (s) and T = 2.0 (s). Same result could be observed from Figure 6.13 for the NS direction that the target sites prediction that their 68% CI includes the recorded PSA have in general higher observation density close to them.

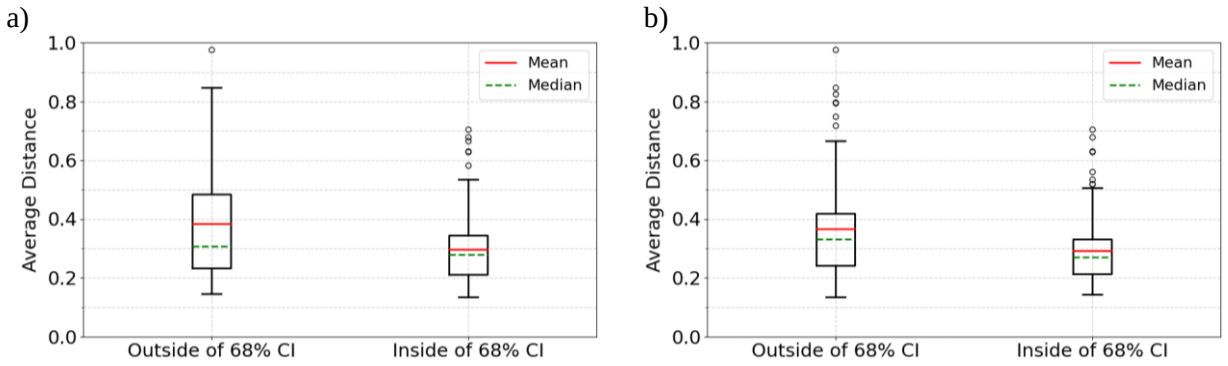


Figure 6.12. Boxplots of the average separation distance from the four closest observed stations versus location of the recorded motion's PSA with respect to the predicted motion's PSA 68% confidence interval for a) T = 0.4 (s) and b) T = 2.0 (s) along East-West direction

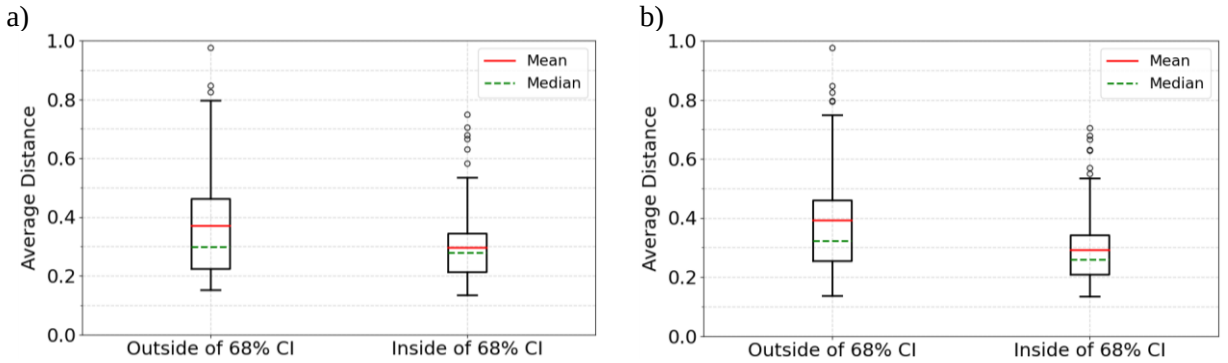


Figure 6.13. Boxplots of the average separation distance from the four closest observed stations versus location of the recorded motion's PSA with respect to the predicted motion's PSA 68% confidence interval for a) T = 0.4 (s) and b) T = 2.0 (s) along North-South direction

The effect of observation density using a representative average separation distance from the four closest neighbors on the accuracy and uncertainty of the predicted motions for the target sites have been investigated.

We study the effect of other parameters such as variation of the topography and uncertainty of the estimated site conditions on the accuracy and uncertainty of the estimated motions. The slope of the sites is obtained from a global slope database (Fischer et al. 2008), and the uncertainty (log-normal standard deviation) of the estimated $V_{s_{30}}$ are obtained using the proxy-based model (Ahdi et al. 2020). Figure 6.14 demonstrates the scatter plot of the variation (logarithmic standard deviation) of the estimated $V_{s_{30}}$ uncertainty for each target site and its four closest neighbors (i.e., the standard deviation of totally five logarithmic standard deviations) versus response spectrum NRMSE along EW, NS, and RotD50 spectrum.

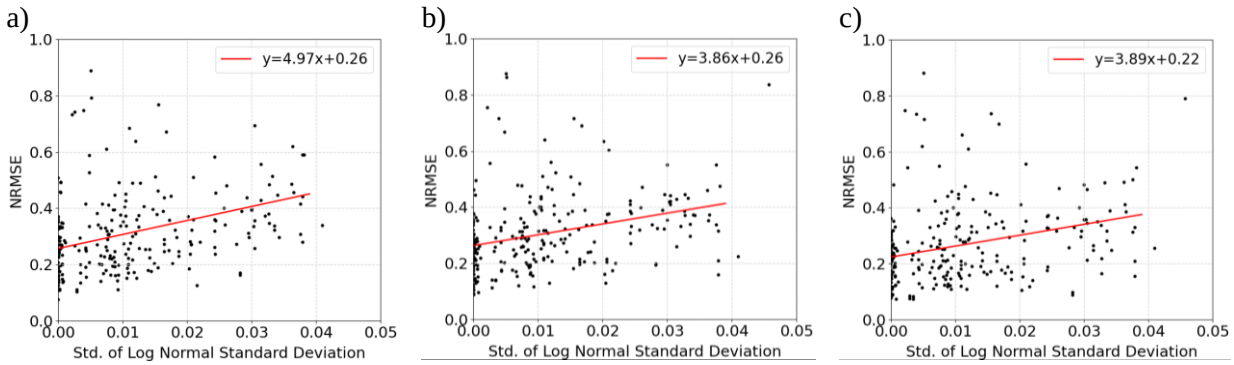


Figure 6.14. Scatter plot of the variation of logarithmic standard deviation of estimated $V_{s_{30}}$ for each target site and its four closest neighbors versus normalized root mean square error of response spectrum for a) EW direction, b) NS direction, and c) RotD50

It is observable from Figure 6.14 that the estimation error for both horizontal directions increases for the target sites with a higher variation of the estimated $V_{s_{30}}$ uncertainty. In other words, a target site that $V_{s_{30}}$ of some of their neighbors are predicted confidently, but the others are estimated with higher uncertainty, are more prone to have a higher prediction error. Similarly, Figure 6.15 illustrates the scatter plot of the variation (standard deviation) of the logarithmic standard deviation of estimated $V_{s_{30}}$ for each target site and its four closest neighbors versus response spectrum log-normal standard deviation along EW and NS at T = 2.0 (s). In other words, Figure 6.15 investigates the effect of variation of estimated site condition uncertainty around each target site on the uncertainty of the predicted motions at long period. Figure 6.15 demonstrates that the target sites that the site condition of some of their neighbors are predicted confidently, and the others with higher uncertainty are prone to have higher uncertainty over long periods. This means that the localized surrounding observations need to have estimated site conditions with almost similar confidence to result in accurate and certain predictions.

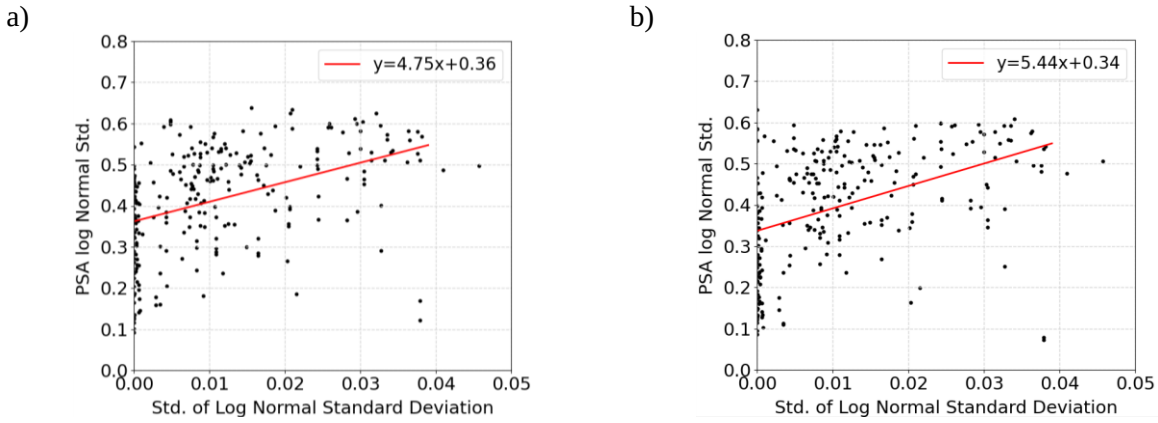


Figure 6.15. Scatter plot of the variation (standard deviation) of logarithmic standard deviation of estimated $V_{s_{30}}$ for each target site and its four closest neighbors versus response spectrum log normal standard deviation along a) EW and b) NS directions for T = 2.0 (s)

Figure 6.16 also demonstrates the scatter plot of the average estimated slope of each target site and its four closest neighbors versus response spectrum NRMSE along EW, NS, and the RotD50 spectrum. Figure 6.16 clearly illustrates that the slope surrounding each target site does not affect the estimation error significantly. Thus, being on top of steep or flat surfaces for the sites does not necessarily affect the accuracy of the GPR method estimation.

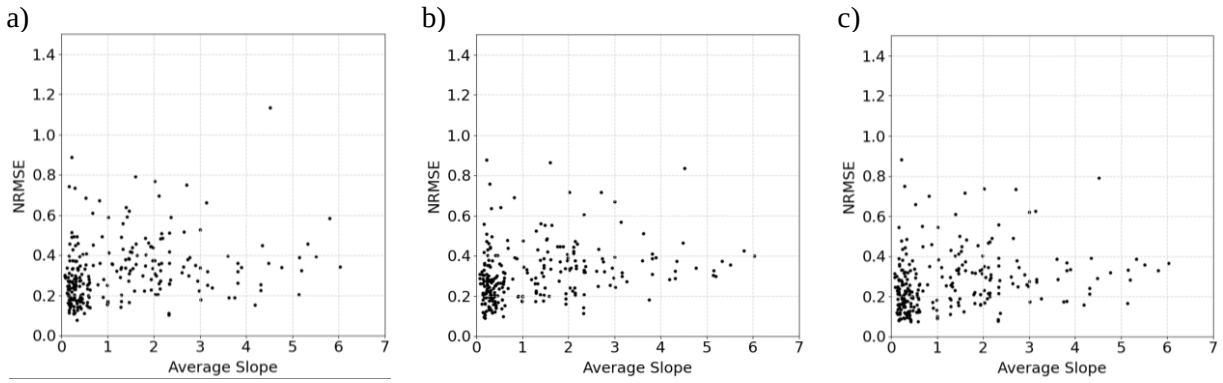


Figure 6.16. Scatter plot of the average estimated slope of each target site and its four closest neighbors versus response spectrum normalized root mean square error for the directions a) East-West, b) North-South, and c) RotD50

7. Performance Evaluation of The Proposed GPR Method on a Combination of Network Datasets

The CSN recorded motions for the 2019 **M**7.1 Ridgecrest earthquake are studied to investigate the sensitivity of the accuracy and uncertainty of the predictions to various parameters in Chapter 6. In addition, the performance of the GPR model using the optimum hyper-parameter (regularization factor, $\hat{\lambda}$) for different observation densities is investigated in Chapter 4. It was observed that one of the key parameters to decrease the error and uncertainty of the predictions is the density of the observations surrounding the target site. Therefore, we aim to investigate the potential improvement of the prediction using the added observations from other networks. There are various recording networks in CA, such as the California Strong Motion Instrumentation Program in the department of California Geological Survey (CGS-CSMIP) (Ryerson et al. 2021), the California Institute of Technology (Caltech), and the National Strong Motion Project (NP). We refer to the combination of all these networks as the California Integrated Seismic Network (CISN) hereafter within this study.

First, we implement LOO ground motion prediction for each CISN station. In other words, we conditionally simulate the ground motion time series at each CISN site using all other CISN stations (excluding the target site). Second, we repeat the same procedure to predict the ground motion time series at each CISN site, using all other CISN and CSN sites' recorded ground motions as observation. Comparing the predicted motions resulting from the two mentioned steps with the recorded (so-called "exact") ones reveals the improvement of the GPR methodology by feeding more observations.

7.1. Ground Motion Simulation for M7.1 Ridgecrest Earthquake Using CISN and CSN Networks

We selected 121 ground-level stations from the CISN network that recorded the 2019 M7.1 Ridgecrest earthquake in Los Angeles. These 121 recording sites are distributed over an approximately 3,100 km^2 region, while the 252 ground level recording sites of CSN are located in a much denser area over a 460 km^2 region. The distribution of the CISN and CSN sites over Los Angeles is shown in Figure 7.1a. As illustrated in Table 4.2, the implemented $\hat{\lambda}$ for prediction through the GPR methodology is dependent on the observation density around each target site. As discussed previously, we aim to predict the ground motions at each CISN site location using 1) all other CISN and all CSN sites as observation; and 2) just all CISN sites as observation. We divided the whole 121 CISN sites domain into three different subdomains; 1) Inner Domain, 2) Middle

Domain, and 3) Exterior Domain (see Figure 7.1.b). The observation density existing in each domain and subsequently the required $\hat{\lambda}$ are different. On the other hand, there is one observation density for the second scenario, where the observation for each CISN target site includes only the rest of the CISN stations. Tables 7.1 and 7.2 show the density of observation for each domain and implemented $\hat{\lambda}$ (based on Table 4.2) for the target sites at the corresponding region for scenario 1 (CISN plus CSN observations) and scenario 2 (CISN observations), respectively. Tables 7.1 and 7.2 demonstrate how the sites located in the exterior domain (yellow domain in Figure 7.1b) require a lower $\hat{\lambda}$=0.2 for having more observations (CISN + CSN sites) rather than the scenario with a smaller number of observations (CISN only) which needs $\hat{\lambda}$=0.4. In addition, it is shown that the inner domain of the CISN sites requires a smaller $\hat{\lambda}$=0.05 for having both observations of CSN and CISN sites. The transition between the inner and exterior regions contains the middle domain (purple domain in Figure 7.1b), which requires $\hat{\lambda}$=0.1.

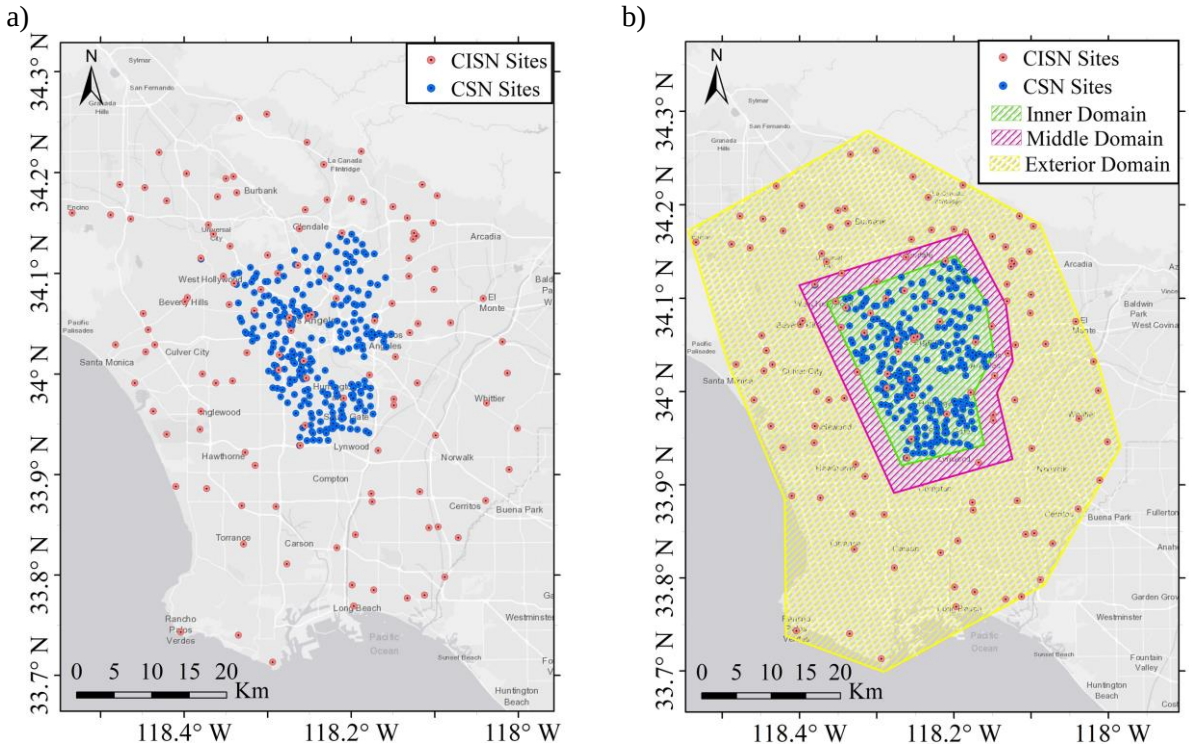


Figure 7.1 a) Distribution of the CSN and CISN sites recorded 2019 **M**7.1 Ridgecrest earthquake within Los Angeles, and b) Division of CISN network in Los Angeles into three different sub domains with various density of observations

Table 7.1. The implemented $\hat{\lambda}$ for implementing GPR model at different domains within the CISN network using CISN and CSN sites as observation

Observations	Target Domain	Area (km^2)	Observation Density (site/km^2)	$\hat{\lambda}$
CISN + CSN	Inner	464	0.57	0.05
CISN + CSN	Middle	764	0.36	0.10
CISN + CSN	Exterior	3103	0.12	0.20

Table 7.2 . The implemented $\hat{\lambda}$ for implementing GPR model within the CISN network using CGS sites as observation

Observations	Area (km^2)	Observation Density (site/km^2)	$\hat{\lambda}$
CISN	3103	0.04	0.40

Figure 7.2 illustrates the distribution of the CISN sites and the NRMSE between the predicted and recorded ground motions' RotD50 spectrum for having just CISN sites as observation. It is worth noting that these predictions follow the LOO procedure such that for each site as a target site, all the rest stations are used as observation using the $\hat{\lambda}$ mentioned in Table 7.2. As a result, the average RotD50 spectrum NRMSE among all CISN stations is 0.47. Similarly, the average response spectrum NRMSE along both EW and NS directions is 0.54.

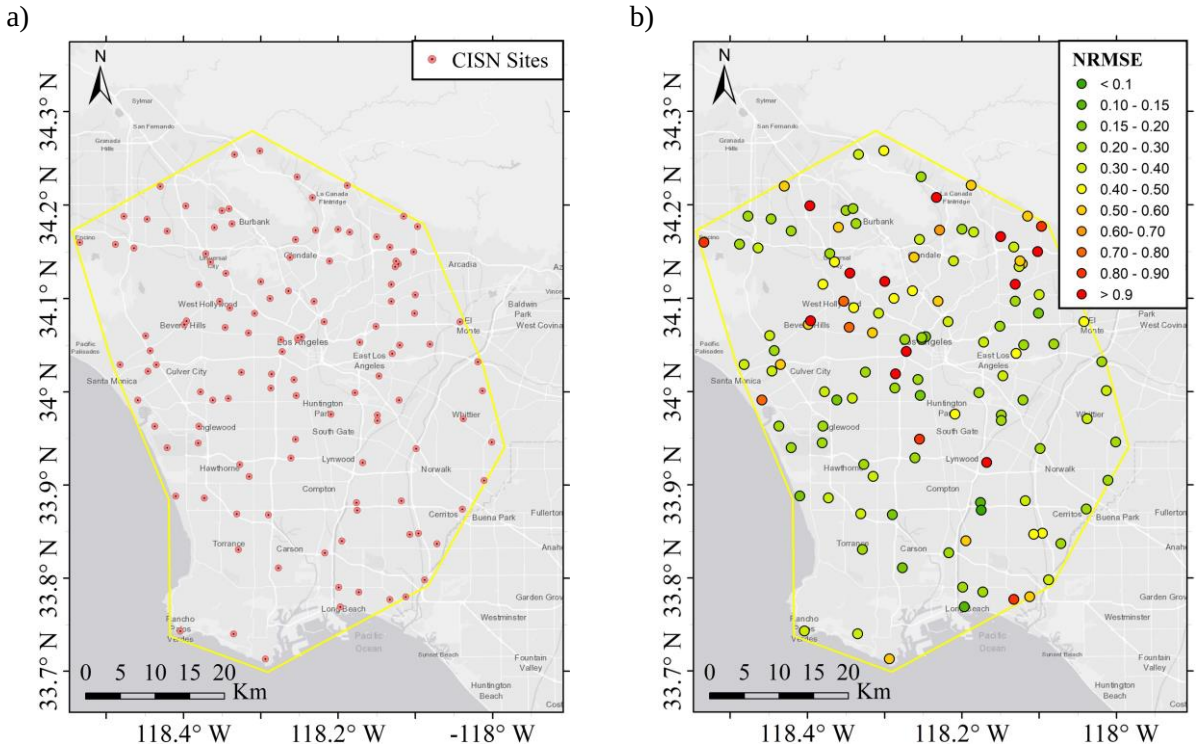


Figure 7.2 Distribution of a) all CISN sites as observation and b) the RotD50 spectrum normalized root mean suqare error at each CISN site as a target site within the **M**7.1 Ridgecrest earthquake dataset

We also predict the ground motion time series at each CISN site using both CSN and CISN sites as observation. Several steps are required to make both networks recorded motions consistent. First, the earliest Universal Time Coordinated (UTC) start time, t_{start}, of the records among all CSN and CISN recorded motions is chosen as the time origin. We conduct zero-padding at the beginning of all records to have them starting at the t_{start}. Then, the latest UTC end time, t_{end}, of the records among both CSN and CISN recorded motions is chosen as the end time of all records, and a zero-padding at the end of all records is conducted to have all of them end at the t_{end}. Finally, The largest time step of the recorded motions among all records, dt = 0.02 (s), is selected as the target time

step of all records, and a down-sampling procedure is implemented over those records with a smaller time step. Lastly, all the CISN recorded motions are rotated to be parallel to the direction of horizontal components of CSN records, which are in the direction of EW and NS, respectively. By taking these steps, all the CSN and CISN records will start and end simultaneously with the same time step. Figure 7.3 depicts the distribution of the CISN and CSN stations (observations) and the RotD50 spectrum NRMSE at each CISN network's site using all CISN and CSN sites as observations.

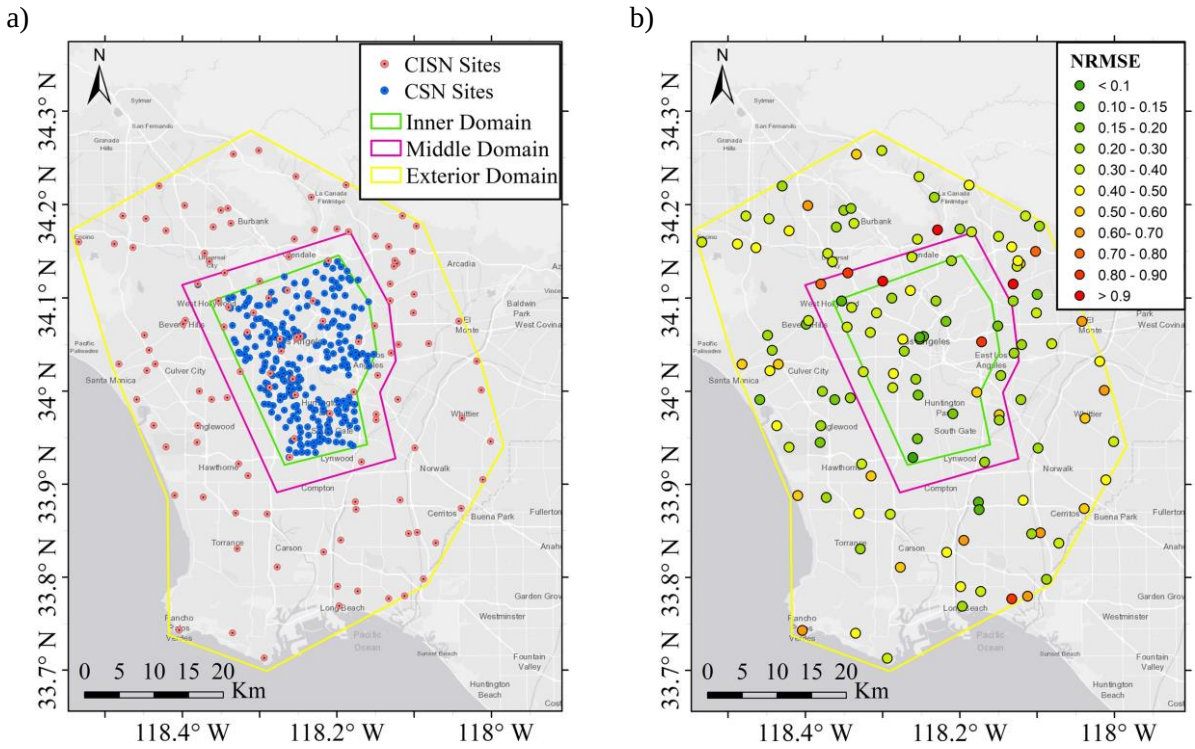


Figure 7.3 Distribution of a) all CISN and CSN sites as observation and b) the RotD50 spectrum normalized root mean suqare error at each CISN site as a target site within the **M**7.1 Ridgecrest earthquake dataset

The average RotD50 spectrum NRMSE for all CISN sites is 0.39 (Figure 7.3). This means that 252 added CSN observations decreased prediction error by 17%. Similarly, the average response spectrum NRMSE along EW and NS directions are 0.43 and 0.44, respectively. Therefore, the prediction error along EW and NS decreased by 20% and 18.5%, respectively. To compare the estimation error between two scenarios, we plot the PSA NRMSE along EW, NS, and RotD50 spectrum for both scenarios within Figures 7.4 through 7.6, respectively. Figure 7.4 indicates that the prediction error for EW decreased after having more observations of the CSN dataset, especially within the inner domain. Furthermore, it is shown that most of the higher prediction errors (red stations in Figure 7.4) are in the outer domain, where there are not any added observations. Similar results are observed for the NS direction's response spectrum and RotD50, shown in Figures 7.5 and 7.6, respectively. Table 7.3 compares the prediction error along each horizontal direction and RotD50 spectrum within different domains of the CISN for two scenarios.

a) b)

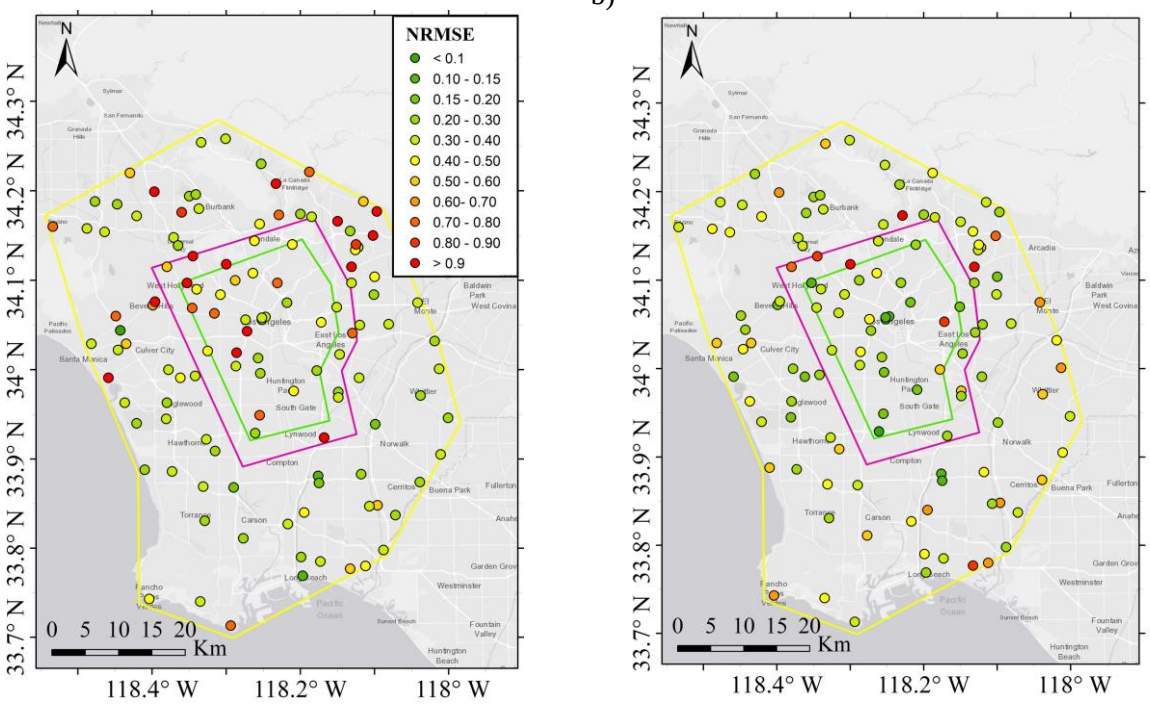


Figure 7.4. Distribution of response spectrum NRMSE along EW direction for having a) CISN sites and b) all CISN and CSN sites as observations within the **M**7.1 Ridgecrest earthquake dataset

a) b)

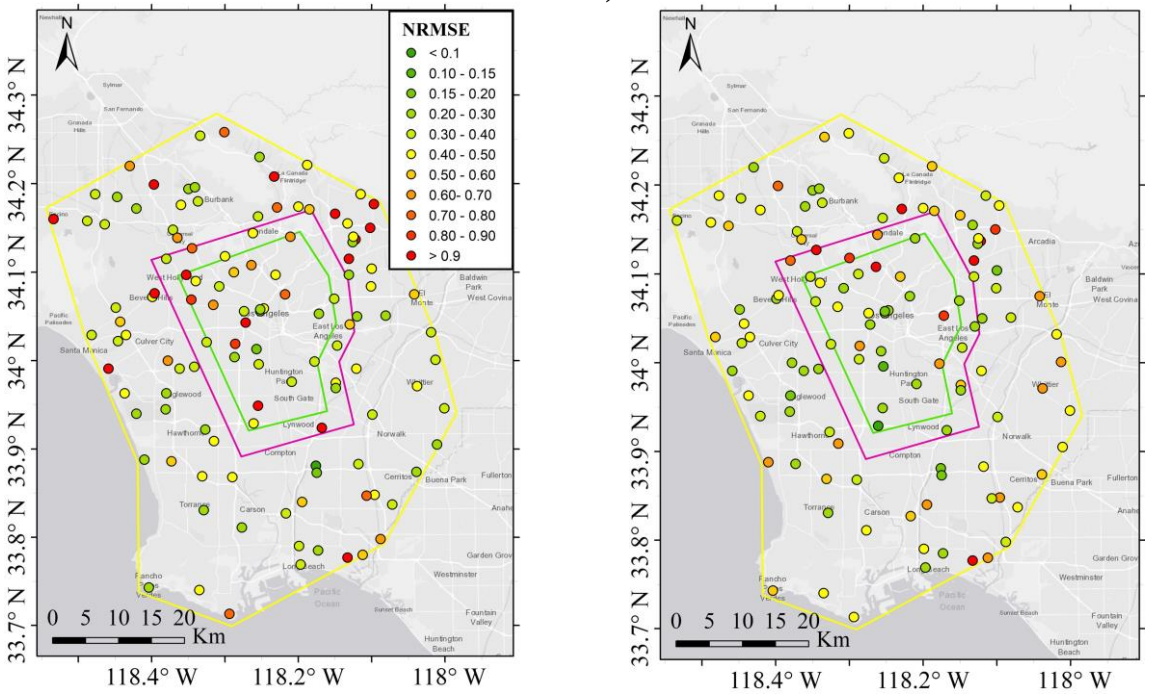


Figure 7.5 Distribution of response spectrum NRMSE along NS direction for having a) CISN sites and b) all CISN and CSN sites as observations within the **M**7.1 Ridgecrest earthquake dataset

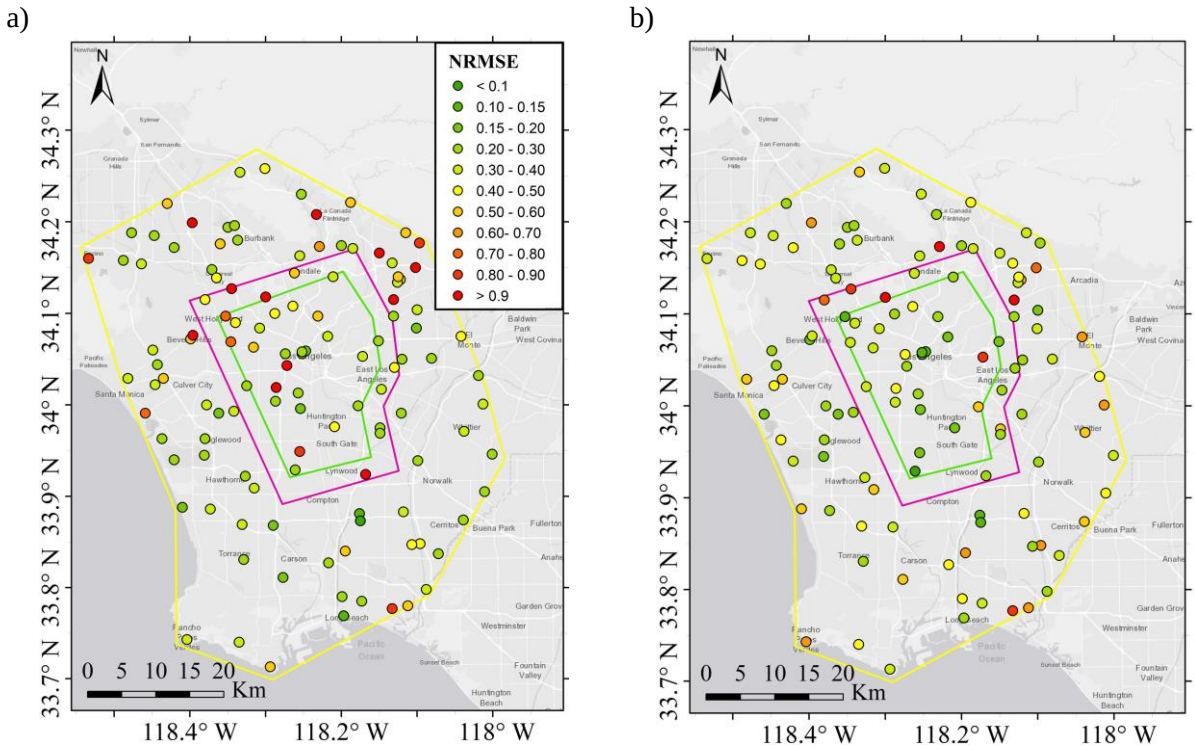


Figure 7.6 Distribution of RotD50 NRMSE for having a) CISN sites and b) all CISN and CSN sites as observations within the **M**7.1 Ridgecrest earthquake dataset

Table 7.3 . The Prediction Error along both horizontal directions as well as RotD50 spectrum within different Domains

Domain	Observations	East-West		North-South		RotD50	
		Average NRMSE	Error Reduction*	Average NRMSE	Error Reduction	Average NRMSE	Error Reduction
Inner	CISN + CSN	0.33	43%	0.37	35%	0.29	42%
	CISN	0.58		0.57		0.50	
Middle	CISN + CSN	0.56	23%	0.50	11%	0.47	23%
	CISN	0.73		0.56		0.61	
Exterior	CISN + CSN	0.45	10%	0.46	13%	0.41	9%
	CISN	0.50		0.53		0.45	

* Error Reduction shows the reduction in the prediction error in two scenarios (i.e., due to the added CSN sites)

Table 7.3 demonstrates that the added CSN sites decreased the average prediction error among all sites within each domain. Table 7.3 shows that added CSN sites decreased the NRMSE along EW and NS directions for the CISN sites within the inner domain by 43% (Figure 7.4) and 35% (Figure 7.5), respectively. Table 7.3 illustrates the added CSN sites have the most and least effects on the prediction error for the inner and outer domains' target sites, respectively.

We also aim to investigate the added observations from CSN sites' effect on the prediction errors at both short and long-period content of the motions. To do so, we plotted the distribution of the

relative error at two periods, T = 0.4 (s) and T = 2.0 (s), for EW, NS, and RotD50 spectra within Figures 7.7 through 7.12. Tables 7.4 and 7.5 demonstrate the average errors for short and long period content, T = 0.4 s and T = 2.0 s, respectively. As is shown in both Tables 7.4 and 7.5, the error reduction at both short and long periods, due to the added CSN sites as observation, is highest within the inner domain and lowest for the outer domain. In addition, the added observations improved the prediction in the short period (T = 0.4 s) more than the long period (T = 2.0 s) within the inner domain. The reason might be the density of the added (CSN) observations between previously existing CISN sites, which updated the GPR model with information about the short-length waves to make the short-period predictions more accurate. In addition, the long period prediction without added CSN observations was quite accurate within inner domain. Thus, the additional observations are not necessary to improve the current long-period prediction.

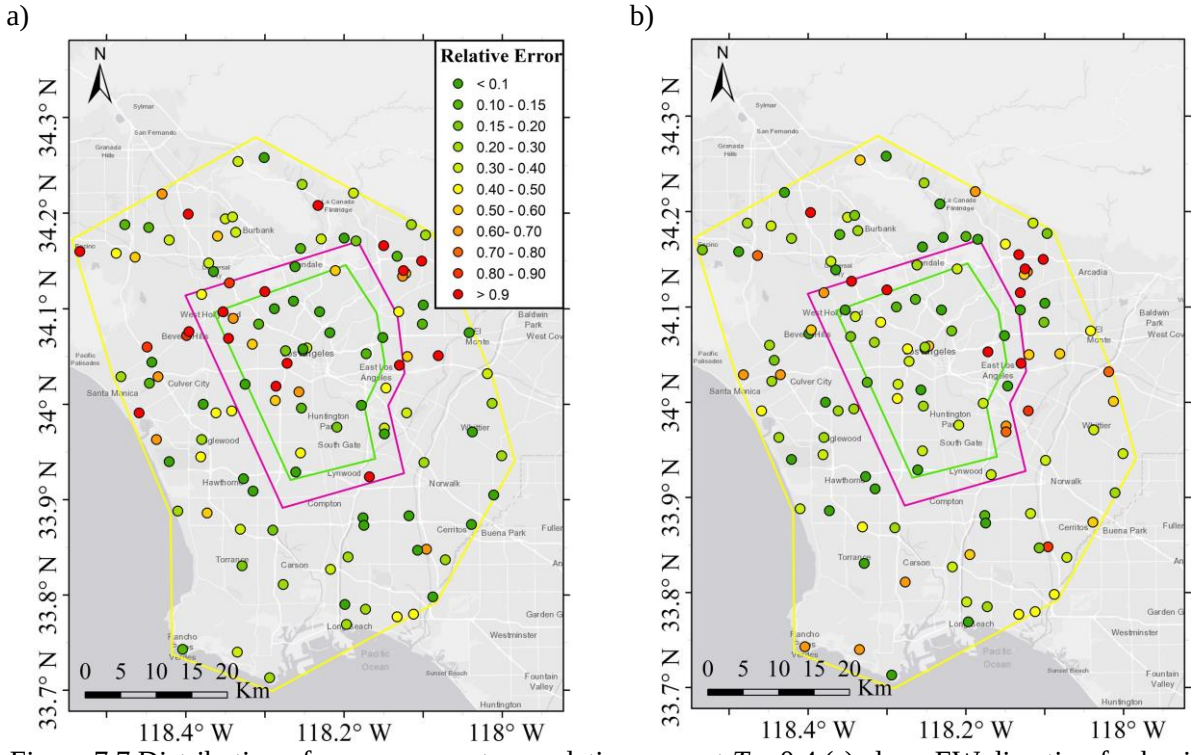


Figure 7.7 Distribution of response spectrum relative error at T = 0.4 (s) along EW direction for having a) CISN sites and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

a) b)

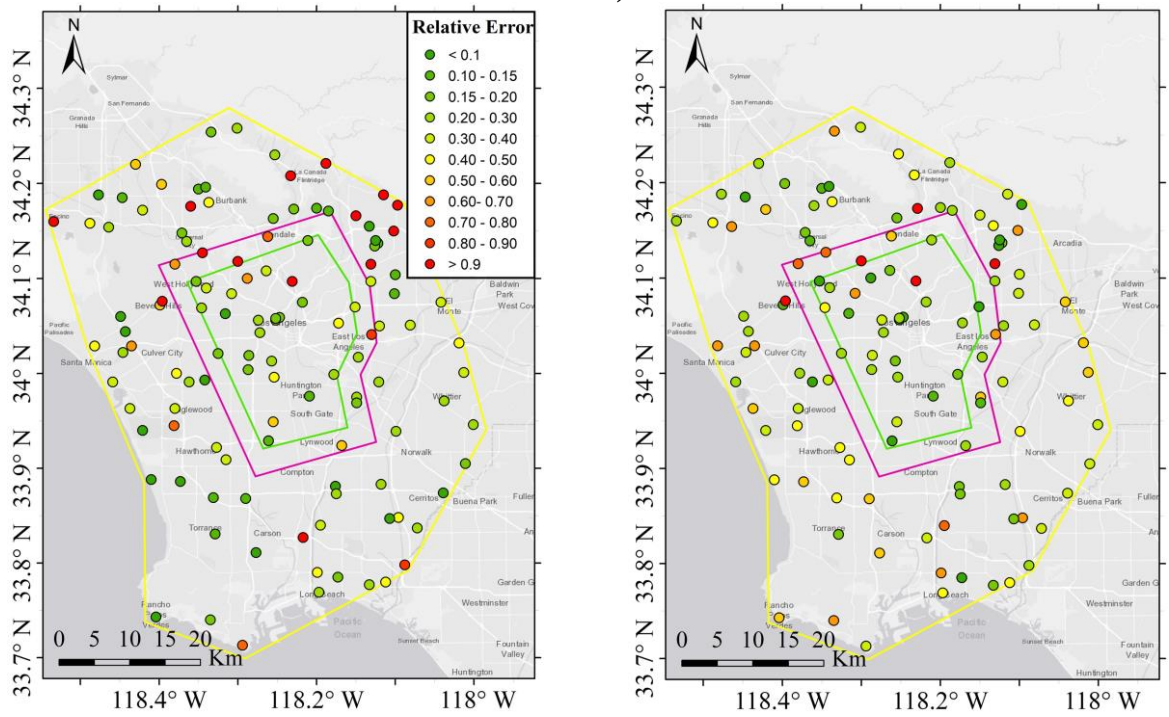


Figure 7.8 Distribution of response spectrum relative error at T = 2.0 (s) along EW direction for having a) CISN sites and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

a) b)

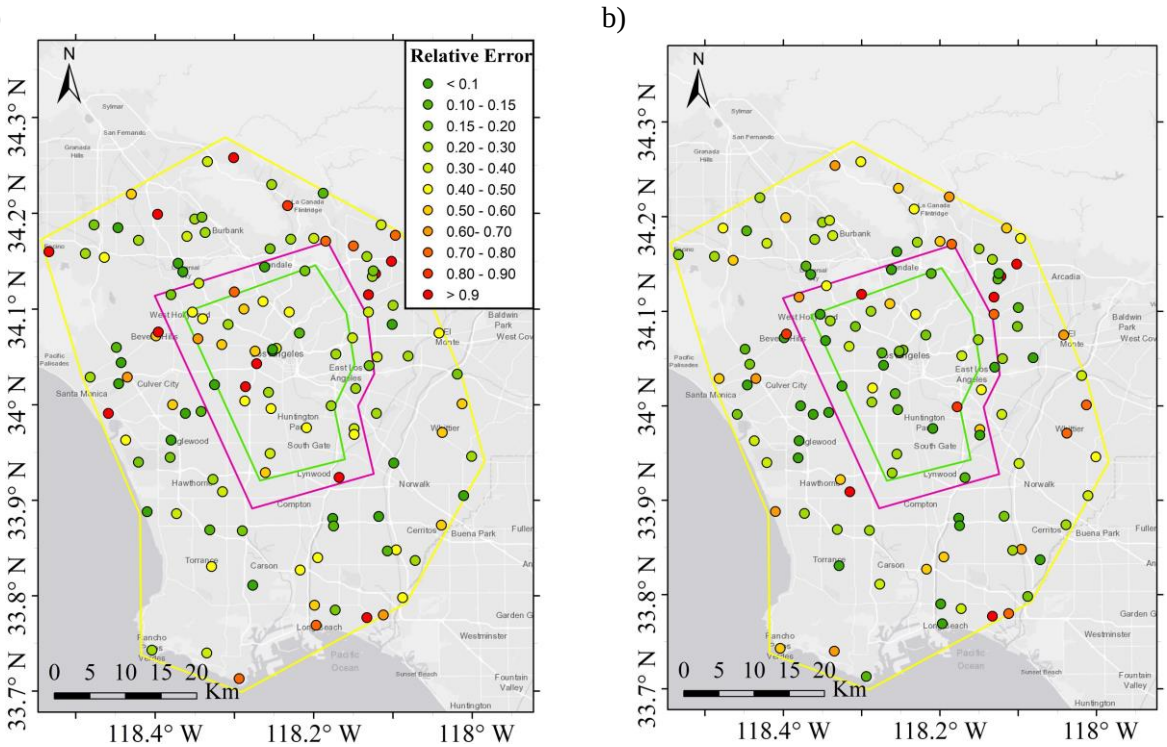


Figure 7.9 Distribution of response spectrum relative error at T = 0.4 (s) along NS direction for having a) CISN sites and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

a) b)

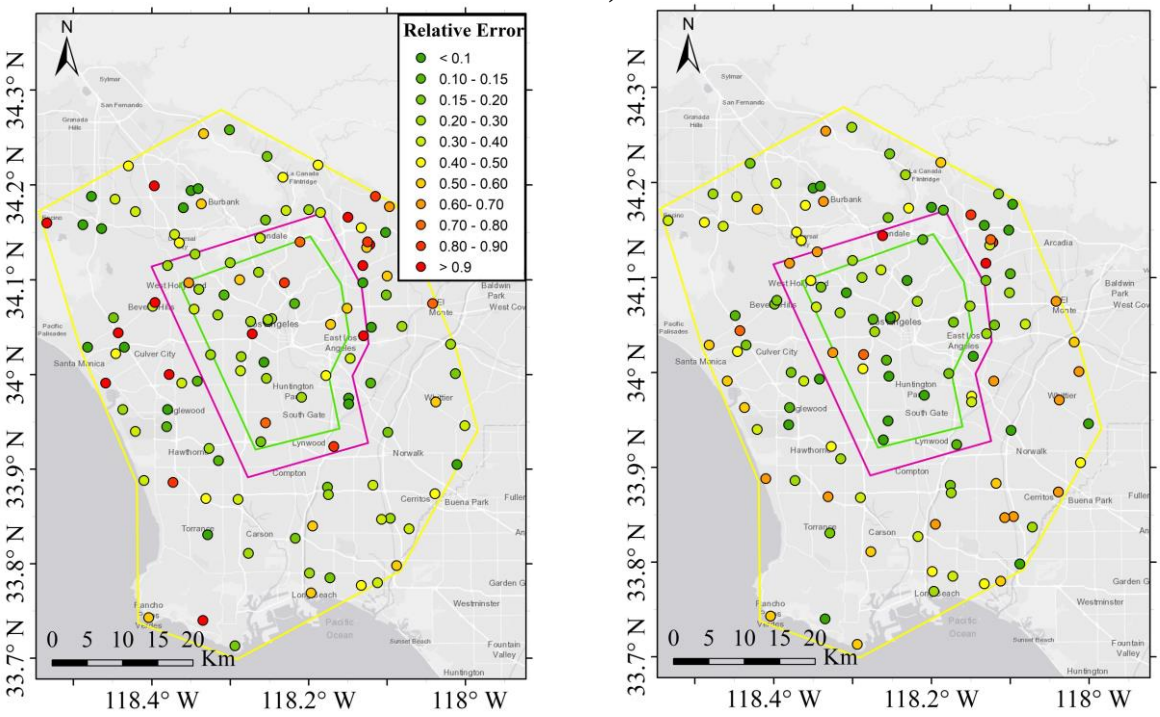


Figure 7.10 Distribution of response spectrum relative error at T = 2.0 (s) along NS direction for having a) CISN sites and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

a) b)

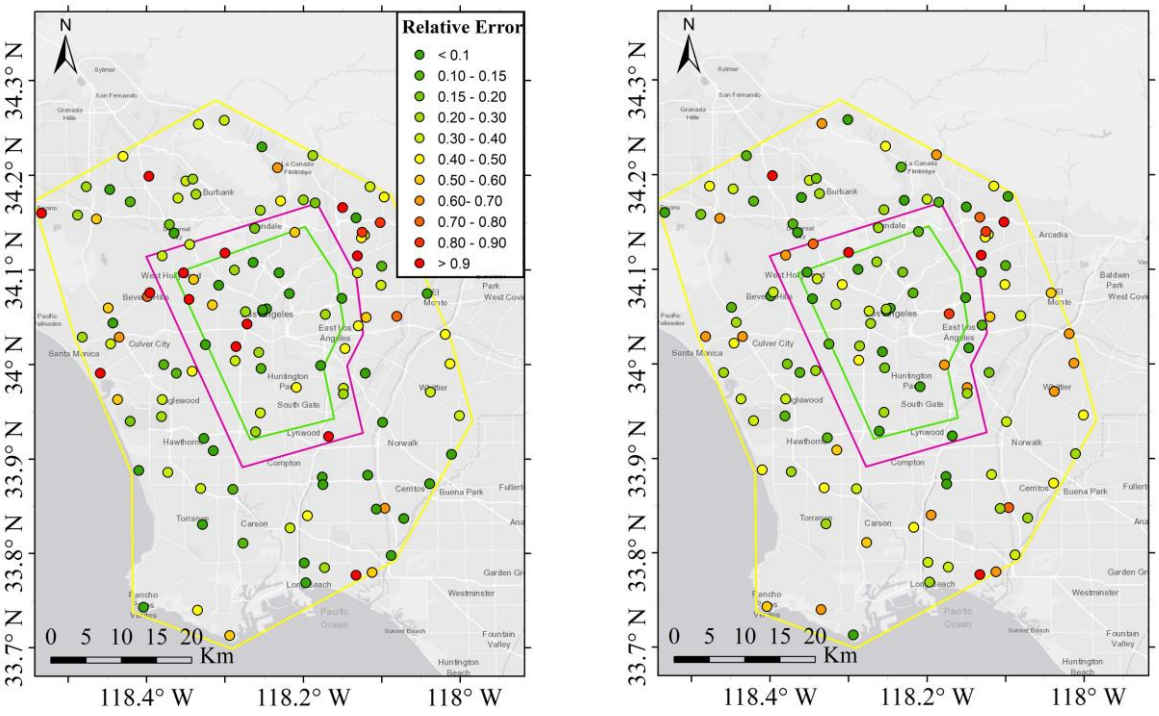


Figure 7.11 Distribution of RotD50 relative error at T = 0.4 (s) along NS direction for having a) CISN sites and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

a) b)

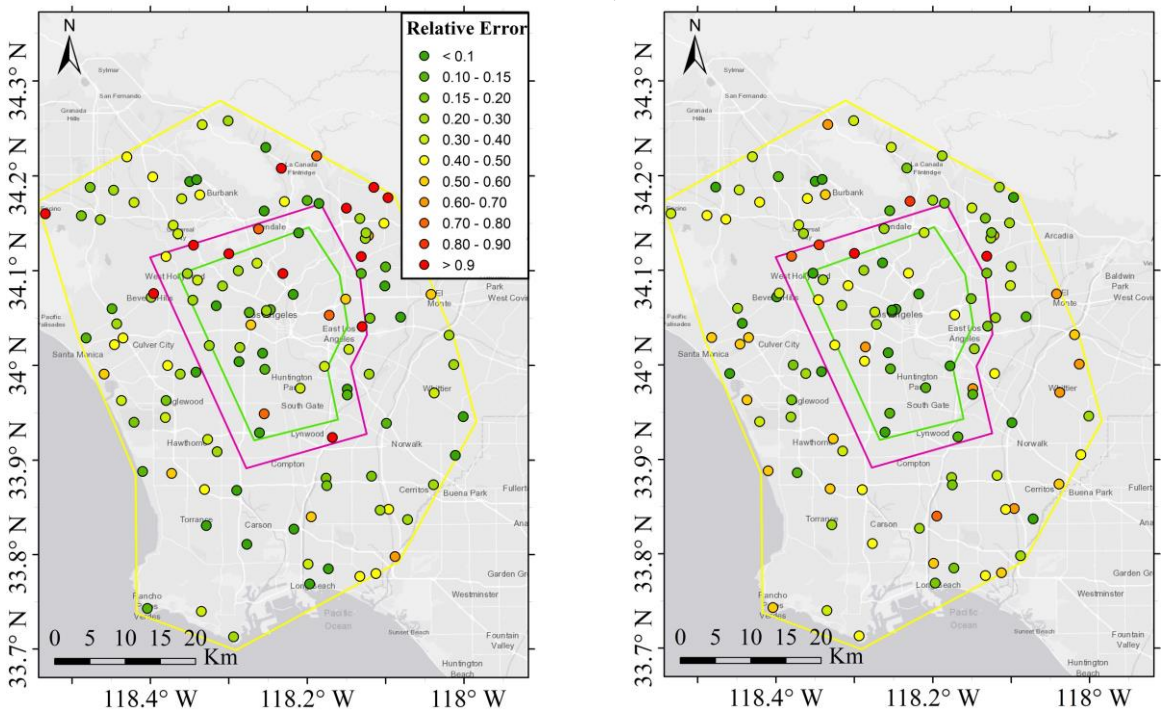


Figure 7.12 Distribution of RotD50 relative error at T = 2.0 (s) along NS direction for having a) CISN sites and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

Table 7.4 . The predicted motions' response spectrum relative error at T = 0.4 (s) along both horizontal directions as well as RotD50 spectrum within different Domains

Domain	Observations	East-West		North-South		RotD50	
		Average NRMSE	Error Reduction*	Average NRMSE	Error Reduction	Average NRMSE	Error Reduction
Inner	CISN + CSN	0.29	48%	0.24	53%	0.23	53%
	CISN	0.56		0.51		0.49	
Middle	CISN + CSN	0.57	11%	0.36	8%	0.36	22%
	CISN	0.64		0.39		0.46	
Exterior	CISN + CSN	0.38	15%	0.39	11%	0.38	2%
	CISN	0.45		0.44		0.39	

* Error Reduction shows the reduction in the prediction error in two scenarios (i.e., due to the added CSN sites)

Table 7.5 . The predicted motions' response spectrum relative error at T = 2.0 (s) along both horizontal directions as well as RotD50 spectrum within different Domains

Domain	Observations	East-West		North-South		RotD50	
		Average NRMSE	Error Reduction	Average NRMSE	Error Reduction	Average NRMSE	Error Reduction
Inner	CISN + CSN	0.25	30%	0.21	48%	0.22	31%
	CISN	0.36		0.40		0.32	
Middle	CISN + CSN	0.62	25%	0.42	2%	0.51	25%
	CISN	0.83		0.43		0.68	
Exterior	CISN + CSN	0.37	14%	0.37	16%	0.35	5%
	CISN	0.43		0.44		0.37	

We also demonstrate the logarithmic standard deviation of one hundred generated ground motion realizations' PSA at two periods, T = 0.4 (s) and T = 2.0 (s), before and after added CSN observations in Tables 7.6 and 7.7, respectively.

Table 7.6 . The predicted motions' response spectrum log normal standard deviation at T = 0.4 (s) along EW and NS directions within different Domains

Domain	Observations	East-West		North-South	
		Average *ln* Std.[1]	Uncertainty Reduction[2]	Average *ln* Std.	Uncertainty Reduction
Inner	CISN + CSN	0.52	-2%	0.49	2%
	CISN	0.51		0.50	
Middle	CISN + CSN	0.57	-11%	0.55	-4%
	CISN	0.51		0.53	
Exterior	CISN + CSN	0.56	-10%	0.56	-12%
	CISN	0.51		0.50	

[1] Average of logarithmic standard deviation of PSA at T among all stations
[2] Reduction of average logarithmic standard deviation of PSA at T among all stations due to added CSN sites (negative values mean uncertainty has been increased)

Table 7.7 . The predicted motions' response spectrum log normal standard deviation at T = 2.0 (s) along EW and NS directions within different Domains

Domain	Observations	East-West		North-South	
		Average *ln* Std.	Uncertainty Reduction	Average *ln* Std.	Uncertainty Reduction
Inner	CISN + CSN	0.49	2%	0.49	8%
	CISN	0.50		0.53	
Middle	CISN + CSN	0.55	-1%	0.56	-3%
	CISN	0.54		0.54	
Exterior	CISN + CSN	0.57	-14%	0.56	-12%
	CISN	0.50		0.50	

As is shown in Table 7.6, the added CSN sites as observation do not necessarily decrease the uncertainty of predicted motions at short periods. It is recognized that the short-periods' uncertainty can even increase for the sites outside the CSN network (middle or exterior domains). The reason is there is higher randomness for recorded short-period content of the ground motions; thus, the added input observation cannot necessarily lead to more certain short-period predictions. On the other hand, the uncertainty of the longer period content of the motions decreases inside the added observation CSN network (inner domain). Yet, the uncertainty of the estimated ground motions in longer periods can even increase for the target sites outside the added observation network as the inner observations cannot provide much information about exterior predictions.

We chose five sites within the inner domain of the CISN network to show the improvement of the predicted results after adding CSN sites as observations. Figure 7.13 shows the selected five sites for result illustration. These five sites are picked as samples to compare the predicted motions'

velocity-time series and RotD50 spectrum with the recorded ones. Figure 7.14 demonstrates the predicted motions' RotD50 response spectrum and velocity time series along EW direction for CISN plus CSN and only CISN sites as observation. It is observable within parts (a) and (b) of Figure 7.14 how the amplitude of the velocity time series is over-predicted for having just CISN sites as observation. The added CSN sites as observation to the GPR model made the predicted ground motion time series match the recorded one. Similarly, the response spectrum of the prediction becomes closer to the recorded ones having more observations from CSN sites, especially within the shorter periods (Figures 7.14b and 7.14c). Readers can refer to Appendix A to observe more results for other test sites within Figure 7.13.

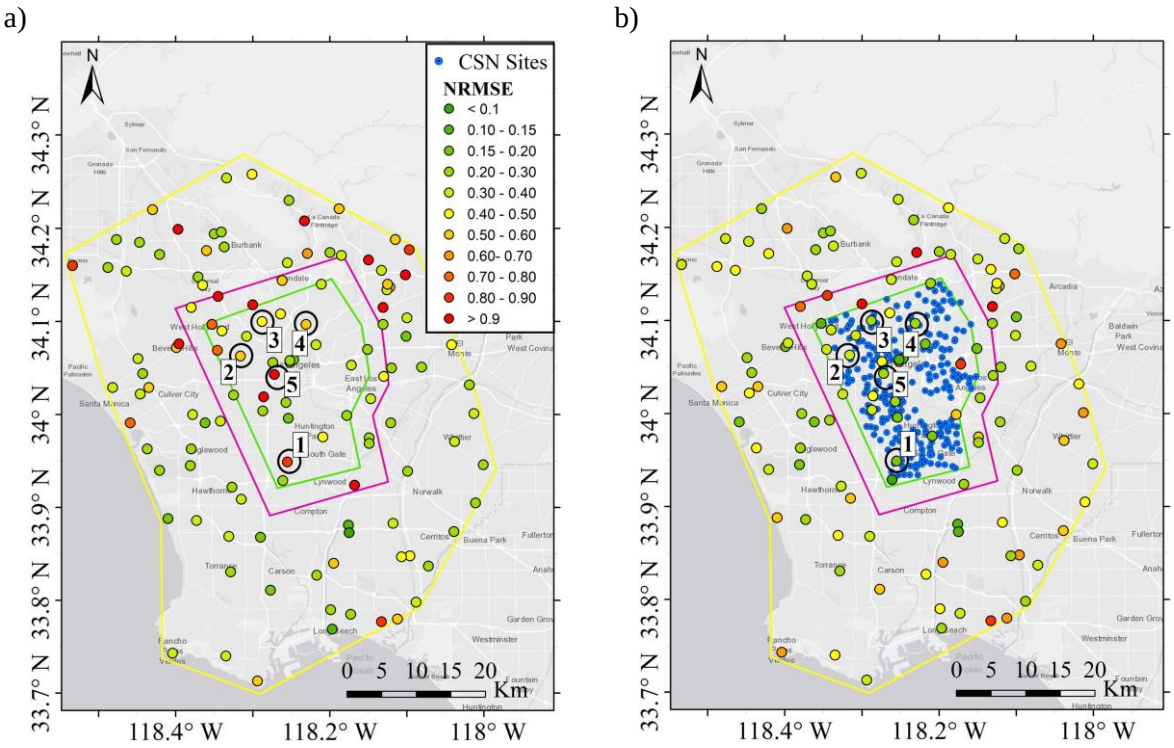


Figure 7.13 Distribution of RotD50 NRMSE for having a) CISN and b) all CISN and CSN sites as observations and the five chosen CISN test sites within the **M**7.1 Ridgecrest earthquake dataset

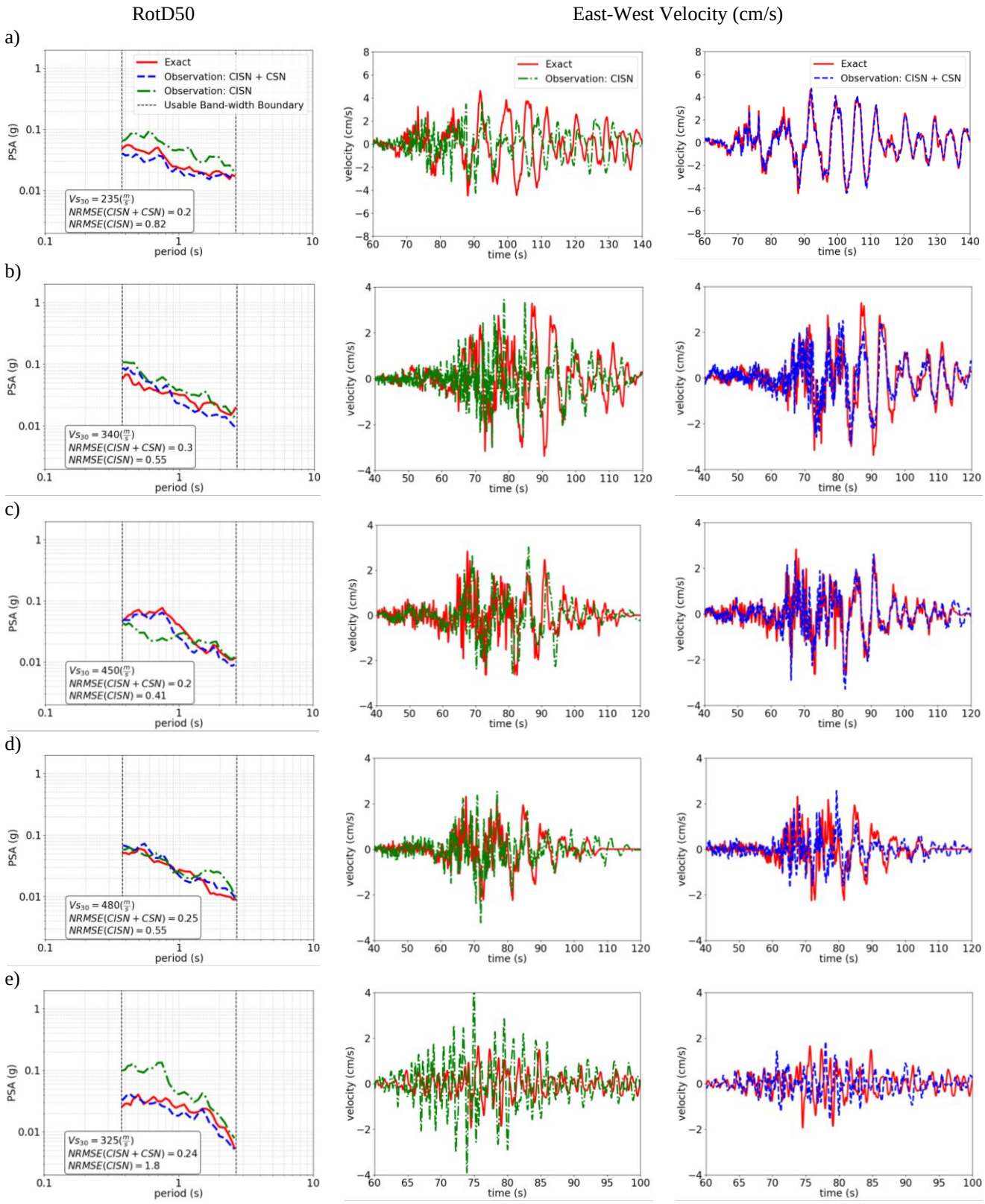


Figure 7.14 The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation as well as the exact motions along East-West direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within the CISN for **M**7.1 Ridgecrest earthquake

7.2. Simulation of Ground Motion for M4.5 South El Monte Earthquake Using CISN and CSN Networks

We investigated the performance of the GPR, having more added observations for the 2020 **M**4.5 South El Monte earthquake that happened on September 18th, 2020. Table 7.8 summarizes the features of the South El Monte earthquake used in this study.

Table 7.8. The 2020 M4.5 South El Monte earthquake features (USGS, 2020)

Date	UTC time	M_w	Epicenter	Depth
September 19th, 2020	06:38:46	4.5	South El Monte	16.9 km

We used 95 and 215 ground-level sites within Los Angeles and recorded the **M**4.5 South El Monte earthquake from CISN and CSN, respectively. It is worth noting the number of stations obtained after removing the stations with processed records with a too narrow usable bandwidth. The reason for filtering out of those stations is that the ultimate predicted motions' usable bandwidth is the mutual usable bandwidth among all observed motions. Therefore, it is required to detect those observations with too narrow bandwidth and remove them. The mutual usable bandwidth among all CSN and CISN stations are (0.11 s – 0.57 s) and (0.10 s – 0.55 s) for EW and NS directions, respectively. The distribution of the CISN and CSN stations over the Los Angeles area is shown in Figure 7.15 a. We used the similar three defined subdomains, Inner, Middle, and Exterior, to allocate the corresponding regularization factor, $\hat{\lambda}$, for each region (Figure 7.15 b).

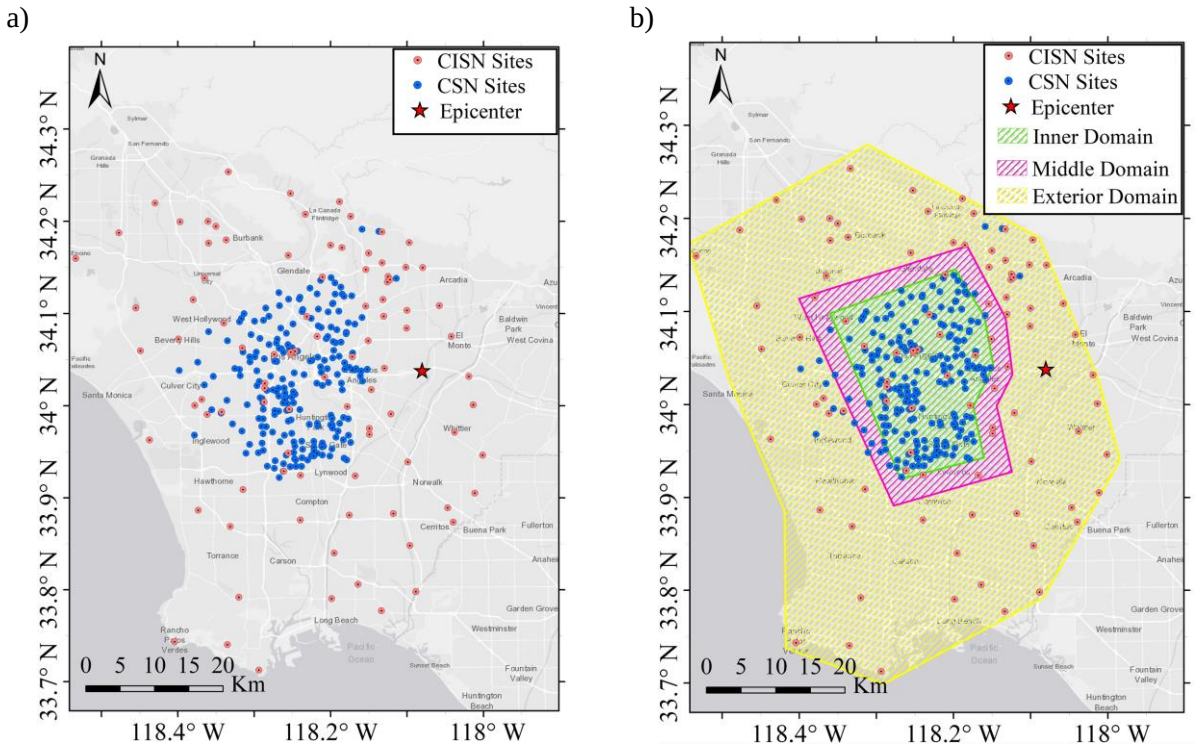


Figure 7.15 a) Distribution of the CSN and CISN sites recorded 2020 **M**4.5 South El Monte earthquake within Los Angeles, and b) Division of CISN network in Los Angeles into three different sub domains with various density of observations

We have implemented two different analyses aiming to estimate the whole ground motion time series at each CISN site location using 1) All other CISN and all CSN sites' motions as observation, and 2) just all other CISN sites as observation. For the first scenario, where we divided the whole CISN network into three different subdomains based on their observation density, we must set three values of $\hat{\lambda}$ for each subdomain using the corresponding density and Table 4.2. On the other hand, we used just one value of $\hat{\lambda}$ for the second scenario as just one observation density is considered for the whole CISN network. Tables 7.9 and 7.10 depict the observation density and implemented $\hat{\lambda}$ (based on Table 4.2) for the target sites at each domain. It is worth mentioning that the required $\hat{\lambda}$ for the inner domain for having both CISN and CSN sites and the $\hat{\lambda}$ for having just CISN sites as observation are obtained using logarithmic interpolation and extrapolation within Table 4.2, respectively.

Table 7.9. The implemented $\hat{\lambda}$ for implementing GPR model at different domains within the CISN network using CISN and CSN sites as observation for 2020 **M**4.5 South El Monte earthquake

Observations	Target Domain	Area (km^2)	Observation Density (site/km^2)	$\hat{\lambda}$
CISN + CSN	Inner	464	0.46	0.08
CISN + CSN	Middle	764	0.30	0.10
CISN + CSN	Outer	3103	0.10	0.20

Table 7.10 . The implemented $\hat{\lambda}$ for implementing GPR model within the CISN network using CISN sites as observation for 2020 **M**4.5 South El Monte earthquake

Observations	Area (km^2)	Observation Density (site/km^2)	$\hat{\lambda}$
CISN	3103	0.03	0.50

Figure 7.16 shows the distribution of CISN sites and the calculated NRMSE between the predicted and recorded motion' RotD50 spectrum for having just CISN sites as observation. These predictions follow the LOO procedure in which for each site as a target site, all the rest CISN stations are used as observation. The average RotD50 spectrum NRMSE among all CISN sites is 0.79. Similarly, the average NRMSE among all sites for prediction along EW and NS directions response spectra are 0.88 and 0.95, respectively.

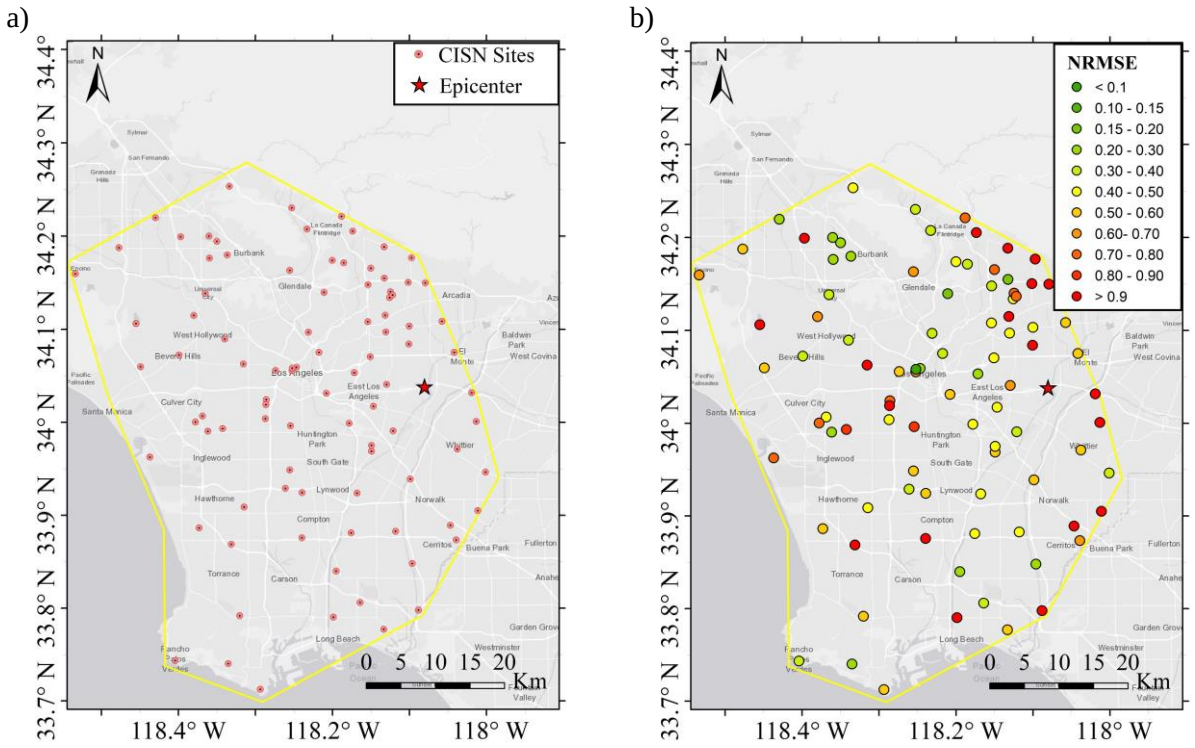


Figure 7.16 Distribution of a) all CISN sites as observation and b) the RotD50 spectrum normalized root mean suqare error at each CISN site as a target site for the **M**4.5 South El Monte earthquake

We estimated the ground motion time series at each CISN site location as a target site using all CISN and all CSN sites as observation. The regularization factor used for prediction is chosen based on the corresponding domain in which the target site is located (cf. Table 7.9). Figure 7.17 shows the distribution of the CISN and CSN stations (observations) and the RotD50 spectrum NRMSE between the predicted and recorded motions using all the rest CISN and CSN sites as observations. The average RotD50 spectrum NRMSE among all CISN sites is 0.76. Similarly, the average NRMSE among all sites for prediction along EW and NS directions response spectra are 0.86 and 0.94, respectively. We plot the response spectra NRMSE distribution along EW, NS, and RotD50 for both scenarios to better compare the predictions' NRMSE within different subdomains before and after added observations in Figures 7.18 through 7.20, respectively.

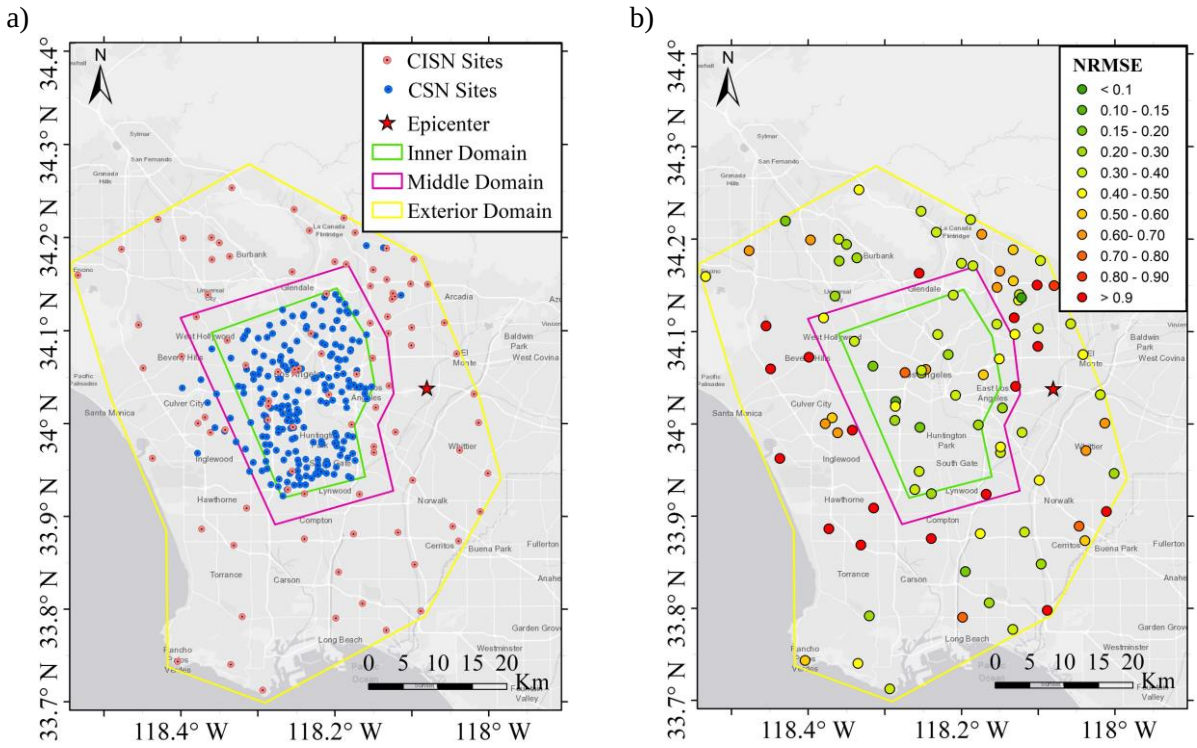


Figure 7.17 Distribution of a) all CISN and CSN sites as observation and b) the RotD50 spectrum normalized root mean suqare error at each CISN site as a target site for the **M**4.5 South El Monte earthquake

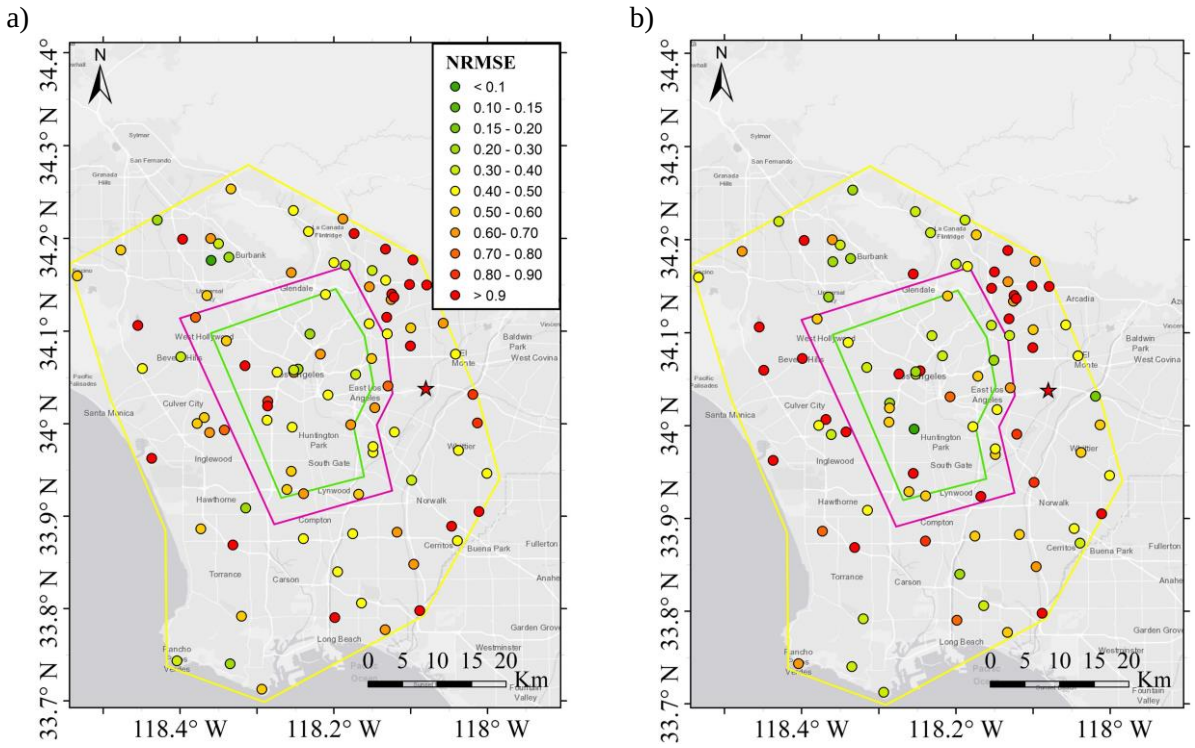


Figure 7.18 Distribution of response spectrum NRMSE along EW direction for having a) CISN sites and b) all CISN and CSN sites as observations within the **M**4.5 South El Monte earthquake dataset

a) b)

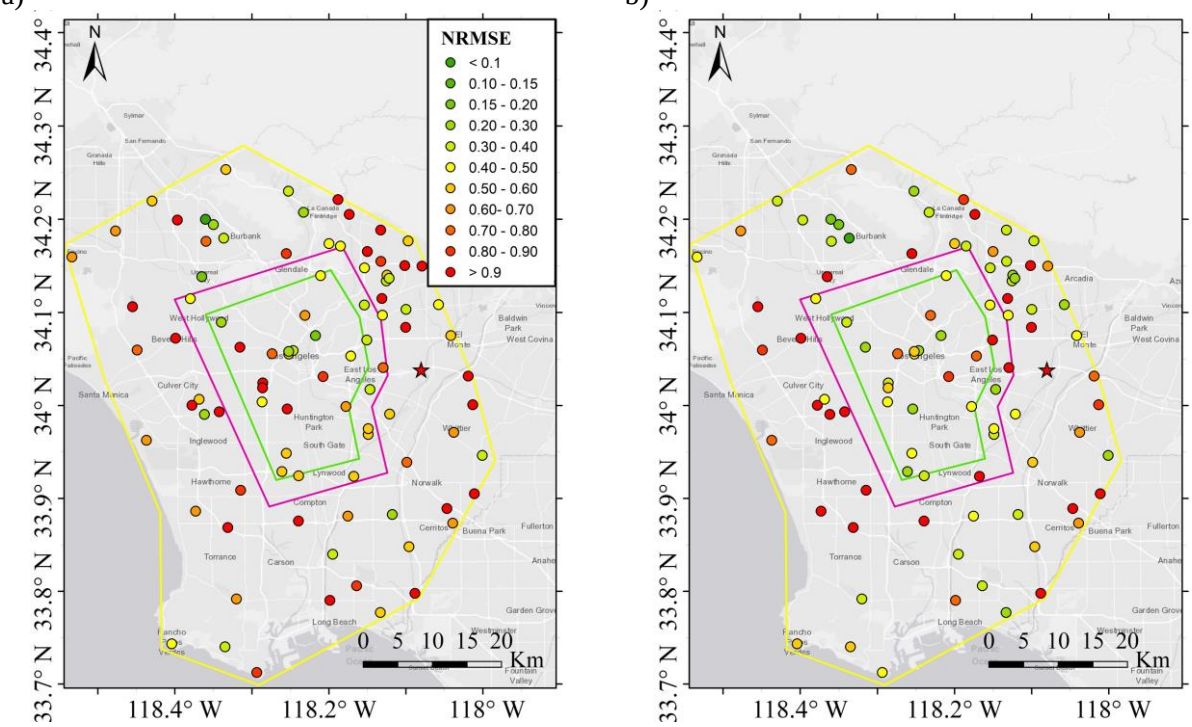


Figure 7.19 Distribution of response spectrum NRMSE along NS direction for having a) CISN sites and b) all CISN and CSN sites as observations within the **M**4.5 South El Monte earthquake dataset

a) b)

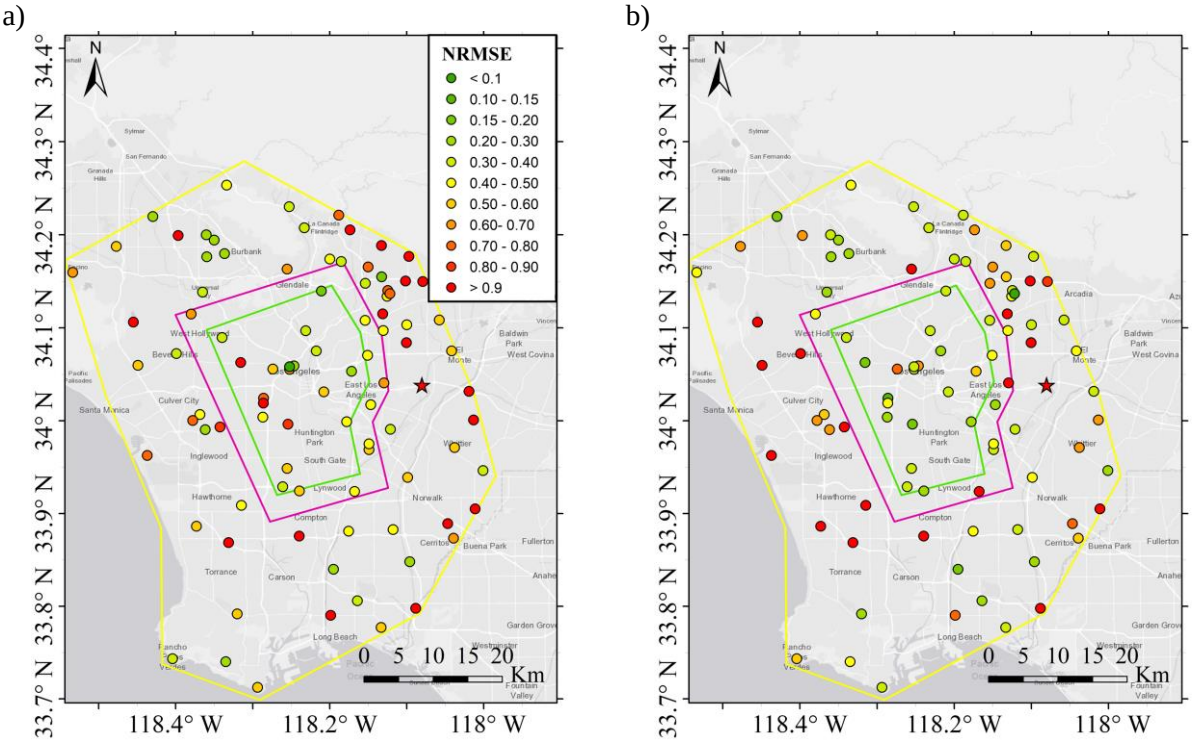


Figure 7.20 Distribution of RotD50 response spectrum NRMSE for having a) CISN sites and b) all CISN and CSN sites as observations within the **M**4.5 South El Monte earthquake dataset

Figure 7.18 shows that the prediction error for the EW direction does not decrease considerably from having a lower number of observations (Figure 7.18a) to having more observations added by CSN (Figure 7.18b), especially within the inner domain. On the other hand, Figure 7.19 demonstrates that the error prediction along the NS direction decreases considerably after having more observation (Figure 7.19). Therefore, adding more observation provides additional informative data for the GPR about the spatial variation of the ground motion along the NS direction over the epicentral region. However, these added observations did not provide significant new information along the EW direction. We summarized the average prediction error (among all test sites) in Table 7.11.

Table 7.11 . The Prediction Error along both horizontal directions as well as RotD50 spectrum within different Domains for the **M**4.5 South El Monte earthquake

Domain	Observations	East-West		North-South		RotD50	
		Average NRMSE	Error Reduction	Average NRMSE	Error Reduction	Average NRMSE	Error Reduction
Inner	CISN + CSN	0.54	7%	0.46	25%	0.35	30%
	CISN	0.58		0.61		0.50	
Middle	CISN + CSN	0.60	-3%	0.60	-20%	0.50	4%
	CISN	0.58		0.50		0.52	
Outer	CISN + CSN	0.98	2%	1.10	0%	0.90	0%
	CISN	1.0		1.10		0.90	

Table 7.11 illustrates how the NS direction experienced a higher error reduction due to the added observation with respect to the EW direction. It is recognizable that the average prediction error is almost unchanged along the EW direction within the inner, middle, and exterior domains due to the added CSN observation. The error reduction for the NS direction within the inner domain is considerable, although the average error reduction within the middle and exterior domains are almost negligible. It is worth mentioning that the middle domain for the **M**4.5 South El Monte earthquake dataset includes nine stations, which could affect the average prediction results to make a reliable conclusion, yet the RotD50 prediction error clearly illustrates that the added CSN sites as observation mostly improved the GPR model's performance within the inner domain and rarely for the middle and exterior domains.

We generated one hundred random realizations of ground motions at each target site (within the CISN network) for both scenarios, with and without CSN observations. These generated random realizations enabled us to have estimated motions' uncertainty (standard deviation of response spectra) at different periods. Table 7.12 shows the logarithmic standard deviation of predicted motions' response spectrum at the period $T = 0.5$ (s), before and after added CSN observations. It should be noted that the longest usable period that we are able to investigate its uncertainty is T=0.5 (s), yet, it is observed from Figures 6.10a and 6.11a that the uncertainty of the estimated motions at the short period ($T = 0.5$ s) is not much affected by the number of observations.

Table 7.12. The predicted motions' response spectrum log normal standard deviation at $T = 0.5$ (s) along EW and NS directions within different Domains for the **M**4.5 South El Monte earthquake

Domain	Observations	East-West		North-South	
		Average *ln* Std.[1]	Uncertainty Reduction	Average *ln* Std.	Uncertainty Reduction
Inner	CISN + CSN	0.53	-12%	0.54	-10%
	CISN	0.47		0.49	
Middle	CISN + CSN	0.56	-3%	0.56	-3%
	CISN	0.54		0.54	
Outer	CISN + CSN	0.57	-11%	0.55	-10%
	CISN	0.51		0.50	

[1] Average of logarithmic standard deviation of PSA at *T* among all stations

Table 7.12 claims that the added observation from CSN causes a small amount of increase in the uncertainty of the predicted motions at the short period content (at 2.0 Hz frequency). This demonstrates that the added observation did not affect the prediction uncertainty of short-length waves. They could also cause uncertainty increment as the added observations are not adequately dense with respect to those wavelengths corresponding to the short period content. Thus, as there is higher randomness in the higher frequency content of the motions, the added observation cannot necessarily improve the uncertainty of the predicted content. This observation is consistent with the previous observation from the 2019 **M**7.1 Ridgecrest earthquake dataset. In general, the added observation could decrease the uncertainty of the low frequencies content prediction within the added observation network (Table 7.7). However, such an effect cannot necessarily be expected for the higher frequencies' prediction (Tables 7.6 and 7.12).

We chose five stations within the inner domain of the CISN (Figure 7.21) to illustrate the predicted results for both scenarios with and without the added CSN sites as additional observations. These five sites are chosen just as samples to compare the predicted motions' velocity time series and PSA spectrum along the NS direction with the corresponding recorded (exact) ones. Although there are 95 sites within all inner, middle, and outer regions, we chose five of them within the inner domain where the added CSN sites exist. Figure 7.22 demonstrates the predicted motions' PSA and velocity time series along the NS direction for both scenarios. It is observable from Figure 7.22 that how the amplitude of the velocity-time series becomes closer to the exact one for having more observations from CSN.

a) b)

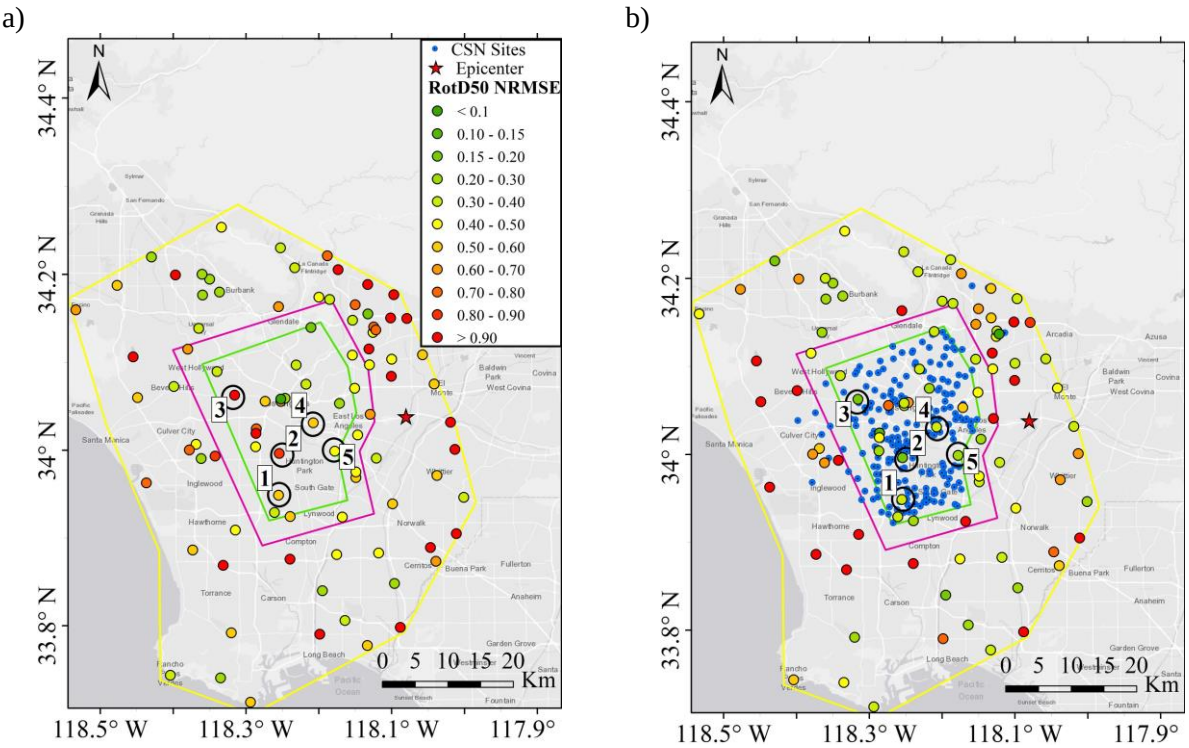


Figure 7.21 Distribution of RotD50 NRMSE for having a) CISN and b) all CISN and CSN sites as observations and the five chosen CISN test sites within the **M**4.5 South El Monte earthquake dataset

It is seen in Figure 7.21 that the majority of the CISN test sites' predicted results improved for those which are located inside the inner domain. However, there are a few sites where their predicted results either did not change considerably or worsened with respect to the case having just CISN observation. The readers can refer to Appendix B to see the prediction results for other CISN test sites. It is worth noting that the predicted ground motions along NS improved more than the other perpendicular direction, EW, within the inner domain.

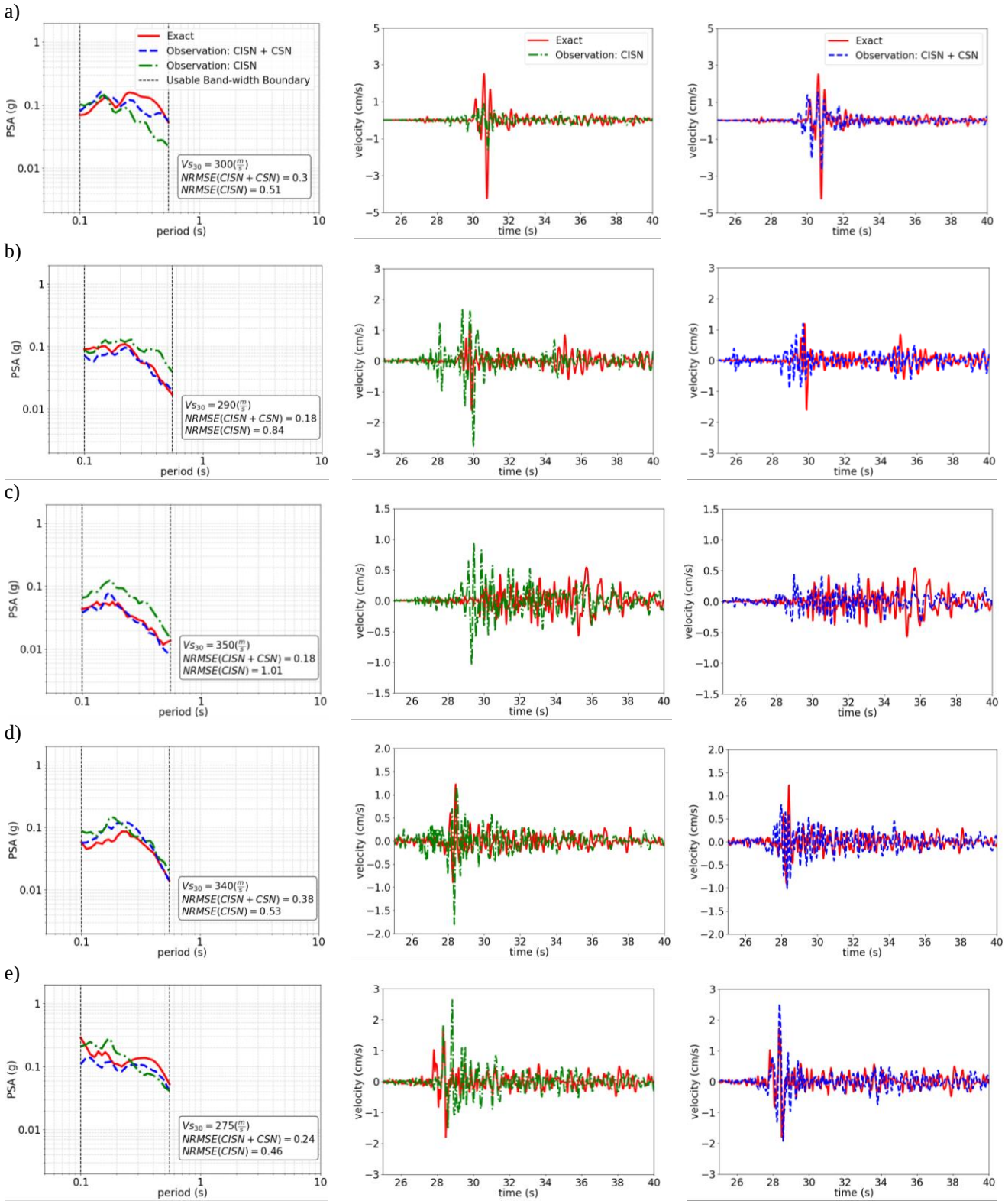


Figure 7.22 The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation as well as the exact motions along North-South direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within the CISN for **M**4.5 South El Monte earthquake

Concluding Remarks

A fine-tuned optimized hyperparameter of the GPR model, the regularization factor, is introduced based on the observation density within the target network. This hyperparameter enables the users to implement the GPR model for various conditions with a different number of observations. It is shown that the optimized regularization factor, $\hat{\lambda}$, is smaller for a region with higher observation density compared to those with lower densities. We also introduced a methodology to generate samples of ground motion realizations rather than one estimated ground motion using mean DFT coefficients for the target sites. The generated samples of ground motions are shown for the **M**7.0 Hayward fault scenario earthquake physics-based simulated dataset and the 2019 **M**7.1 Ridgecrest earthquake.

We investigated the uncertainty of the predicted motions within the 2019 **M**7.1 Ridgecrest earthquake dataset using the randomly generated ground motions. It was observed that the number of observations close to the target site plays a vital role in the prediction error, especially at long periods. The prediction error generally was higher for the target sites with lower observations close to them. In addition, the higher density of observation close to the target site can decrease the uncertainty of the predicted motions over long periods. However, the estimated short period content's uncertainty is less affected by the number of observations close to the target location. It was observed that the recorded motions response spectra could be captured by the 68% confidence interval of the predicted motion's spectrum for the sites with a higher number of observations close to them. It was also shown that the higher uncertainty variation for the estimated site condition at the target sites and its neighboring observed sites could increase the prediction error and uncertainty. We concluded that the slope variation of the target site and the closest neighbors do not have a significant effect on the prediction error.

The effect of added observations from other seismic networks on the prediction error and uncertainty is also studied. To do so, we predicted the ground motions of the 2019 **M**7.1 Ridgecrest and 2020 **M**4.5 South El Monte earthquakes recorded by CISN (within Los Angeles) sites using two sets of observation; 1) The whole CISN dataset and 2) The whole CISN and CSN datasets. The results illustrated that the prediction for the target sites located within the added observation network could considerably improve. This improvement occurred at both horizontal components for the Ridgecrest earthquake. The prediction error reduction due to the added network happened along one horizontal component (NS) more considerably for the **M**4.5 South El Monte earthquake, while the other perpendicular direction (EW) experienced less improvement. Both earthquakes showed that the prediction uncertainty at short periods might not change significantly due to the added observations from other networks.

Appendix A

The estimated ground motion results for 2019 **M**7.1 Ridgecrest earthquake dataset at CISN sites using 1) just all CISN sites and 2) all CSN and CISN sites as observation are shown in Figure A2. This is the test results for 10 sites that are not shown in Figures 7.13 and 7.14.

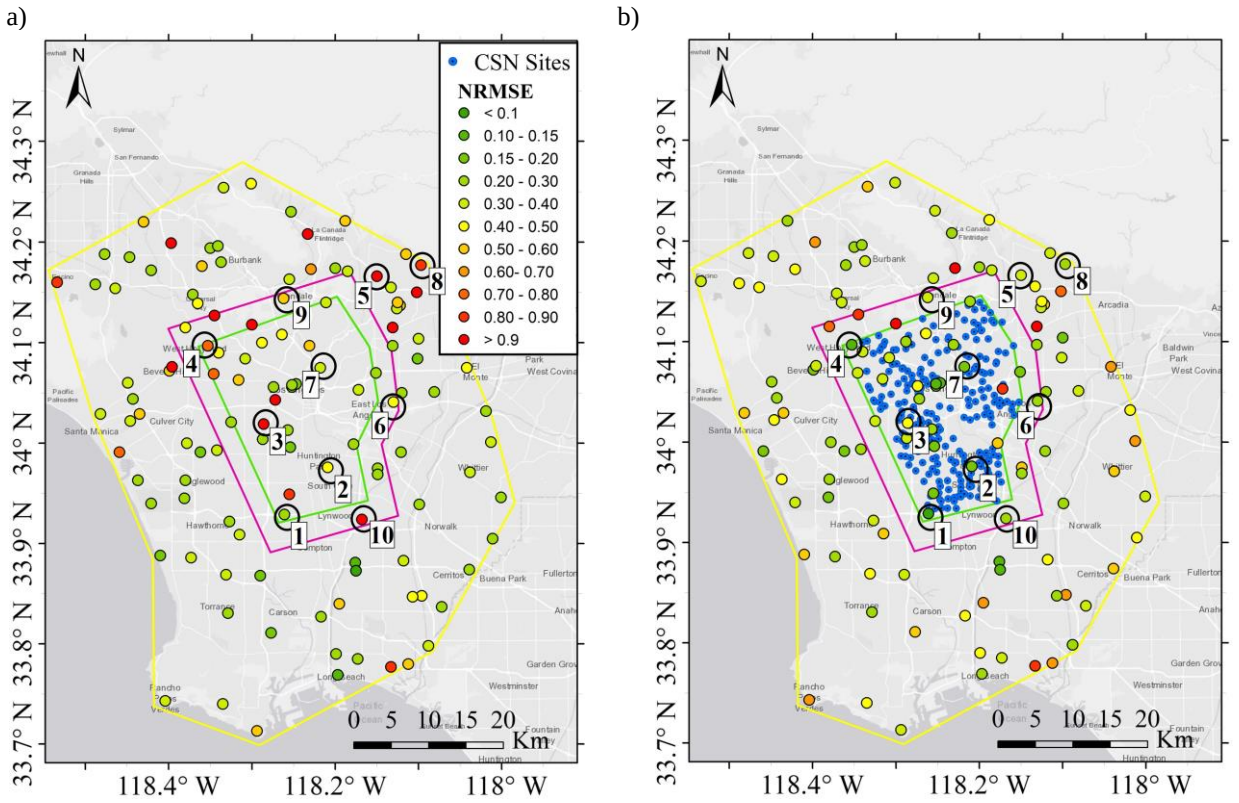
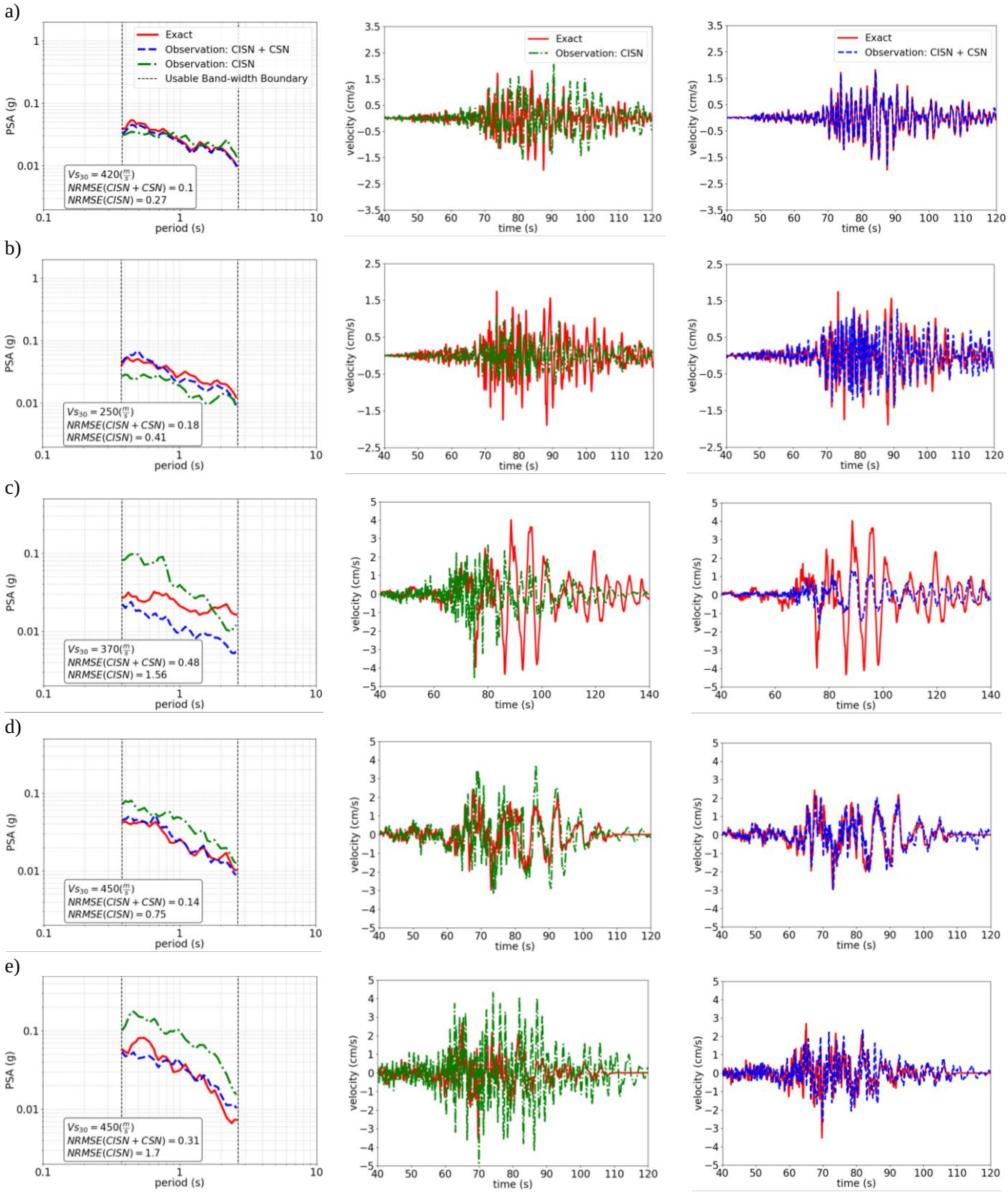


Figure A1. Distribution of RotD50 NRMSE for having a) just CISN sites and b) all CISN and CSN sites as observations and ten chosen CISN test sites within the **M**7.1 Ridgecrest earthquake dataset


RotD50
East-West Velocity (cm/s)
a)
b)
c)
d)
e)
Exact
Observation: CISN + CSN
Observation: CISN
Usable Band-width Boundary
$Vs_{30} = 420(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.1$
$NRMSE(CISN) = 0.27$
$Vs_{30} = 250(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.18$
$NRMSE(CISN) = 0.41$
$Vs_{30} = 370(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.48$
$NRMSE(CISN) = 1.56$
$Vs_{30} = 450(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.14$
$NRMSE(CISN) = 0.75$
$Vs_{30} = 450(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.31$
$NRMSE(CISN) = 1.7$
PSA (g)
period (s)
velocity (cm/s)
time (s)

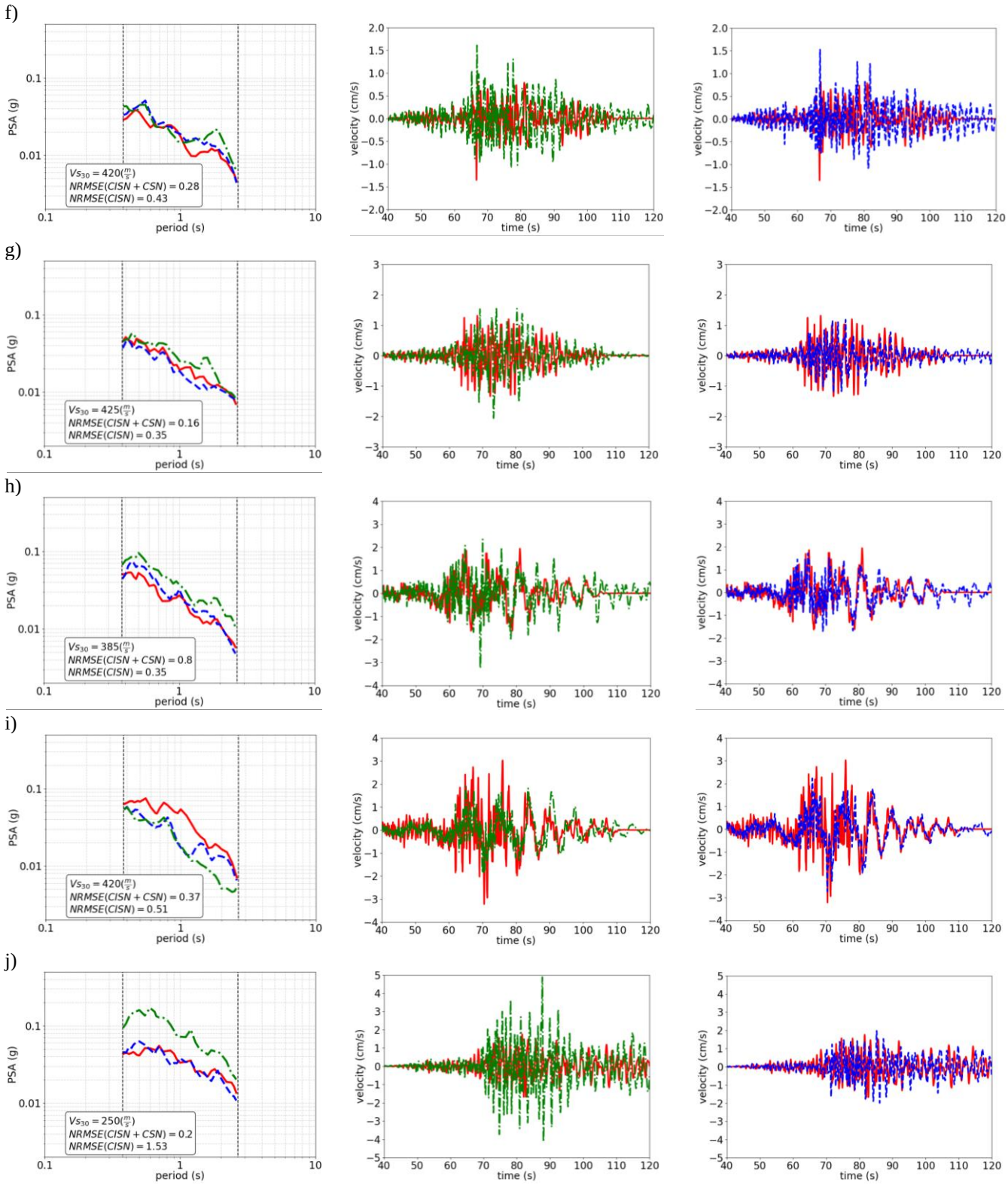


Figure A2. The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation and exact motions along East-West direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, e) No. 5, f) No. 6, g) No. 7, h) No. 8, i) No. 9, and j) No. 10 within CISN network for **M**7.1 Ridgecrest earthquake

Appendix B

The estimated ground motion results for 2020 **M**4.5 South El Monte earthquake dataset at CISN site locations using with 1) just all CISN sites and 2) all CSN and CISN sites shown in (Figure B1) are depicted in the Figure B2. This is the test results for 10 sites that are not shown in Figures 7.21 and 7.22.

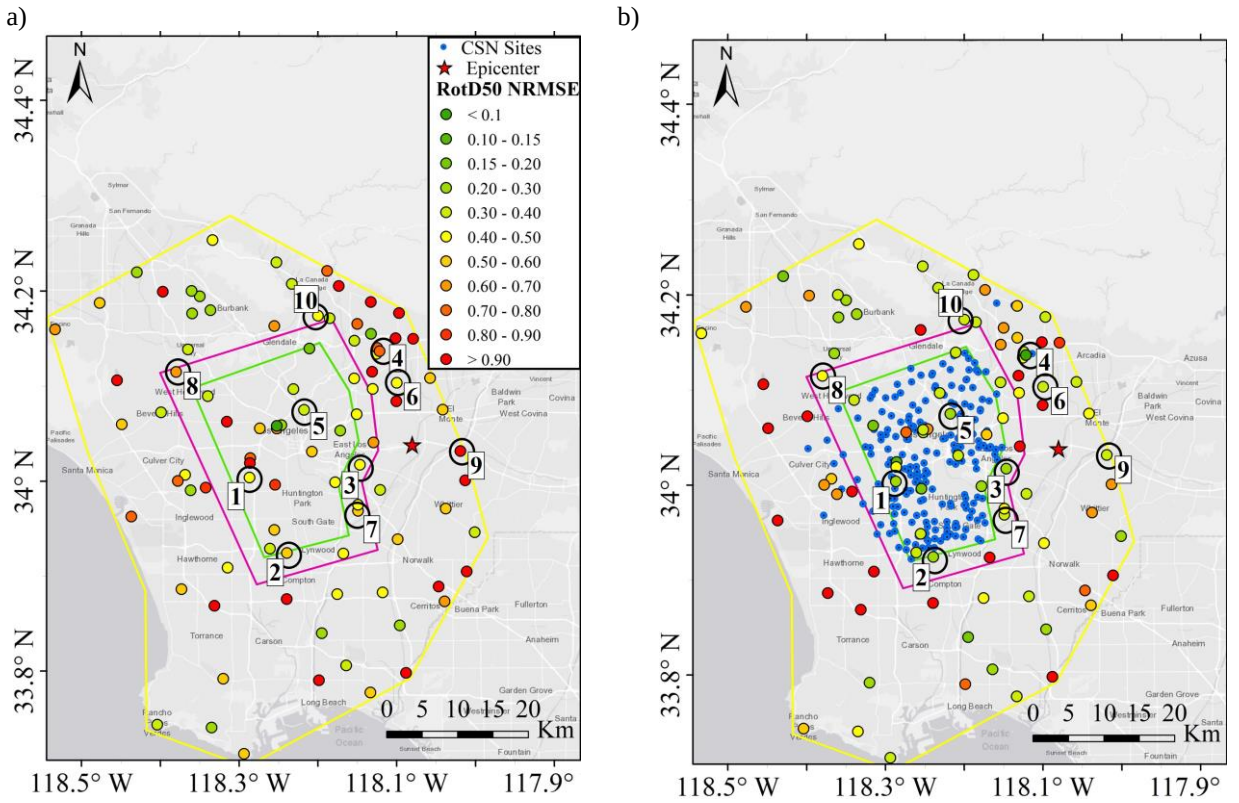
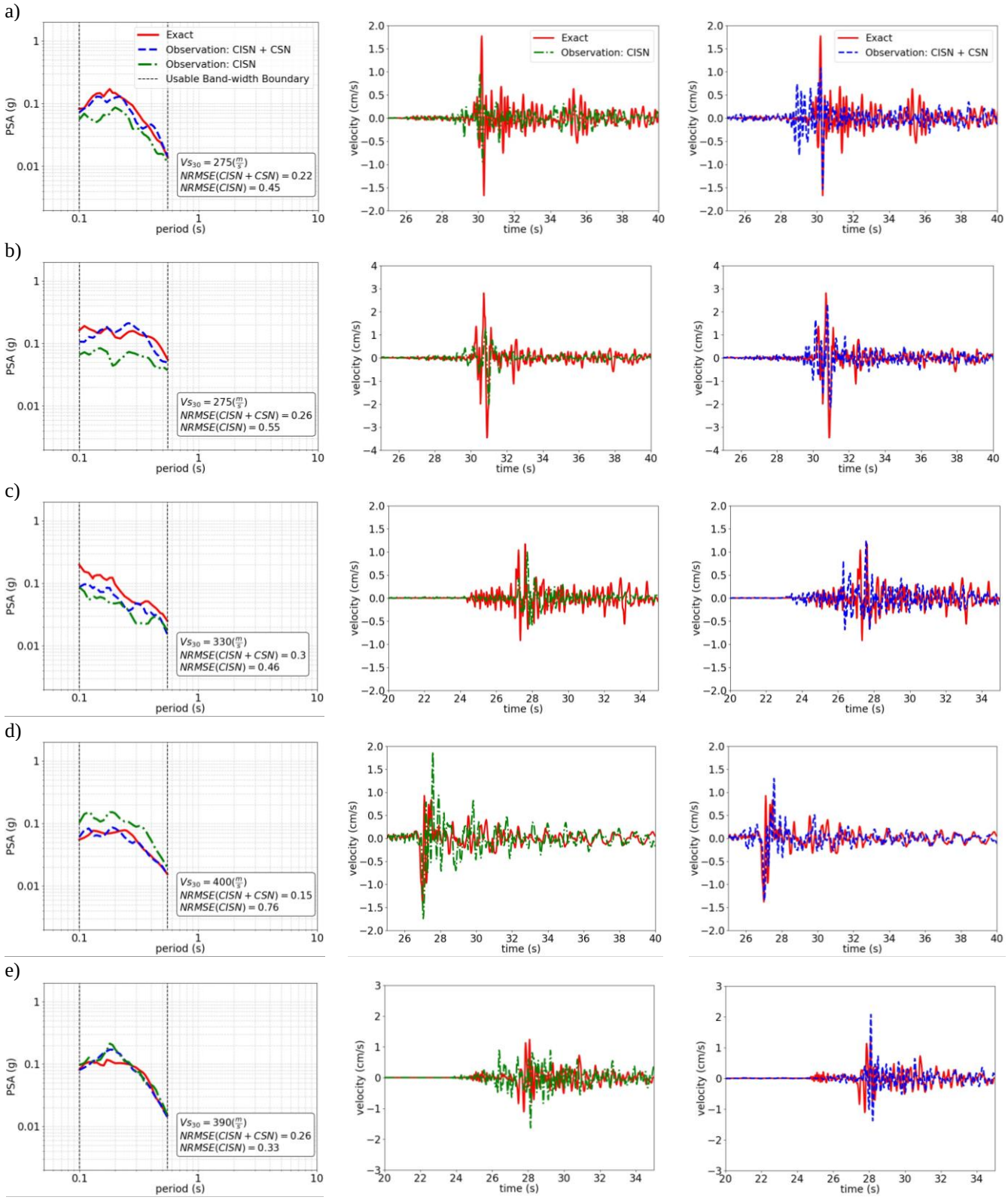


Figure B1. Distribution of RotD50 NRMSE for having a) just CISN sites and b) all CISN and CSN sites as observations and ten chosen CISN test sites within the **M**4.5 South El Monte earthquake dataset

RotD50 North-South Velocity (cm/s)



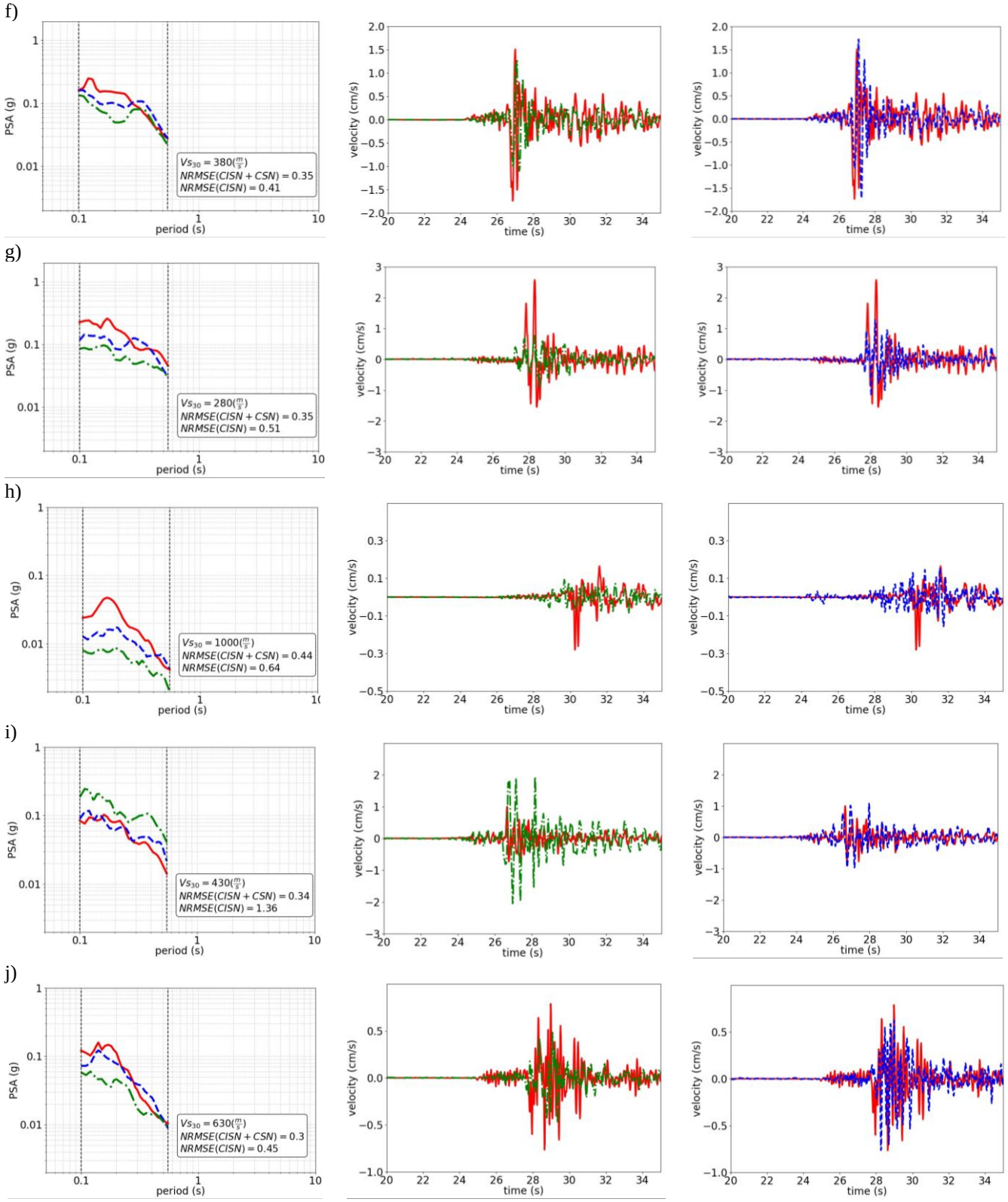


Figure B2. The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation and exact motions along North-South direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, e) No. 5, f) No. 6, g) No. 7, h) No. 8, i) No. 9, and j) No. 10 within CISN network for **M**4.5 South El Monte earthquake

References

Aagaard, B.T., Brocher, T.M., Dolenc, D., Dreger, D., Graves, R.W., Harmsen, S., Hartzell, S., Larsen, S. and Zoback, M.L., 2008a. Ground-motion modeling of the 1906 San Francisco earthquake, Part I: Validation using the 1989 Loma Prieta earthquake. *Bulletin of the Seismological Society of America*, *98*(2), pp.989-1011.

Aagaard, B.T., Brocher, T.M., Dolenc, D., Dreger, D., Graves, R.W., Harmsen, S., Hartzell, S., Larsen, S., McCandless, K., Nilsson, S. and Petersson, N.A., 2008b. Ground-motion modeling of the 1906 San Francisco earthquake, part II: Ground-motion estimates for the 1906 earthquake and scenario events. Bulletin of the Seismological Society of America, 98(2), pp.1012-1046.

Abrahamson, N.A., Schneider, J.F. and Stepp, J.C., 1991. Empirical spatial coherency functions for application to soil-structure interaction analyses. Earthquake spectra, 7(1), pp.1-27.

Adanur, S., Altunisik, A.C., Soyluk, K., Dumanoglu, A.A. and Bayraktar, A., 2016. Contribution of local site-effect on the seismic response of suspension bridges to spatially varying ground motions. Earthquakes and Structures, 10(5), pp.1233-1251.

Ahdi, S.K., Mazzoni, S., Kishida, T., Wang, P., Nweke, C.C., Kuehn, N.M., Contreras, V., Rowshandel, B., Stewart, J.P. and Bozorgnia, Y., 2020. Engineering characteristics of ground motions recorded in the 2019 Ridgecrest earthquake sequence. Bulletin of the Seismological Society of America, 110(4), pp.1474-1494.

Ancheta, T.D., Darragh, R.B., Stewart, J.P., Seyhan, E., Silva, W.J., Chiou, B.S.J., Wooddell, K.E., Graves, R.W., Kottke, A.R., Boore, D.M. and Kishida, T., 2014. NGA-West2 database. Earthquake Spectra, 30(3), pp.989-1005.

Atkinson, G.M. and Assatourians, K., 2015. Implementation and validation of EXSIM (a stochastic finite-fault ground-motion simulation algorithm) on the SCEC broadband platform. Seismological Research Letters, 86(1), pp.48-60.

Baker, J.W. and Chen, Y., 2020. Ground motion spatial correlation fitting methods and estimation uncertainty. Earthquake Engineering & Structural Dynamics, 49(15), pp.1662-1681.

Bayless, J. and Abrahamson, N.A., 2019. An empirical model for the interfrequency correlation of epsilon for Fourier amplitude spectra. Bulletin of the Seismological Society of America, 109(3), pp.1058-1070.

Boore, D.M., 2010. Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. Bulletin of the Seismological Society of America, 100(4), pp.1830-1835.

Boore, D.M., 2003. Simulation of ground motion using the stochastic method. Pure and applied geophysics, 160(3), pp.635-676.

Chen, Y. and Baker, J.W., 2019. Spatial correlations in CyberShake physics-based ground-motion simulations. Bulletin of the Seismological Society of America, (6), pp.2447-2458.

Clayton, R.W., Kohler, M., Guy, R., Bunn, J., Heaton, T. and Chandy, M., 2020. CSN-LAUSD network: A dense accelerometer network in Los Angeles Schools. Seismological Research Letters, 91(2A), pp.622-630.

Daub, E.G., Arabnejad, H., Mahmood, I. and Groen, D., 2021. Uncertainty quantification of dynamic earthquake rupture simulations. Philosophical Transactions of the Royal Society A, 379(2197), p.20200076.

Fan, J. and Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), pp.1348-1360.

Filippitzis, Filippos, Monica D. Kohler, Thomas H. Heaton, Robert W. Graves, Robert W. Clayton, Richard G. Guy, Julian J. Bunn, and K. Mani Chandy. 2021. “Ground Motions in Urban Los Angeles from the 2019 Ridgecrest Earthquake Sequence.” Earthquake Spectra. https://doi.org/10.1177/87552930211003916.

Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuizen, H.T., Verelst, L. and Wiberg, D., 2008. Global agro-ecological zones assessment for agriculture (GAEZ 2008). IIASA, Laxenburg, Austria and FAO, Rome,

Italy, 10.

Fraser, W.A., Wald, D.J. and Lin, K.W., 2008. Using ShakeMap and ShakeCast to Prioritize Post-Earthquake Dam Inspections. In Geotechnical Earthquake Engineering and Soil Dynamics IV (pp. 1-10).

Gentile, R. and Galasso, C., 2020. Gaussian process regression for seismic fragility assessment of building portfolios. Structural Safety, 87, p.101980.

Jayaram, N. and Baker, J.W., 2009. Correlation model for spatially distributed ground-motion intensities. Earthquake Engineering & Structural Dynamics, 38(15), pp.1687-1708.

Kameda, H. and Morikawa, H., 1992. An interpolating stochastic process for simulation of conditional random fields. Probabilistic Engineering Mechanics, 7(4), pp.243-254.

Kohler, M.D., Magistrale, H. and Clayton, R.W., 2003. Mantle heterogeneities and the SCEC reference three-dimensional seismic velocity model version 3. Bulletin of the Seismological Society of America, 93(2), pp.757-774.

Kohler, M.D., Filippitzis, F., Heaton, T., Clayton, R.W., Guy, R., Bunn, J. and Chandy, K.M., 2020. 2019 Ridgecrest Earthquake Reveals Areas of Los Angeles That Amplify Shaking of High-Rises. Seismological Society of America, 91(6), pp.3370-3380.

Konakli, K. and Der Kiureghian, A., 2012. Simulation of spatially varying ground motions including incoherence, wave-passage and differential site-response effects. Earthquake Engineering & Structural Dynamics, 41(3), pp.495-513.

Kwon, O.S. and Elnashai, A., 2006. The effect of material and ground motion uncertainty on the seismic vulnerability curves of RC structure. Engineering structures, 28(2), pp.289-303.

Li, R. and Sudjianto, A., 2005. Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. Technometrics, 47(2), pp.111-120.

Lin, K., Wald, D.J., Kircher, C.A., Slosky, D., Jaiswal, K. and Luco, N., 2018, June. USGS shakecast system advancements. In 11th National Conference on Earthquake Engineering, pp. 3458-3468.

Loos, S., Lallemant, D., Baker, J., McCaughey, J., Yun, S.H., Budhathoki, N., Khan, F. and Singh, R., 2020. G-DIF: A geospatial data integration framework to rapidly estimate post-earthquake damage. Earthquake Spectra, 36(4), pp.1695-1718.

Mangalathu, S. and Jeon, J.S., 2020. Ground motion-dependent rapid damage assessment of structures based on wavelet transform and image analysis techniques. Journal of Structural Engineering, 146(11).

Nweke, C.C., Wang, P., Brandenberg, S.J. and Stewart, J.P., 2018. Reconsidering basin effects in ergodic site response models.

Oppenheim, Alan V., Alan S. Willsky, and S. Hamid Nawab. “Signals and systems Prentice Hall.” Inc., Upper Saddle River, New Jersey 7458 (1997).

Otake, R., Kurima, J., Goto, H. and Sawada, S., 2020. Deep Learning Model for Spatial Interpolation of Real-Time Seismic Intensity. Seismological Society of America, 91(6), pp.3433-3443.

Petrone, F., Abrahamson, N., McCallen, D., Pitarka, A. and Rodgers, A., 2021. Engineering evaluation of the EQSIM simulated ground-motion database: The San Francisco Bay Area region. Earthquake Engineering & Structural Dynamics, 50(15), pp.3939-3961.

Rasmussen, Carl Edward, and Christopher KI Williams. “Gaussian Processes for Machine Learning”, the MIT Press. Cambridge, Mass (2006).

Rodda, G.K. and Basu, D., 2018. Coherency model for translational and rotational ground motions. Bulletin of Earthquake Engineering, 16(7), pp.2687-2710.

Rodgers, A.J., Anders Petersson, N., Pitarka, A., McCallen, D.B., Sjogreen, B. and Abrahamson, N., 2019. Broadband (0–5 Hz) fully deterministic 3D ground-motion simulations of a magnitude 7.0 Hayward fault earthquake: Comparison with empirical ground-motion models and 3D path and site effects from source normalized intensities. Seismological Research Letters, 90(3), pp.1268-1284.

Ryerson, F.J., Bartling, W., Dietrich, W., DePaolo, D.J., Knudsen, K., Rodgers, A., Zafir, Z. and Darling, G., 2021. External Review Committee Report on the California Geological Survey (No. LLNL-TR-827950). Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States).

Shao, J., 1993. Linear model selection by cross-validation. Journal of the American statistical Association, 88(422), pp.486-494.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and computers in simulation, 55(1-3), pp.271-280.

Southern California Earthquake Data Center (2022). [Online], avail- able at https://service.scedc.caltech.edu/SCSNStationMap/station.html (last accessed June 2022).

Tamhidi, A, N Kuehn, S F Ghahari, A J Rodgers, E Taciroglu, and Y Bozorgnia. 2022. "Earthquake Ground Motion Conditioned Simulation Using Sparsely Distributed Observed Motions for Analysis and Design of Lifeline Structures," In Lifelines 2022 Conference, American Society of Civil Engineers.

Tamhidi, A., Kuehn, N. M., Kohler, M. D., Ghahari, F., Taciroglu, E., & Bozorgnia, Y., "Ground-Motion Time-Series Interpolation within the Community Seismic Network using Gaussian Process Regression: Application to the 2019 Ridgecrest Earthquake". Poster Presentation at 2020 SCEC Annual Meeting, (2020).

Tamhidi, A., Kuehn, N., Bozorgnia, Y., Taciroglu, E., & Kishida, T. "Prediction of Ground-Motion Time-Series at an arbitrary location using Gaussian Process Interpolation: Application to the Ridgecrest Earthquake". Poster Presentation at 2019 SCEC Annual Meeting, (2019).

Tamhidi, Aidin, Nicolas Kuehn, S. Farid Ghahari, Arthur J. Rodgers, Monica D. Kohler, Ertugrul Taciroglu, and Yousef Bozorgnia. 2021. "Conditioned Simulation of Ground-Motion Time Series at Uninstrumented Sites Using Gaussian Process Regression." Bulletin of the Seismological Society of America, 1–17. https://doi.org/10.1785/0120210054.

Tian, L., Gai, X., Qu, B., Li, H. and Zhang, P., 2016. Influence of spatial variation of ground motions on dynamic responses of supporting towers of overhead electricity transmission systems: An experimental study. Engineering Structures, 128, pp.67-81.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp.267-288.

U.S. Geologic Survey (2018). 3-D geologic and seismic velocity models of the San Francisco Bay region, available at https://www.usgs.gov/natural-hazards/earthquake-hazards/science/3-d-geologic-and-seismic-velocity-models-san-francisco (last accessed June 2022).

U.S. Geological Survey (2019). M7.1-2019 Ridgecrest Earthquake Sequence, available at https://earthquake.usgs.gov/earthquakes/eventpage/ci38457511/executive (last accessed June 2022.)

U.S. Geological Survey (2020). M4.5-3km WSW of South El Monte, CA, available at https://earthquake.usgs.gov/earthquakes/eventpage/ci38695658/executive (last accessed June 2022.)

Vehtari, A., Gelman, A. and Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing, 27(5), pp.1413-1432.

Wald, D., Lin, K.W., Porter, K. and Turner, L., 2008. ShakeCast: Automating and improving the use of ShakeMap for post-earthquake decision-making and response. Earthquake Spectra, 24(2), pp.533-553.

Worden, C.B., Thompson, E.M., Baker, J.W., Bradley, B.A., Luco, N. and Wald, D.J., 2018. Spatial and spectral interpolation of ground-motion intensity measure observations. Bulletin of the Seismological Society of America, 108(2), pp.866-875.

Zentner, I., 2013. Simulation of non-stationary conditional ground motion fields in the time domain. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 7(1), pp.37-48.

Zerva, A., 2009. Spatial variation of seismic ground motions: modeling and engineering applications. Crc Press.

Zerva, A., Falamarz-Sheikhabadi, M.R. and Poul, M.K., 2018, June. Issues with the use of spatially variable seismic ground motions in engineering applications. In European Conference on Earthquake Engineering

Thessaloniki, Greece (pp. 225-252). Springer, Cham.

Zerva, A. and Zervas, V., 2002. Spatial variation of seismic ground motions: an overview. Appl. Mech. Rev., 55(3), pp.271-297.