UCLA
UCLA Electronic Theses and Dissertations

Title

Earthquake Resilient Smart Cities: A Framework for Collection and Utilization of Highly Granular Field Data for Seismic Performance Characterization of Soft-Story Buildings

Permalink

https://escholarship.org/uc/item/8290p1kg

Author

Tamhidi, Aidin

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Earthquake Resilient Smart Cities:

A Framework for Collection and Utilization of Highly Granular Field Data for Seismic Performance Characterization of Soft-Story Buildings

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy in Civil Engineering

by

Aidin Tamhidi

2022

© Copyright by

Aidin Tamhidi

2022

ABSTRACT OF THE DISSERTATION

Earthquake Resilient Smart Cities: A Framework for Collection and Utilization of Highly Granular Field Data for Seismic Performance Characterization of Soft-Story Buildings

by

Aidin Tamhidi

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2022

Professor Yousef Bozorgnia, Chair

The traditional post-earthquake damage assessment of buildings is done visually through building-by-building inspections, which for a major event might take weeks to months to be completed. The main goal of this study is to establish a framework to quantify the post-event seismic performance of a portfolio of existing structures, in this case soft-story (SS) residential buildings, in near-real-time and an ultra-high resolution. Two main ingredients are required; 1) input ground motion time-series at the sites of the buildings that excite the structures, and 2) well-established nonlinear structural computer models representing the buildings. By having these two ingredients, nonlinear dynamic analyses of the existing structures can be performed, and seismic performance of the portfolio buildings can be quantified. The number of currently available recording instruments is sparse. Therefore, using the Gaussian Process Regression (GPR), a novel approach was developed to generate ground motion time series at un-instrumented target sites (i.e., the sites of the buildings) employing the surrounding recorded ground motions. The methodology has been validated for the physics-based scenario earthquakes in Northern California as well as two events recorded in Southern California: the 2019 **M**7.1 Ridgecrest and 2020 **M**4.5 South El

Monte earthquakes. In addition, an approach was introduced to generate multiple realizations of ground motions at each target site. The results illustrated that the instrumentation density closer to the target site is crucial in producing accurate and reliable ground motions, especially at long periods.

The second ingredient of the framework needs an accurate nonlinear computer structural model for soft-story buildings. There is an inventory of *OpenSees* models constructed to represent the 13,500 SS buildings in Los Angeles. This inventory includes thirty-two models categorized through four main key features: first-story wall layout, the number of stories, floor plan dimension, and wood-frame shear wall material. I have utilized machine learning methodologies to classify the first two features via visual recognition. I extracted images for 2,681 buildings within Los Angeles via Google Street View (GSV) service through an automated algorithm to train the Convolutional Neural Network (CNN) models for that purpose. The accuracy of the trained model is 80.1% and 91.4% for first-story wall layout and the number of stories classification tasks, respectively. I, then, utilized the OpenStreetMap (OSM) to detect the floor plan area of the target buildings to classify the third required feature. Having both a structural model and input ground motion time series, one can conduct a nonlinear response history analysis to assess the seismic performance of each specific building. As an application of the framework, I generated ground motion time series at over 2,000 SS buildings' locations within Los Angeles for a scenario **M**6.7 earthquake based on the 2020 **M**4.5 El Monte rupture mechanism. I also estimated the structural features through the GSV extracted images to identify the "closest" available *OpenSees* model. Lastly, a map of the estimated damage state for the 2,000 SS buildings is produced. The results demonstrated that more than 50% of the SS buildings are either severely damaged or collapsed after such an earthquake in Los Angeles, while approximately 15% of them experienced slight

damage. The methodology and framework developed in this research can be implemented to assess, in a near-real-time, seismic performance of a large portfolio of structures in moderate-to-severe earthquake events.

The dissertation of Aidin Tamhidi is approved.

Henry J Burton

Monica D Kohler

Nicolas Martin Kuehn

Yizhou Sun

Ertugrul Taciroglu

Yousef Bozorgnia, Committee Chair

University of California, Los Angeles

2022

To my lovely mother, Dr. Gisoo Ghaem, and to the soul of my father, Masoud Tamhidi,

to whom I owe everything …

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgement

The study presented in this manuscript was supported by the University of California, Los Angeles (UCLA) Graduate Fellowship, National Science Foundation (award no. 2025310), California Department of Transportation, and Pacific Gas and Electric Company which are all appreciated. Any opinions, findings, conclusions, or recommendations expressed are those of mine and do not necessarily reflect those of supporting agencies.

I sincerely appreciate the endless support of my Ph.D. advisor, Prof. Yousef Bozorgnia, whose encouragements, passion, and guidance were always my motivation during my education. I want also to thank Dr. Nicolas Kuehn, Prof. Mohamad Alipour, Prof. Monica Kohler, Prof. Tadahiro Kishida, Prof. Pengfei Wang, Prof. Ertugrul Taciroglu, Prof. Chukwuebuka Nweke, Dr. S. Farid Ghahari, Dr. Zhengxiang Yi, Dr. Richard Guy, Dr. Robert Graves, and Dr. Arthur Rodgers for their generous help and guidance during this study.

Last but not least, I want to express my heartfelt gratitude to my family, whom I have not seen in over five years due to my studies: my mother, father, and brother. My parents dedicated their lives and made personal sacrifices to help me obtain my Ph.D. I owe them everything in my life. I also appreciate my beloved partner, Parastoo, and my friends' support during my Ph.D. life's hard times.

VITA

2010-2014	B.Sc. in Civil Engineering, Sharif University of Technology, Tehran, Iran
2014-2016	M.Sc. in Structural Engineering, Sharif University of Technology, Tehran, Iran
2018-2022	Ph.D. Candidate in Structural and Earthquake Engineering, Civil and Environmental Engineering, University of California, Los Angeles

CHAPTER 1. Introduction

1.1 Motivation and Background

In the aftermath of an earthquake, the traditional post-earthquake damage assessment of buildings is carried out visually through building-by-building inspections. This procedure is generally done by a group of structural engineers inspecting the buildings within a region and classifying them through tagging suggested by Applied Technology Council (ATC) 20 guidelines (ATC-20, 1995). This process can be very labor-intensive, judgmental, and time-consuming. In general, these procedures might take months. For example, the inspection process took about two months to be completed for the 1994 Northridge earthquake (Trifunac and Todorovska, 1997). Thus, it is highly desirable to rapidly assess the performance or damage states on an ultra-high scale within a region following a moderate-to-major earthquakes (see, e.g., Ranf et al., 2007; Earle et al., 2010; Mangalathu and Jeon, 2020). Such information can be utilized by government emergency first responders, stakeholders, state officials, building owners, and insurance companies to allocate their resources rapidly and efficiently. There are studies aimed at developing methodologies to investigate the distribution of the damage to buildings after an earthquake (see e.g., Boatwright et al., 2015; Loos et al., 2020; Lu et al., 2021). Many previous studies aimed to develop such framework employed the ground motion intensity measure (GMIM) combined with the building fragility curves to assess the damage state or performance of the buildings. However, a thorough nonlinear response history analysis of buildings using site-specific structural models and ground motion time series has less been accomplished. Thus, such a methodology can reveal more detailed information regarding the building damage states after an earthquake.

1.2 Objectives and Contributions

Wood-frame buildings are among the most common types of dwellings in the United States cities, especially in the Western US. It is also very common that first stories of multi-unit wood-frame buildings are widely open for parking space, wide doors, or retails, which imposes an opening on the wall layout of the building. This opening within the first-story wall layout causes a considerable reduction in the lateral stiffness and strength of the first story compared to the immediate above story, resulting in a structural system known as a soft- and/or weak-story (SS) (FEMA P-807, 2012). Soft-story buildings are vulnerable to damage and possibly collapse during moderate to severe earthquake shakings (e.g., Holmes and Sommers, 1996). As an illustration, about two-thirds of 49,000 collapsed or damaged buildings in Los Angeles due to the 1994 **M**6.7 Northridge earthquake were soft-story structures (Public Policy Institute of California, 2006). A timely assessment of the distribution of damages to these vulnerable buildings requires two main components: 1) input ground motion time-series exciting at the location of the building, and 2) a reliable well-established computer structural model representing existing buildings. The main objective of this study is to develop and provide methodologies to construct the main required components for the rapid post-earthquake damage assessment for the SS buildings. The developed methodology here can also be expanded to other types of structures. More specifically, the major objectives and contributions of this study are summarized below.

1. The first aim is to develop a model to simulate the ground motion time-series at the sites where there is no recording instrumentation. 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 of site-specific structures, the input ground motion time series must be estimated for locations with no recording instruments. This goal is addressed by Gaussian Process Regression (GPR) to employ the surrounding observed ground motions to construct the time-series at any desired un-instrumented location. The model has been trained using physics-based dense simulated ground motions, and the performance of the model is evaluated upon different real earthquakes.

2. The constructed time series through the GPR model is uncertain. It is vital to quantify this uncertainty to know the reliability of the simulated motions with respect to different situations, such as instrumentation density, site condition, and corresponding structural period. The second goal of this study is to quantify the validity and uncertainty of the generated ground motions based on various governing factors. This goal is investigated by introducing a methodology to generate random realizations of the generated ground motions at each un-instrumented target site and quantify the variation of the simulations given the target site's features. The whole process of uncertainty quantification is assessed using two sets of recent earthquakes recorded in Southern California.
3. The next key component of the damage assessment is to have a reliable computer structural model for the target SS building. A total of thirty-two archetypical building models were developed by Burton et al. (2019) to represent 13,000 identified SS buildings within Los Angeles. As the next goal of the study, I aim to identify the main features required to develop the "closest" structural model for each target SS building. To address this issue, an automated methodology to extract buildings' images from Google Street View (GSV) service is developed. Then, these images are used to train Convolutional

Neural Networks (CNNs) to detect and classify the main key components of the SS buildings through image recognition. The GSV is a large-scale yet inexpensive source of urban imagery which can provide a rich source of visual information regarding the buildings to establish structural models. The methodology is developed and validated for a portion of the detected SS buildings within Los Angeles.

4. Eventually, the entire framework must be implemented and evaluated. I aim to perform the whole procedure, including time series generation and structural model classification, on a subset of Los Angeles SS buildings as the testbed. For that purpose, an earthquake recorded at the Community Seismic Network (CSN) and California Integrated Seismic Network (CISN) sites was used. Then, the ground motions and structural features are estimated using the results of previous items. The framework developed in this research can be implemented to generate a real-time map of structural performance at an ultra-high resolution. The process can also be implemented for other types of structures at any city located in high seismic zones.

1.3 Organization and Outline

The current study and thesis are written in six chapters. The contents and body of each chapter are summarized here.

Chapter 2 outlines the motivation of developing ground motion simulation methods. It demonstrates the main methods of ground motion generation, using physics-based and coherency function. The developed GPR method which can deploy the observed ground motions efficiently is presented. The formulation and theoretical background, assumptions, and datasets to train and examine the GPR model are introduced. Lastly, the generated motions through the GPR model are compared to those of recorded ones for three different earthquakes.

Chapter 3 introduces the objective to quantify the uncertainty and reliability of the produced time series. The time-series generated at each un-instrumented location has a variation depending on the governing factors such as instrumentations' density surrounding the target site. A methodology to generate realizations of the generated ground motions at the target sites is introduced, which enables the quantification of uncertainty at different structural periods. The effect of adding more observations from other seismic networks on the accuracy and variation of generated ground motions are investigated.

Chapter 4 starts with outlining the thirty-two *OpenSees* structural models developed by Burton et al. (2019) for Los Angeles. The key elements needed to develop the computer structural models are introduced. An automated algorithm to estimate these essential features for each target building is presented. Through this algorithm, the CNN models are trained for classifying buildings' features through GSV images. A portion of the SS buildings are employed as the training set while another subset makes the test set. Finally, the CNN models' performance in image and building-level classification tasks are evaluated.

Chapter 5 outlines the whole framework to estimate the damage state of each SS building in a real-time manner. The target domain including approximately 2,000 SS buildings is presented as the testbed. The application of the CNN models on these target buildings and their estimated features are shown. A scenario earthquake of **M**6.7 is constructed to excite Los Angeles city. To do so, a previously recorded ground motions for the 2020 **M**4.5 South El Monte earthquake recorded at the CSN and CISN stations are amplified. Eventually, the trained GPR model is deployed to generate ground motion time series at each target SS locations to excite the chosen closest *OpenSees* structural model. The statistics of the structural damage distribution over Los Angeles is presented.

Chapter 6 summarizes the conclusions and findings, presents the limitations and assumptions of the study, and introduces the potential future studies.

CHAPTER 2. Conditioned Simulation of Ground Motion Time Series

2.1 Introduction and Background

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, 2020). Thus, for site-specific post-earthquake performance assessment of specific structures, the input ground motion time series must be estimated for locations with no recording instruments. Currently, "ShakeCast" and "ShakeMap" developed by the 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, a nonlinear response history analysis requires ground motion time series as input. Therefore, a reliable generation of the ground motion time series is required (e.g., 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., 2021).

In general, there are two methods for generating ground motion time series: 1) physics-based simulations, which use finite-fault and seismic velocity models that can account for source,

path, and site effects (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 complexities of such approaches are a concern since they demand precise information on the site attributes and fault patterns. As a result, these procedures are less impractical for rapid post-earthquake damages assessment (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 (e.g., Abrahamson et al., 1991). Furthermore, detailed information regarding the site properties and wave propagation characteristics might be needed through these methods, which brings the similar issue of being demanding and time-consuming. Therefore, both methods have their critical challenges for a rapid post-earthquake structural damage assessment in a near-real-time manner.

In this chapter, a Gaussian Process Regression (GPR) method, also known as Kriging (Rasmussen and Williams 2006), is employed to generate the ground motion time series at target sites where there are no available recording instruments. This method is able to construct the ground motion time series decently at the target site using limited input information such as geographical coordinates and the average shear wave velocity in the uppermost 30 m, $V_{s_{30}}$, from each site. Therefore, it is able to estimate the motion time series with lower computational costs. The GPR method spatially interpolates the real and imaginary parts of the observed frequency content of the neighboring motions using an assumed covariance function to establish the ground motion time series at the target sites.

2.2 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 et al., 1997; Tamhidi et al., 2021; Tamhidi et al., 2022b)

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

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} 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} 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). Here, I assume that $\mathcal{R}e_k$ and $\mathcal{I}m_k$ (at k^{th} frequency, k = 0, …, N-1) are random Gaussian variables for any location, *s*, within a region. I also consider that $\mathcal{R}e_k$ at location *s*, is spatially correlated to $\mathcal{R}e'_k$ at location s' where *s* and s' are neighbors. A similar assumption for $\mathcal{I}m_k$ is taken. In this study, I aim to implement GPR as a method to estimate the values of $\mathcal{R}e_k$ (and $\mathcal{I}m_k$) at the k^{th} frequency (k = 0, …, N-1) using the corresponding $\mathcal{R}e'_k$ (and $\mathcal{I}m'_k$) from the surrounding station observations. It is assumed that there is a statistically insignificant correlation between $\mathcal{R}e_k$ (or similarly $\mathcal{I}m_k$) and $\mathcal{R}e_j$ (or similarly $\mathcal{I}m_j$) at the same location, *s*, for different frequencies *k* and *j*, where $k \neq j$, in order to construct the mean estimated ground motion time series. It is worth noting that the mean estimated

values for multivariate Gaussian variables (here $\mathcal{R}e$ and $\mathcal{I}m$) are independent of the inter-frequency correlation between amplitudes at various frequencies; yet the inter-frequency correlations of the DFT coefficients need to be accounted for generating random ground motion realizations.

2.3 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, 2022a; Sajedi and Liang, 2020; Sheibani et al., 2020). 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 vector observed values 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, I 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

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

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

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).

2.4 Model Development

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. I consider a four-element input vector of $\boldsymbol{x} = \{x_1, x_2, x_3, \log(V_{s_{30}})\}$ which is shown to be adequately informative about the features of the sites to estimate the ground motions (Tamhidi et al, 2021). $V_{s_{30}}$ stands for the time-average shear-wave velocity in the uppermost 30 m of soil, 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. I 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. The L2 and SCAD penalty functions are shown in Eqs. (2.10) and (2.11).

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

$$p_\lambda(\theta) = \begin{cases} \lambda\theta & \theta \le \lambda \\ -\dfrac{\lambda^2+\theta^2-2a\lambda\theta}{2(a-1)} & \lambda < \theta \le a\lambda \\ \dfrac{\lambda^2(a+1)}{2} & a\lambda < \theta \end{cases} \tag{2.11}$$

The GPR model is completely defined once the optimum parameters, $\hat{\gamma}$, having penalized log-likelihood function (cf. Eq. 2.9) maximized. The regularization factor, λ, needs to be defined before optimizing these parameters through maximizing $Q(\boldsymbol{\gamma})$. More specifically, λ governs the derivation of optimized parameters $\hat{\theta}$, $\hat{\mu}$, and $\widehat{\sigma_f}$. As a hierarchical view, one can recognize θ, μ, and σ_f as the parameters of the GPR model, while λ is its hyper-parameter. The process of optimization of this hyper-parameter is elaborated next.

2.4.1 Model Training using Physics-Based Simulation of the 1906 M7.9 San Francisco Earthquake

It is common to use data-driven methods such as Cross-Validation (CV) to find the optimum hyper-parameter values (Shao 1993), here the regularization factor, $\hat{\lambda}$. In our case, the "data" to be used in the "data-driven" methodology is a set of "observed" ground motions which is a subset of physics-based simulated ground motions for the 1906 **M**7.9 San Francisco earthquake. Here, I used broadband ground motions generated using Graves's hybrid simulation wave propagation code (Aagaard et al. 2008a). These ground motion time series were generated at 40,700 locations on a 1.5 km × 1.5 km uniform grid along three orthogonal directions. A minimum $V_{s_{30}}$ value of 760 m/s was used for these simulations. Correction factors were applied for site effects at locations with $V_{s_{30}}$ lower than 760 m/s. Two regions with 104 and 111 sites are selected to find out the optimum regularization factor. These two regions are hereafter referred to as the 'Palo Alto' and 'South Napa' regions (Figure 2.1). The sites within each of the Palo Alto and South Napa regions are randomly split into a training set (80% of the total number of sites) which makes up the "observed" ground motions, while the remaining 20% are considered the test set (target sites) (Figure 2.2).

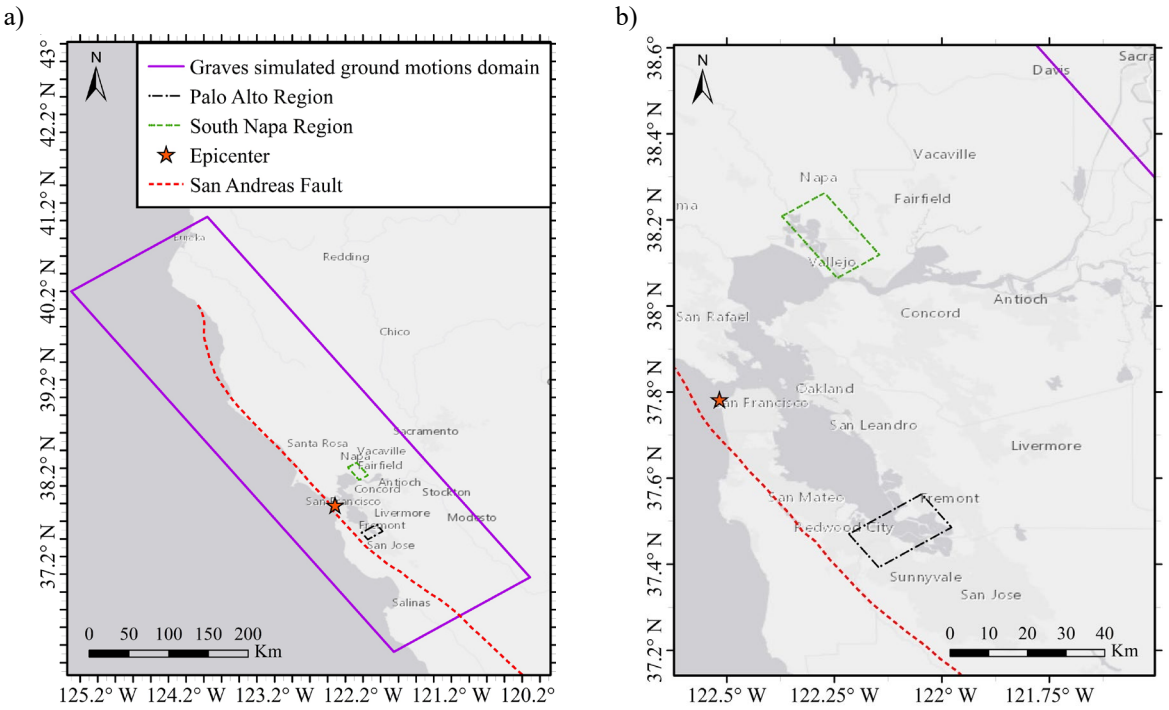


Figure 2.1a) Aagaard et al. (2008a) 1906 M7.9 San Francisco earthquake simulated motions domain and b) Palo Alto and South Napa study regions for training and testing of the GPR model

A five-fold CV procedure is implemented over the training set (observed ground motions) within each region in order to select the best regularization factor, $\hat{\lambda}$, for the corresponding GPR model. The accuracy criterion for this selection is the normalized root mean square error (NRMSE) between the exact (physics-based simulated) and the estimated (conditioned simulated) ground motions' 5%-damped pseudo-spectral acceleration (PSA) at the target site. The NRMSE is computed as

$$NRMSE = \sqrt{\frac{1}{\tau}\sum_{i=1}^{N_{period}} \frac{(PSA_i - \widehat{PSA}_i)^2}{\widehat{PSA}_i^{\,2}}}. \tag{2.12}$$

where, N_{period} is the number of structural periods included in the PSA, and PSA_i and $\widehat{PSA}_i$ are the predicted and exact ground motions' PSA values at the i^{th} structural period, respectively. A lower NRMSE value indicates a greater similarity between the estimated and recorded response spectra.

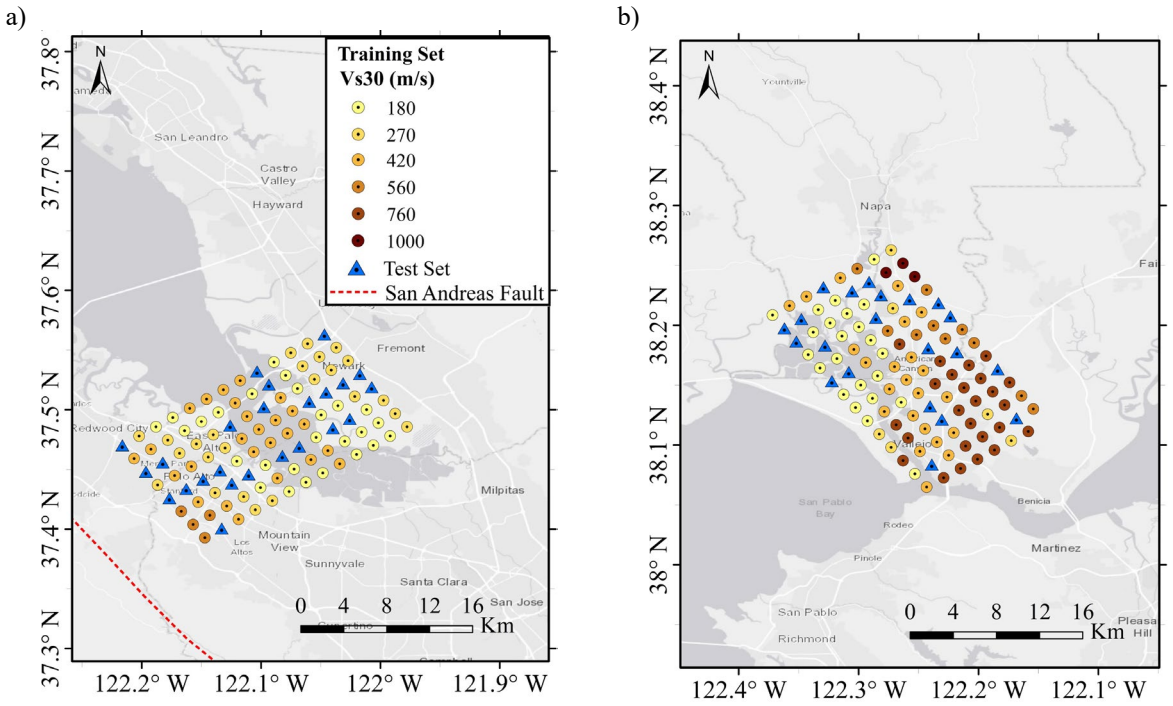


Figure 2.2 Distribution of the training and test sets for the a) Palo Alto and b) South Napa regions in 1906 **M**7.9 San Francisco Physics based sumulated earthquake

The following steps are taken to select $\hat{\lambda}$. First, I randomly split the training (observed) data set into five separate folds. For each λ_{test} to be evaluated, I carry out the following procedure:

1. For each fold i = 1, …, 5:

   1.1 Find the optimum parameters, $\hat{\theta}$, $\hat{\mu}$, and $\hat{\sigma_f}$ for the observed motions within all folds except the i^{th} fold using λ_{test} and maximizing $Q(\boldsymbol{\gamma})$ in Eq. (2.9). These parameters need to be found for each frequency and for both real and imaginary parts of the DFT coefficients.

   1.2 Estimate the ground motion time series at each site within the i^{th} fold using the posterior mean (Eq. 2.5) for the DFT coefficients, using $\hat{\theta}$, $\hat{\mu}$, and $\hat{\sigma_f}$ determined in step 1.1.

1.3 Compute the NRMSE between the estimated (step 1.2) and exact ground motion response spectra (Eq. 2.12) at each site within the i^{th} fold and store their averages as $Error_i$.

2. Take the average of $Error_i$ (i = 1, …, 5) i.e., $Error_{avg}$ and record it as being associated with λ_{test}.

Eventually, I choose the λ_{test} with the lowest $Error_{avg}$ computed in step 2 as the optimized regularization factor, $\hat{\lambda}$. The derived $\hat{\lambda}$ = 0.7 was obtained for both horizontal component directions, Fault Normal (FN) and Fault Parallel (FP). Next, I aim to illustrate the model's performance in predicting the ground motions using the physics-based simulated motions.

2.4.2 Performance Evaluation using the 1906 M7.9 San Francisco Earthquake

The test set's locations shown in Figures 2.1 and 2.2 are targeted to estimate their ground motion time series using the rest observed ground motions in that region and the $\hat{\lambda}$. 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, $\hat{\gamma}$= $(\hat{\theta}, \hat{\mu}, \hat{\sigma_f})$ are obtained at each frequency for the real and imaginary parts of the DFT coefficients using the corresponding $\hat{\lambda}$.
2. The posterior means (Eq. 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 2.3 displays the DFT coefficients' real part $\hat{\theta}$ values for the GPR model implemented within the Palo Alto and South Napa study regions along the FN and FP directions. Figure 2.3

demonstrates the $\hat{\theta}$ growth as a function of increasing frequency. A similar observation exists for the imaginary part $\hat{\theta}$ values. It is recognizable from Eqs (2.7) and (2.8) that covariance (and subsequently correlation) among the observed values decreases with increasing $\hat{\theta}$ (equivalently decrease of length-scale). In other words, there is a lower correlation between the higher-frequency content of the ground motions, which is consistent with the established lagged coherency models. Figure 2.4 demonstrates the distribution of NRMSE for RotD50 spectrum (Boore, 2010) between the estimated and exact ground motions at each test site for both Palo Alto and South Napa study regions.

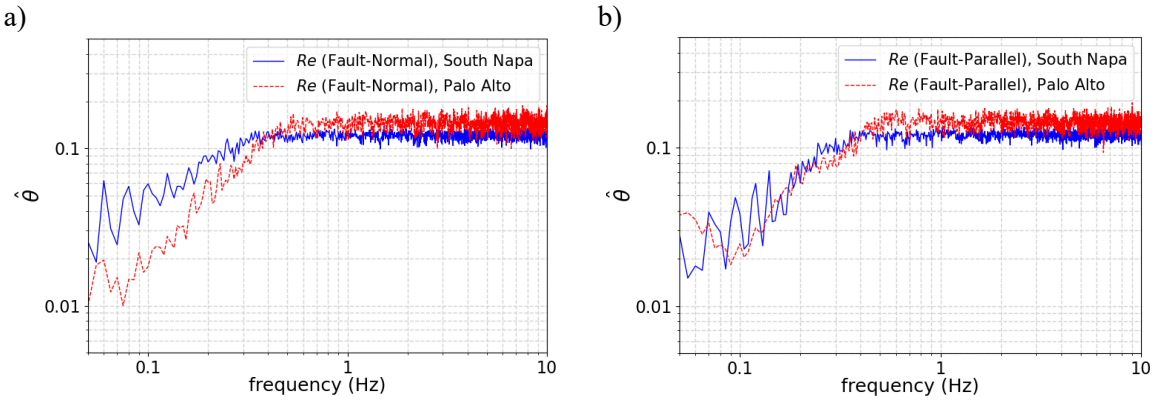


Figure 2.3. $\hat{\theta}$ for real part (Re) covariance functions along a) Fault-Normal and b) Fault-Parallel directions within Palo Alto and South Napa study regions

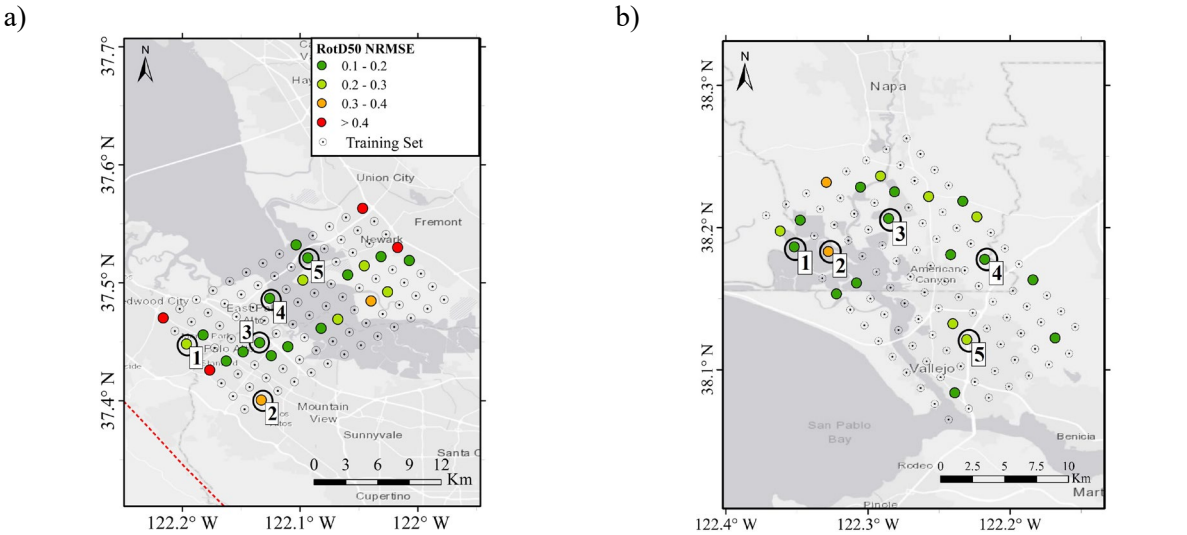


Figure 2.4 The distribution of the test set's NRMSE for the RotD50 spectrum at a) Palo Alto and b) South Napa study regions in 1906 **M**7.9 San Francisco Physics based sumulated earthquake

In Figure 2.4, there are five chosen test sites within each of the Palo Alto and South Napa study regions to illustrate their estimated motions. Figures 2.5 and 2.6 demonstrate the predicted motions' RotD50 response 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. In addition, Table 2.1 summarizes the average test set's NRMSE for response spectra along FN and FP directions and the RotD50 spectrum for both study regions.

Table 2.1 1906 M7.9 San Francisco test set's NRMSE for response spectra along FN and FP directions as well as the RotD50 spectrum

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 evident from Figures 2.5 and 2.6 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 (Figure 2.4a). It is also shown that the GPR model can predict the long period pulses due to the directivity effect reasonably well (Figure 2.5). By comparing the results of Table 2.1, 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).

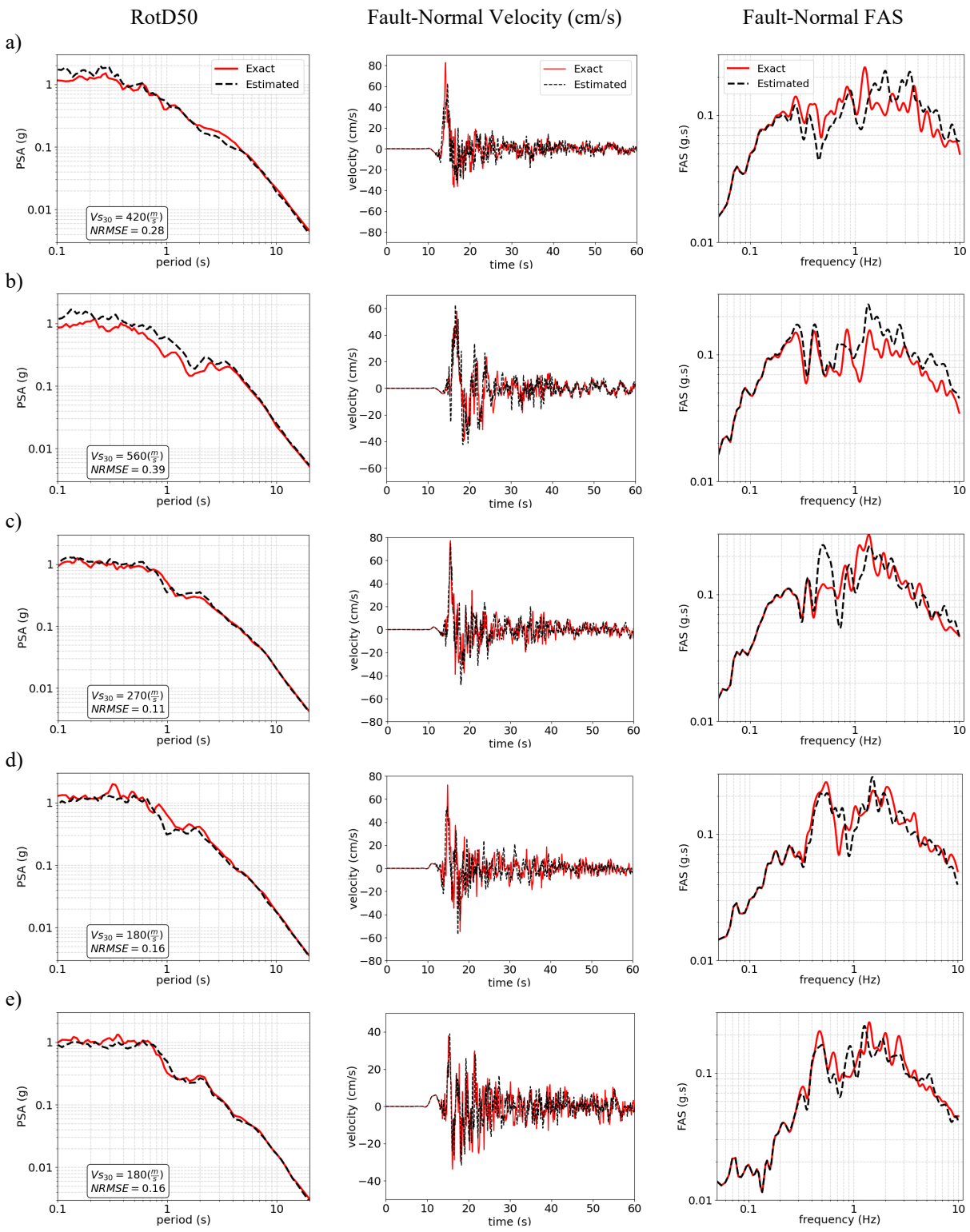


Figure 2.5 The RotD50, time series, and FAS of the predicted and the exact motions along FN direction for test sites: a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within Palo Alto region

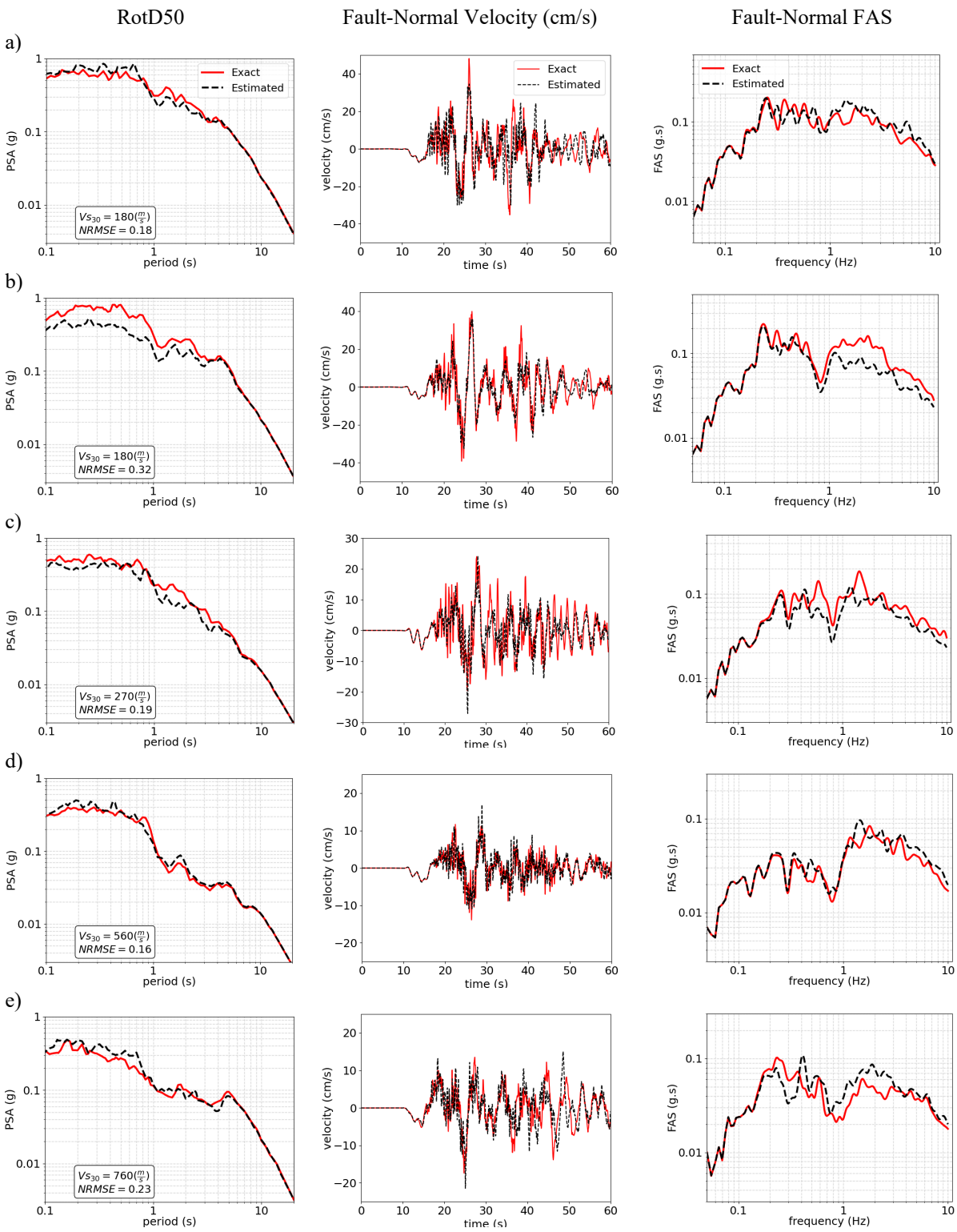


Figure 2.6 The RotD50, time series, and FAS of the predicted and the exact motions along FN direction for test sites: a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within South Napa region

2.4.3 Performance Evaluation using M7.0 Hayward Fault Earthquake Simulated Motions

Tamhidi et al. (2021) evaluated 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 2.7 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.

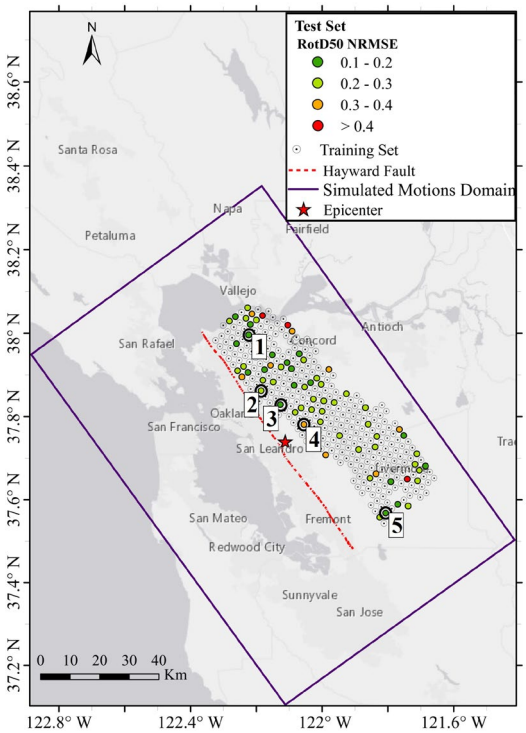


Figure 2.7 Distribution of the training and test set's NRMSE for the 5% damped RotD50 spectrum for the M7.0 Hayward fault scenario earthquake simulated motions

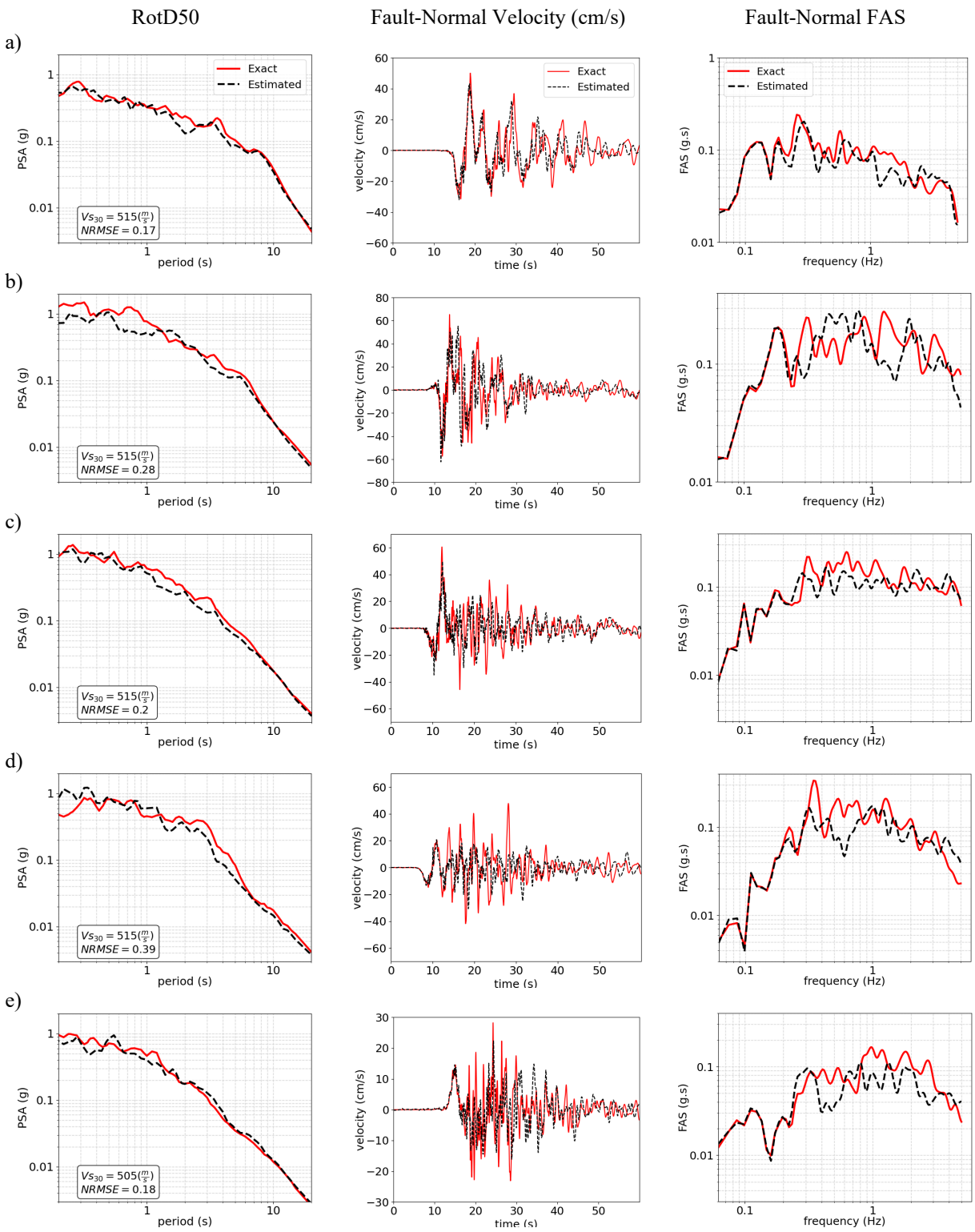


Figure 2.8. The RotD50, time series, and FAS of the predicted and exact motions along FN direction for test sites: a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within M7.0 Hayward fault earthquake

In Figure 2.7, there are five chosen test sites that I illustrated the predicted as well as exact motions' velocity time series and FAS along FN and the RotD50 spectra in Figure 2.8. As shown in Figure 2.8, the RotD50 spectra of the predicted motions are quite close to those of the physics-based simulated motions. Furthermore, Figures 2.8a through 2.8c show that the long period pulses of the ground motion (or Peak Ground Velocity, PGV) are reasonably accurately estimated. Moreover, Figure 2.7 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 observation density.

2.5 Optimum Regularization Factor Based on Observation Density

The CSN stations within Los Angeles (Clayton et al. 2020) are employed to investigate the effect of observation density on $\hat{\lambda}$. I used the recorded earthquake ground motions during the 2019 **M**7.1 Ridgecrest earthquake for that purpose. The 2019 **M**7.1 Ridgecrest earthquake happened on July 6th, 2019, in Searles Valley, 17.9 km Northeast of Ridgecrest, California. I implemented the 252 ground-level recording sites motions of the CSN in this study. Table 2.2 summarizes the 2019 **M**7.1 Ridgecrest earthquake features.

Table 2.2. 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 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.

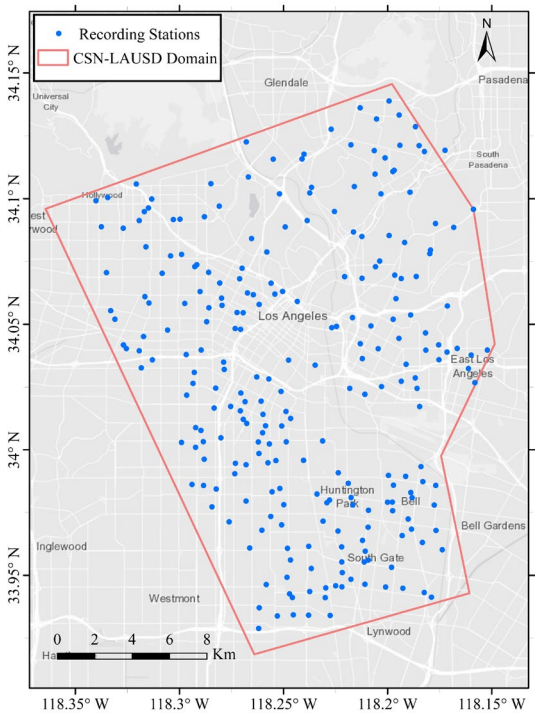


Figure 2.9. Distribution of the 252 CSN-LAUSD sites that recorded 2019 **M**7.1 Ridgecrest earthquake

The L2 penalty function was utilized for the penalized log-likelihood function (cf. Eq. 2.9). I created six datasets with various observation density by randomly selecting out of 252 CSN sites to evaluate the optimum $\hat{\lambda}$ for various observation densities. I made datasets with 252, 201, 151, 100, 50, and 25 recorded ground motions. The distribution of the randomly chosen sites for each

dataset is shown in Figure 2.10. 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).

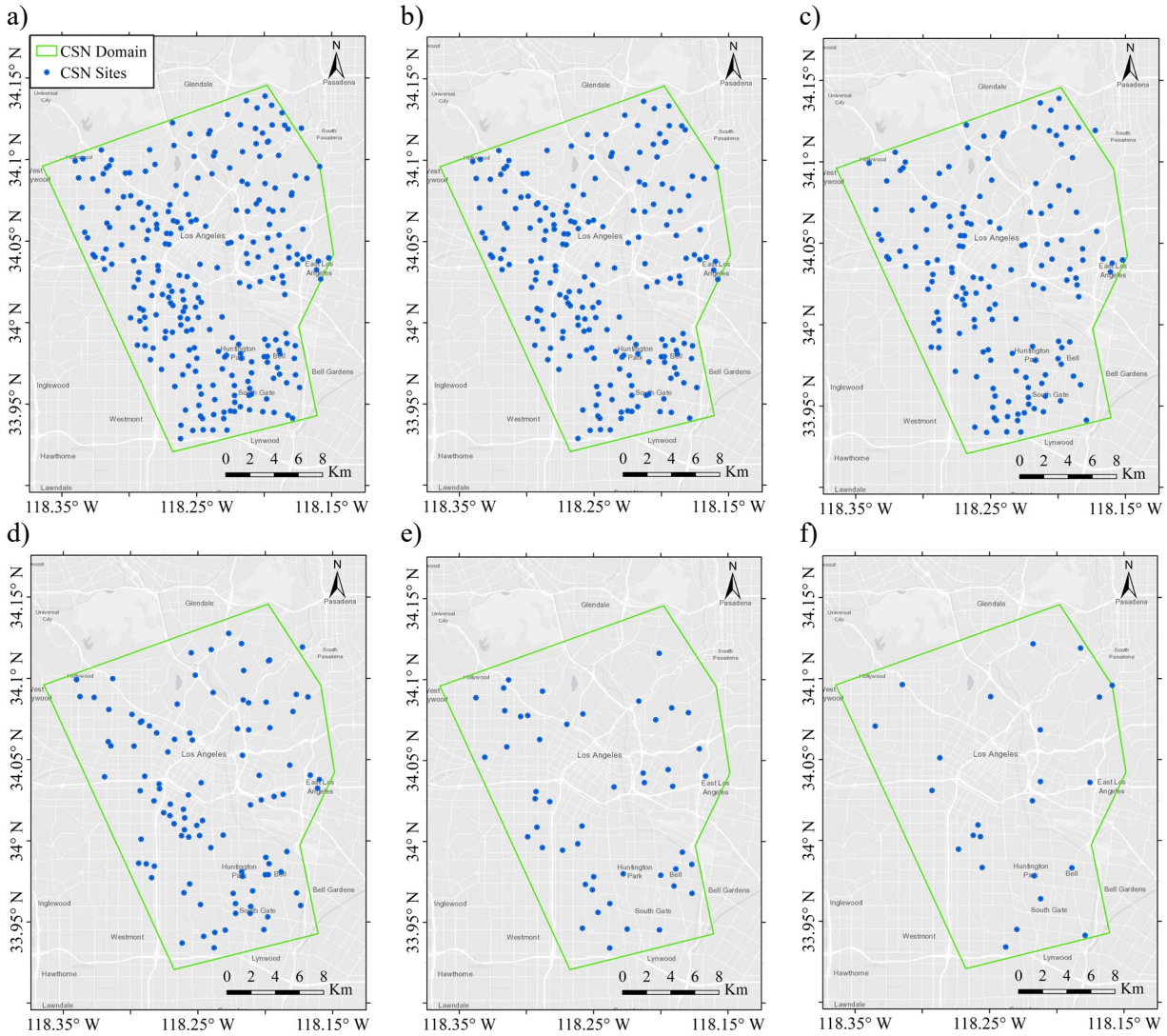


Figure 2.10. Distribution of the randomly chosen subsets from CSN's recorded 2019 **M**7.1 Ridgecrest earthquake motions with a) 252, b) 201, c) 151, d) 100, e) 50, and f) 25 number of sites

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 in Figure 2.10, the Leave One Out (LOO) CV method is used (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, …, 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, …, N_{sites}), i.e., $Error_{avg}$ and store it as the corresponding error to the λ_{test}.

Eventually, I chose the λ_{test} with the lowest $Error_{avg}$ as the $\hat{\lambda}$ for the corresponding dataset. Table 2.3 illustrates the obtained $\hat{\lambda}$ for each of the datasets in Figure 2.10.

Table 2.3. 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

I have employed the corresponding $\hat{\lambda}$ for each of the datasets shown in Figure 2.10 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). demonstrates the distribution of the NRMSE between the exact (recorded) and estimated motions' RotD50 spectrum.

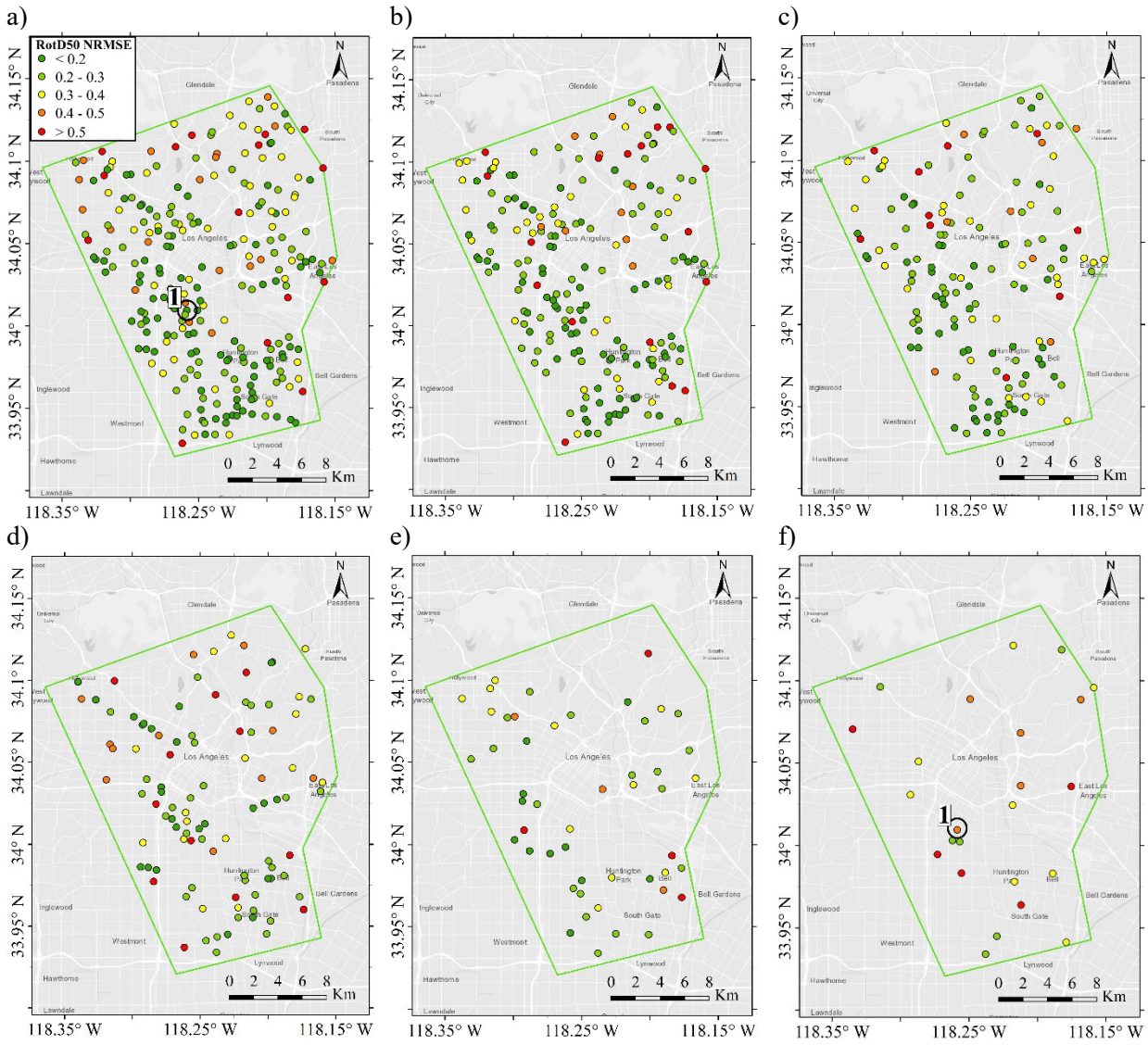


Figure 2.11. Distribution of the RotD50 NRMSE between the recorded and mean estimated ground motions using corresponding $\hat{\lambda}$ for CSN sites recorded M7.1 Ridgecrest earthquake having a) 251, b) 200, c) 150, d) 99, e) 49, and f) 24 number of observed sites. The test site 1 (panels a and f) is chosen as a target site for assessment of the introduced methodology for random generation of ground motions (see Chapter 3).

Figure 2.11 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, it shows that the number of stations with a higher RotD50 NRMSE (yellow, orange, and red stations) increases by decreasing the number of observed sites.

Table 2.4 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 2.11.

Table 2.4. Average and Standard Deviation of NRMSE between recorded and estimated motions' response spectra for each dataset shown in Figure 2.11

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

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.

CHAPTER 3. Uncertainty Quantification and Sensitivity Analysis of Simulated Ground Motions

3.1 Ground Motion Random Realizations

The ground motions generated in Chapter 2 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. It is desirable to quantify the uncertainty of the mean estimated ground motions at any target site (Tamhidi et al., 2022c). The posterior mean vector and covariance matrix for all target sites' k^{th} frequency DFT coefficients, k = 0, …, N-1, are given by Eqs. (2.5) and (2.6). In this study, I generate ground motion time series at one target site at each process. Therefore, Eqs. (2.5) and (2.6) can be converted to Eqs. (3.1) and (3.2), providing the scalar posterior mean, μ_*, and standard deviation, σ_*, for each $\mathcal{R}e_k$ (and $\mathcal{I}m_k$).

$$\mu_* = \hat{\mu} + \boldsymbol{k}_{\boldsymbol{x}_*\boldsymbol{x}} \boldsymbol{K}_{xx}^{-1} (\boldsymbol{f} - \boldsymbol{\mu}) \tag{3.1}$$

$$\sigma_* = \widehat{\sigma_f} - \boldsymbol{K}_{x_*x} \boldsymbol{K}_{xx}^{-1} \boldsymbol{K}_{xx_*} \tag{3.2}$$

In Eqs. (3.1) and (3.2), $\boldsymbol{k}_{\boldsymbol{x}_*\boldsymbol{x}}$ is the vector of covariance among the target site and observed sites' $\mathcal{R}e_k$ (or $\mathcal{I}m_k$). $\boldsymbol{k}_{\boldsymbol{x}_*\boldsymbol{x}}$ and $\boldsymbol{K}_{\boldsymbol{xx}}$ are established using $\widehat{\boldsymbol{\gamma}}$ at the corresponding frequency. The ground motion realizations at each target site are generated following the below steps.

1. At each k^{th} frequency, k = 0, …, N-1:
   1.1. The posterior mean and standard deviation of $\mathcal{R}e_k$ and $\mathcal{I}m_k$ are calculated by Eqs. (3.1) and (3.2).
   1.2. The correlation between the $\mathcal{R}e_k$ and $\mathcal{I}m_k$ at the target site is estimated by the correlation between $\mathcal{R}e_k{}'$ and $\mathcal{I}m_k{}'$ among all observed sites' (whole dataset except target site) ground motions. Consequently, a 2×2 covariance matrix for the ($\mathcal{R}e_k$, $\mathcal{I}m_k$)

is established using the estimated correlation and standard deviations resulting in step 1.1.

1.3. A set of random samples of 2×1 vectors of ($\mathcal{R}e_k$, $\mathcal{I}m_k$) are produced using the estimated 2×1 mean vector (step 1.1) and 2×2 covariance matrix (step 1.2). The sample size is selected so that the average of the generated samples becomes stable and converges to the mean vector determined in step 1.1. These generated ($\mathcal{R}e_k$, $\mathcal{I}m_k$) are then transformed to $|A_k|$ samples.

1.4. The logarithmic mean and standard deviation of $|A_k|$ samples in step 1.3 are obtained.

2. The N×N covariance matrix of $log(|A_k|)$, k = 0, …, N-1, is constructed using the inter-frequency correlation values given by Bayless and Abrahamson (2019a) model and calculated standard deviations in step 1.4.
3. Random Gaussian N×1 vector samples of $log(|A_k|)$ are produced using the N×1 mean vector (step 1.4) and N×N covariance matrix (step 2).
4. The phase spectrum of the mean estimated ground motion is coherent with the nearby observed ground motions. Therefore, the generated samples of FAS in step 3 are combined with the Phase spectrum constructed with the posterior mean DFT coefficients to generate ground motion time series realizations.

I also examined the randomization of Fourier phase spectra, yet the results were not as promising as the outcomes stated in step 4 above (Tamhidi et al., 2022b; 2022c). I employ the 2019 **M**7.1 Ridgecrest earthquake dataset recorded over the CSN sites to evaluate the proposed methodology. Test site 1 shown in Figures 2.11a and 2.11f, is the target site. The geotechnical properties of test site 1 are summarized in Table 3.1, where $Z_{1.0}$ and $Z_{2.5}$ are depths to the V_s = 1 km/s and V_s = 2.5

km/s horizons, respectively, and are estimated using the SCEC CVMS-4 model (Lee et al., 2014). In Table 3.1, R_{rup} is the closest distance to the coseismic rupture.

Table 3.1. Site properties of the test site 1 shown in Figure 2.11

Coordinates (Longitude, Latitude)	Vs_{30} (m/s)	$Z_{1.0}$ (km)	$Z_{2.5}$ (km)	R_{rup} (km)	Hypocentral Distance (km)
(118.258W, 34.009N)	290	0.62	4.48	191.7	204.1

Two different observed sets are considered to generate ground motion realizations at test site 1; first, all 251 CSN sites in Figure 2.11a, and second, all 24 CSN sites in Figure 2.11f. The $\hat{\lambda}$ for each case is chosen from Table 2.3. I generated one hundred ground motion realizations at test site 1. Figure 3.1 indicates the mean estimated and five ground motion time series realizations along the EW direction at the target site 1. Figure 3.1a displays that the mean estimated, and generated realizations of ground motion given 251 observed sites fit closer to the recorded one than those estimated using 24 observed sites in Figure 3.1b. In addition, it is observed from Figure 3.1a that the generated ground motion time series using 251 observed sites exhibit minor variation (uncertainty) at long periods (cf. velocity and displacement time series in Figure 3.1a). However, the higher frequency content of the generated motions shows a greater degree of uncertainty even using 251 observations (cf. accelerations in Figure 3.1a). On the other hand, Figure 3.1b displays that 24 observed sites are insufficiently informative to estimate the long-period content of the motions (cf. velocity and displacement time series in Figure 3.1b). The reason for this is that the average distance between the 24 observed sites is not close enough to predict long waves of motion. Therefore, the shortest period that the GPR model could possibly estimate reliably depends on the separation distance of the observations (instrumentations).

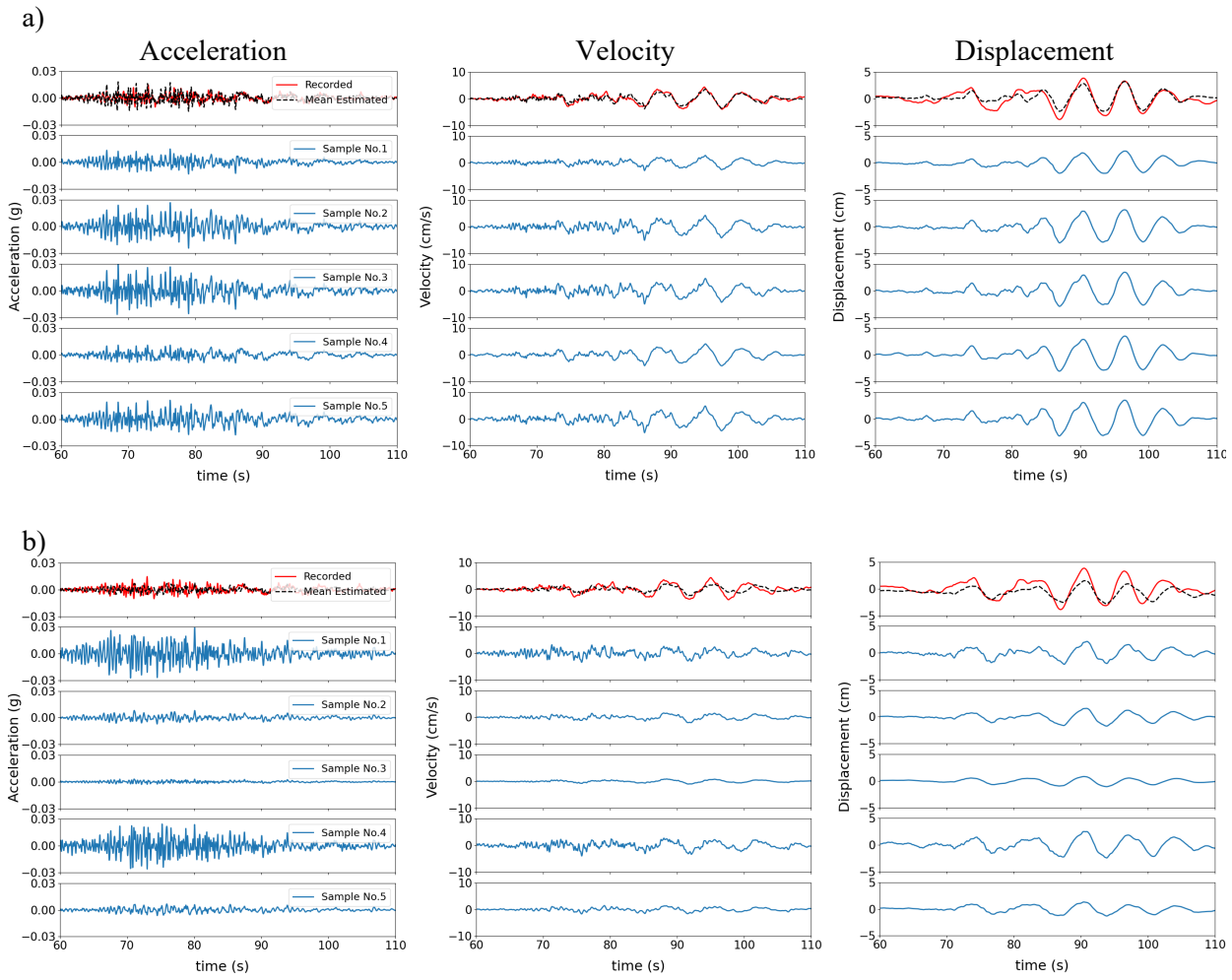


Figure 3.1. Estimated mean and five generated ground motion realizations time series along East-West direction at the test site 1 within the **M**7.1 Ridgecrest earthquake CSN dataset shown in Figure 2.11a and Figure 2.11f using a) 251 and b) 24 observed sites

Figure 3.2 depicts the mean estimated and one hundred ground motion realizations' 5%-damped RotD50 spectra at test site 1, utilizing 251 and 24 observed sites. It is acknowledged that the generated motions' uncertainty is lower at long periods than those at short periods. Moreover, Figure 3.2 indicates that the long-period prediction (longer than 1 second) has minor variation and error for having 251 observed sites than those estimated with 24 observations. On the other hand, neither 251 nor 24 observations are dense enough to provide informative detail of short-length waves corresponding to the short periods. That is why it is seen in Figures 3.2a and 3.2b that the

short periods' variation is high and does not differ significantly from having 251 observations to 24 ones.

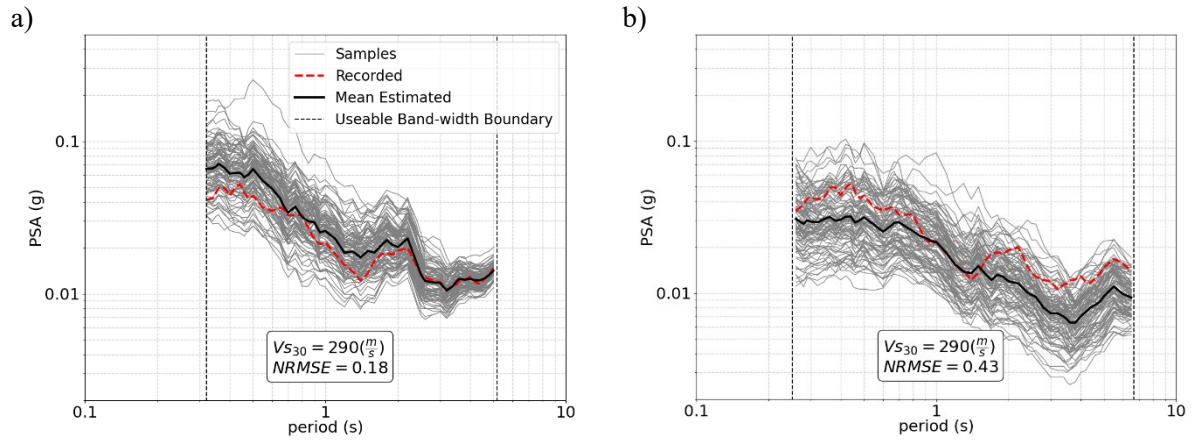


Figure 3.2. The 5%-damped RotD50 spectrum of generated ground motion realizations at the test site 1 using the a) 251 (Figure 2.11a) and b) 24 (Figure 2.11f) observed sites

Figure 3.3 presents the 68% confidence interval (CI), mean $\pm$ standard deviation, for the RotD50 spectra of generated realizations at test site 1 employing 251 and 24 observed sites. In addition, Figure 3.3 demonstrates the average RotD50 spectrum provided by CB14 (Campbell and Bozorgnia, 2014), ASK14 (Abrahamson et al., 2014), and BSSA14 (Boore et al., 2014) ground motion models (GMMs) and their average within-event standard deviation.

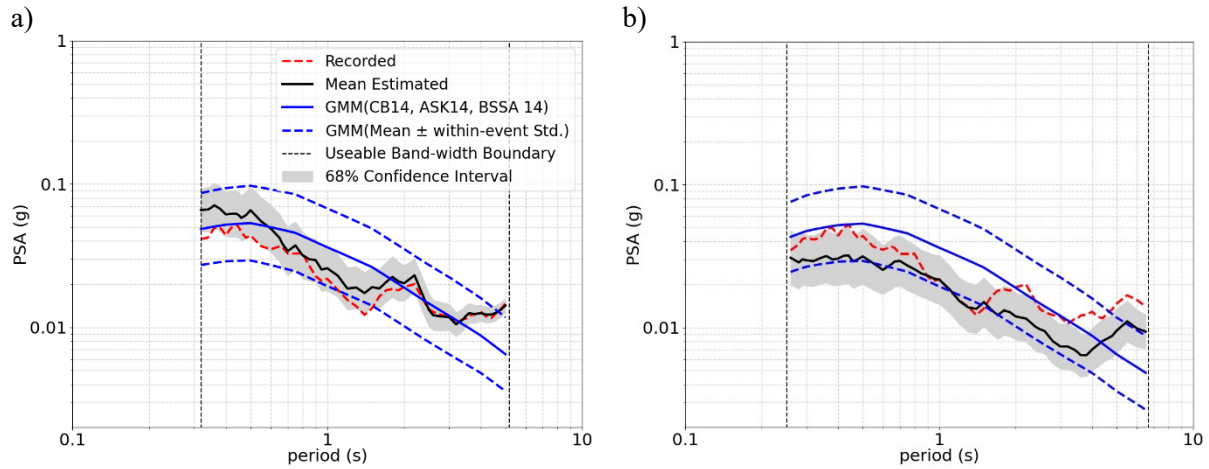


Figure 3.3. The 68% confidence interval of RotD50 spectrum of generated ground motion realizations at the test site 1 using the a) 251 (Figure 2.11a) and b) 24 (Figure 2.11f) observed sites

Figure 3.3a indicates that the recorded ground motion response spectrum falls inside the 68% CI of the generated motions using 251 observations for the majority of periods. Furthermore, Figure 3.3a shows that within-event uncertainty for the average GMMs is greater than that of generated motions utilizing 251 observations. Figure 3.3a also displays that the logarithmic CI of the estimated motions narrows at longer periods. In contrast, the within-event standard deviation of GMMs does not change considerably. In other words, the estimated ground motions' variability is less than that of GMMs, especially at long periods. On the other hand, Figure 3.3b demonstrates that the recorded ground motion response spectrum falls either outside or on the edge of 68% CI for having 24 observed sites. In addition, Figure 3.3b indicates that the standard deviation from short to long periods does not alter considerably for having fewer observations.

Furthermore, I 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, I have generated four maps that indicate whether the estimated motion's 68% CI encompasses the recorded motion's PSA at two periods, $T = 0.4$ (s) and $T = 2.0$ (s). Figure 3.4 illustrates the location of the recorded motion's PSA relative to the 68% CI of the estimated motion's PSA for each target station along the EW direction. In Figure 3.4, 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 3.5 shows the location of the recorded motion's response spectrum with respect to the 68% CI for the NS direction. In Figure 3.5, 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.

a) b)

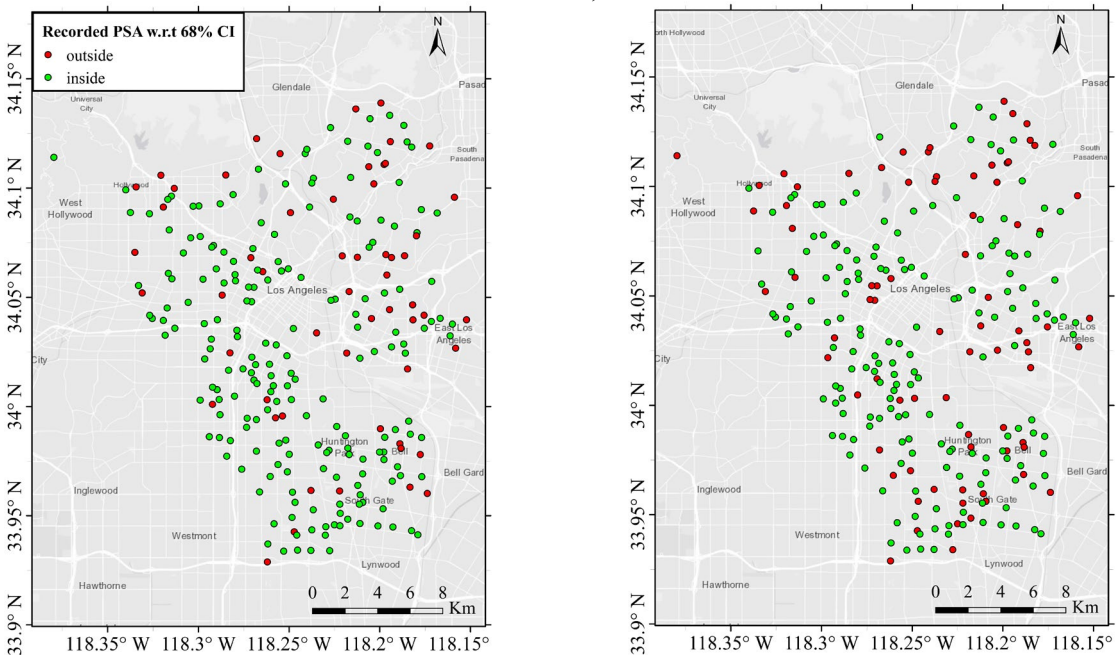


Figure 3.4. Location of the recorded motion's PSA with respect to the 68% Confidence Interval of predicted motion at a) $T = 0.4$ (s) and b) $T = 2.0$ (s) along East-West direction for 2019 **M**7.1 Ridgecrest earthquake dataset

a) b)

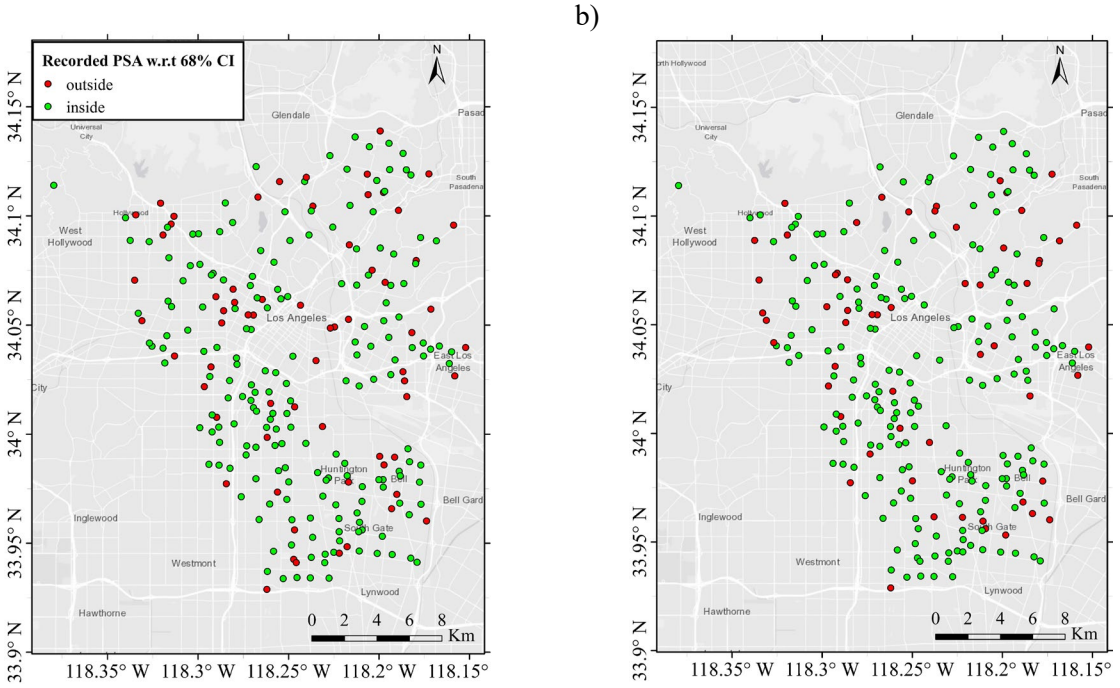


Figure 3.5. Location of the recorded motion's PSA with respect to the 68% Confidence Interval of predicted motion at a) $T = 0.4$ (s) and b) $T = 2.0$ (s) along North-South direction for 2019 **M**7.1 Ridgecrest earthquake dataset

3.2 Uncertainty Quantification and Sensitivity Analysis

The accuracy and uncertainty of the generated time series are quantified in this section. I employed 252 CSN sites' LOO analysis results for the 2019 **M**7.1 Ridgecrest earthquake. The logarithmic standard deviation of one hundred generated PSAs at two periods, $T = 0.4$ s, and $T = 2.0$ s, is obtained as a measure of generated motions' uncertainty at short and long periods, respectively. Figures 3.6 and 3.7 depicts the distribution of the EW and NS PSAs' logarithmic standard deviation horizontal component at $T = 0.4$ and 2.0 s, respectively.

Figures 3.6 and 3.7 illustrate that estimated ground motion realizations at CSN sites in the Los Angeles basin show minor variations at long structural periods ($T = 2.0$s) compared to those located outside the basin. On the other hand, the generated motions' uncertainty at the short period, $T = 0.4$ s, changes insignificantly between CSN sites inside and outside the Los Angeles basin. Comparing results at $T = 0.4$ s with $T = 2.0$ s in Figures 3.6 and 3.7 reveals that the PSA's logarithmic standard deviation at long periods is smaller than those of short periods for sites located on the basin. However, the PSA logarithmic standard deviation does not vary considerably from the short to the long period for sites located outside the basin. This is primarily because the observation density surrounding the target sites in the southern part of the CSN is high enough to produce reliable long-period motions. Furthermore, the sites atop the Los Angeles basin receive more coherent long period motions, as evidenced by (Kohler et al., 2020). Thus, the estimated motions at long-periods are less uncertain for the target sites on the basin. Figure 3.8 illustrates the distribution of the estimated $Z_{1.0}$ and $Z_{2.5}$ using the SCEC CVMS-4 model (Lee et al., 2014).

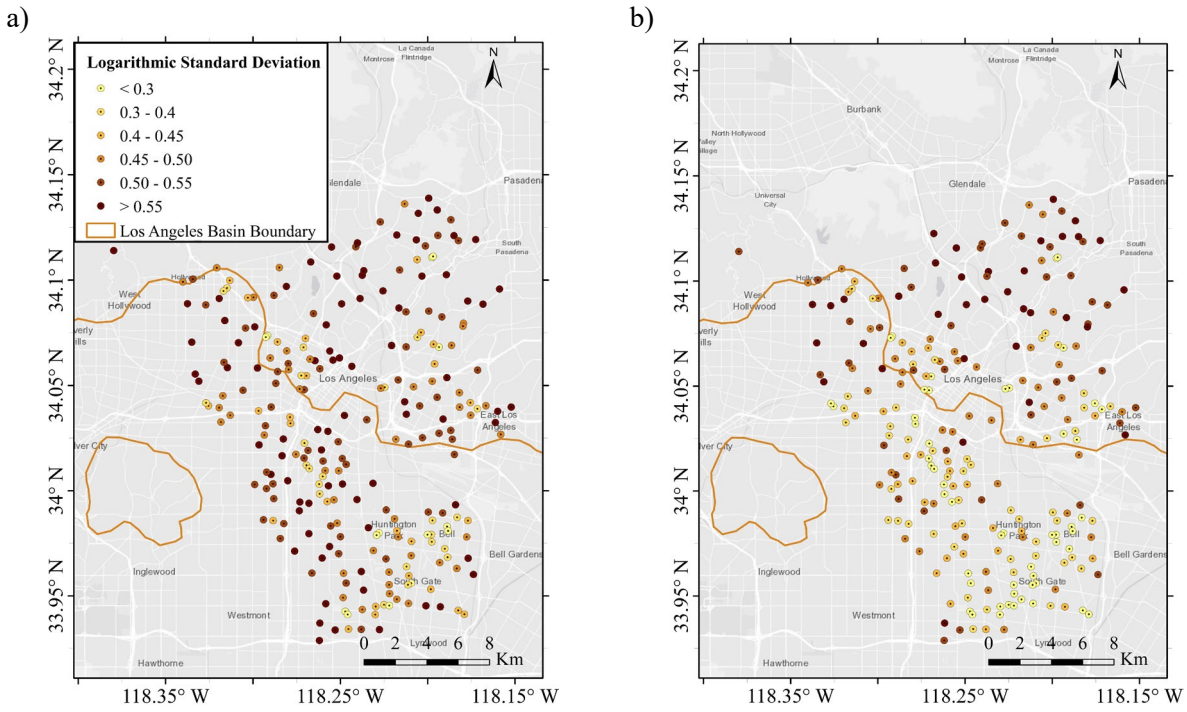


Figure 3.6.The logarithmic standard deviation of the estimated PSA along EW direction at a) 0.4 s and b) 2.0 s for 2019 **M**7.1 Ridgecrest earthquake dataset

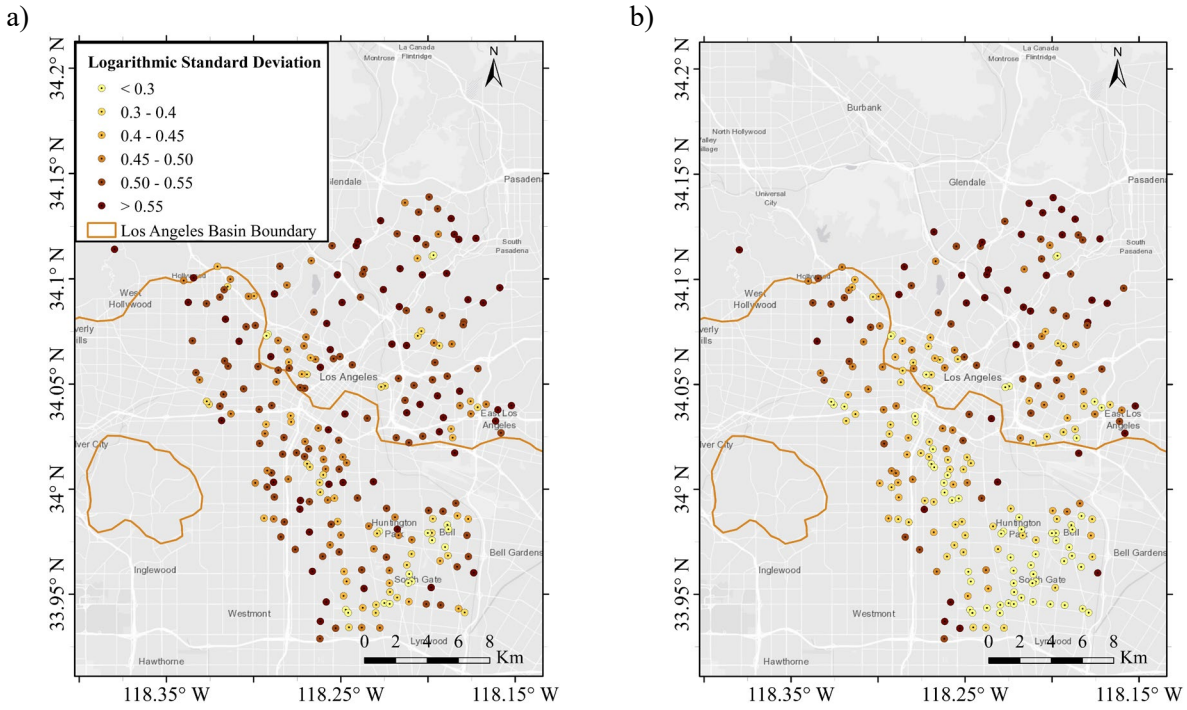


Figure 3.7. The logarithmic standard deviation of the estimated PSA along NS direction at a) 0.4 s and b) 2.0 s for 2019 **M**7.1 Ridgecrest earthquake dataset

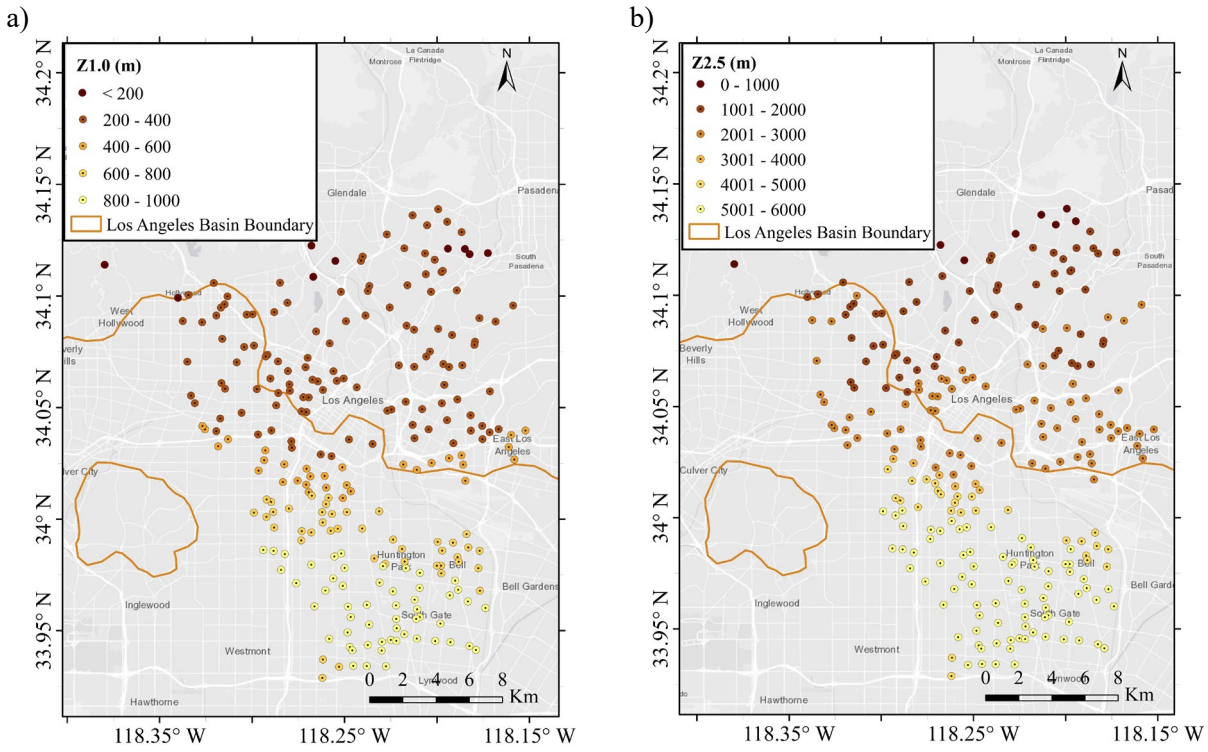


Figure 3.8. The distribution of the estimated a) $Z_{1.0}$ and b) $Z_{2.5}$ for the CSN sites

I evaluated the effect of observation density surrounding each target point on the response spectrum NRMSE. I selected the four nearest observed sites to each target station and computed their average separation distance (using the 4D space defined in Chapter 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 3.9 illustrates the scatter plot of the response spectrum NRMSE (within the usable bandwidth) along EW, NS, and the RotD50 versus the average separation distance. The separation distance (x-axis) is the distance between the feature vectors (with zero mean and unit standard deviation) of the sites within the 4D space of the GPR model, therefore is unitless.

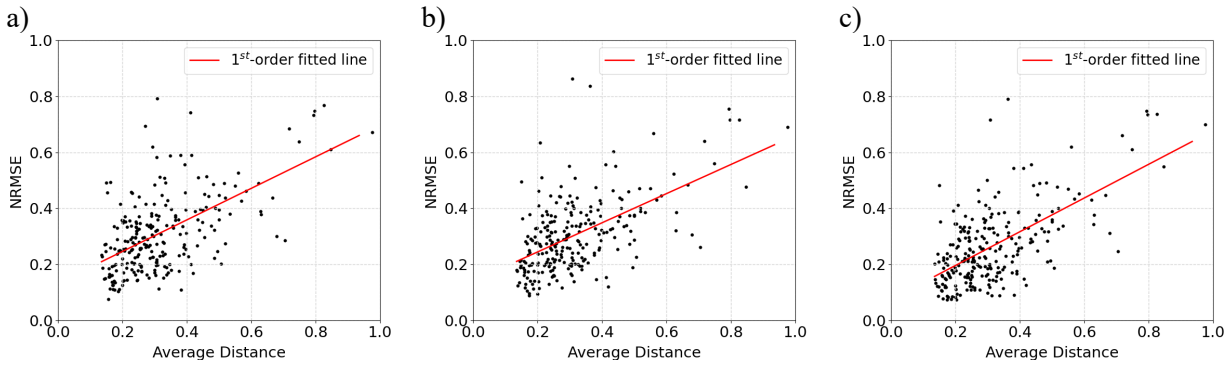


Figure 3.9. Scatter plot of the PSA normalized root mean square error along a) Eas-West and b) North-South, and c) RotD50 spectrum with respect to the average separation distance from four closest observations for 2019 **M**7.1 Ridgecrest earthquake dataset

Figure 3.9 shows that the estimation error increases having larger average distances. In other words, there is a general direct correlation between the prediction error and the density of the observations surrounding the target site. Figure 3.10 shows the scatter plot of PSA's relative error along EW and NS at two periods, T = 0.4 (s) and T = 2.0 (s). The relative error between the estimated and recorded motions' PSA at period T, $Error_T$, is calculated by

$$Error_T = \frac{PSA_T - \widehat{PSA}_T}{\widehat{PSA}_T} \tag{3.3}$$

where PSA_T and $\widehat{PSA}_T$ are the predicted and recorded motion's response spectrum at period T. It is observable from Figure 3.10 that the effect of the observation density on a longer period's error is more than that on a short period's error.

Furthermore, I 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. Figure 3.11 and Figure 3.12 indicate the scatter plot of the PSA's logarithmic standard deviation at T = 0.4 s and T = 2.0 s relative to the average separation distance along EW and NS, respectively.

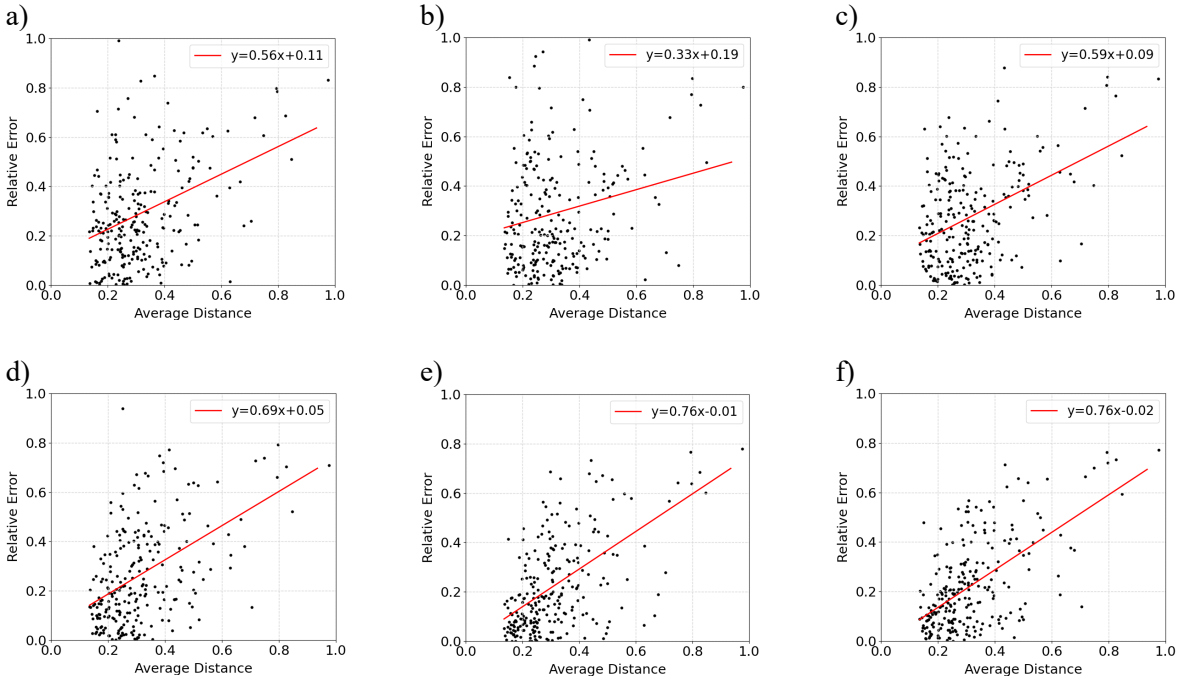


Figure 3.10. Scatter plot of the response spectrum relative error at $T = 0.4$ (s) along a) East-West, b) North-South, and c) RotD50 spectrum, and at $T = 2.0$ (s) along d) East-West, e) North-South, and f) RotD50 spectrum for 2019 **M**7.1 Ridgecrest earthquake dataset

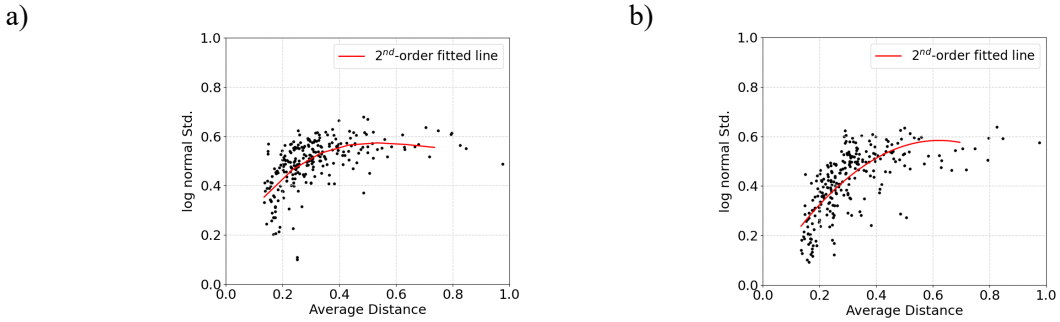


Figure 3.11. Scatter plot of the PSA log normal standard deviation at a) $T = 0.4$ (s) and b) $T = 2.0$ (s) along East-West direction for 2019 **M**7.1 Ridgecrest earthquake dataset

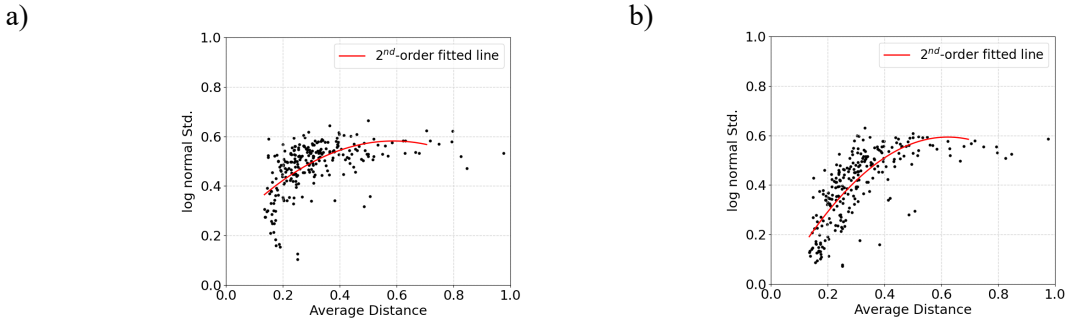


Figure 3.12. Scatter plot of the PSA log normal standard deviation at a) $T = 0.4$ (s) and b) $T = 2.0$ (s) along North-South direction for 2019 **M**7.1 Ridgecrest earthquake dataset

Figures 3.11b and 3.12b depict that estimation uncertainty increases having smaller observation density at long-periods. One can recognize that the uncertainty at the longer periods is generally more sensitive to the observation density compared to the shorter ones. This phenomenon is due to the complexities and intrinsic unpredictability of the short period motions, making added observations less useful to produce reliable short-period waves. Comparing the scatter plots at $T = 0.4$ s and $T = 2.0$ s in Figure 3.11 and Figure 3.12 reveals that the long-period motions have less variability than short-period ones at shorter average separation distances. It demonstrates that for a target site with an average distance of 0.2 from its four nearest neighbors, the estimated logarithmic standard deviation for PSA is around 0.45 and 0.30 at $T = 0.4$ s and $T = 2.0$ s, respectively. Furthermore, Figure 3.11 depicts that both short and long periods' uncertainty saturates for very large average separation distances. In other words, the GPR model produces random estimations with similar variance at short and long periods where there are too few observations.

Figure 3.13 demonstrates the stacked bar plots for the proportion of target sites where the recorded PSA falls inside (or outside) the estimated motions' 68% CI with respect to the average separation distance. The eight spans of average separation distance shown in Figure 3.13 are selected so that each span includes approximately the same number of target sites. Figure 3.13 indicates that the percentage of sites where the recorded PSA locates outside of the 68% CI rises as the average separation distance grows. This pattern becomes more apparent at $T = 2.0$ s. The percentage of sites where their 68% CI includes the recorded PSA decreases steadily for average distances greater than 0.3 and 0.2 for $T = 0.4$ s and $T = 2.0$ s, respectively.

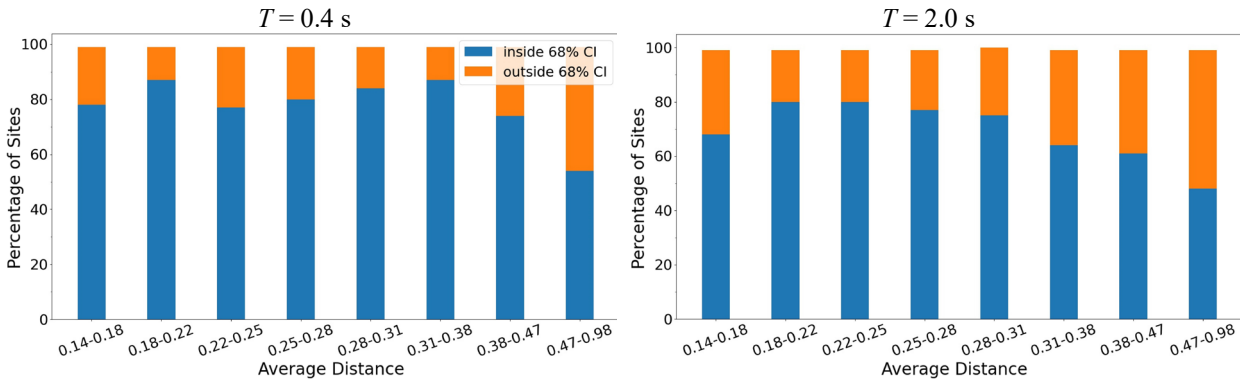


Figure 3.13. Stacked bar plots of the percentage of target sites where the EW recorded PSA falls inside the 68% CI with respect to average separation distance for 2019 M7.1 Ridgecrest earthquake CSN dataset

Figure 3.13 indicates that for the target sites with a greater observation density, the uncertainty of long periods' estimated motions is smaller than that for short periods; however, the probability that the recorded spectrum falls within 68% CI is almost the same at both short and long periods. In other words, for target sites with average distances of less than 0.3, the PSA realizations are about 80% likely to capture the recorded spectra within their mean ± standard deviation bandwidth.

I 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 (logarithmic standard deviation) of the estimated $V_{s_{30}}$ are obtained using the proxy-based model (Ahdi et al. 2020). Figure 3.14 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 a totally five logarithmic standard deviations) versus response spectrum NRMSE along EW, NS, and RotD50 spectrum.

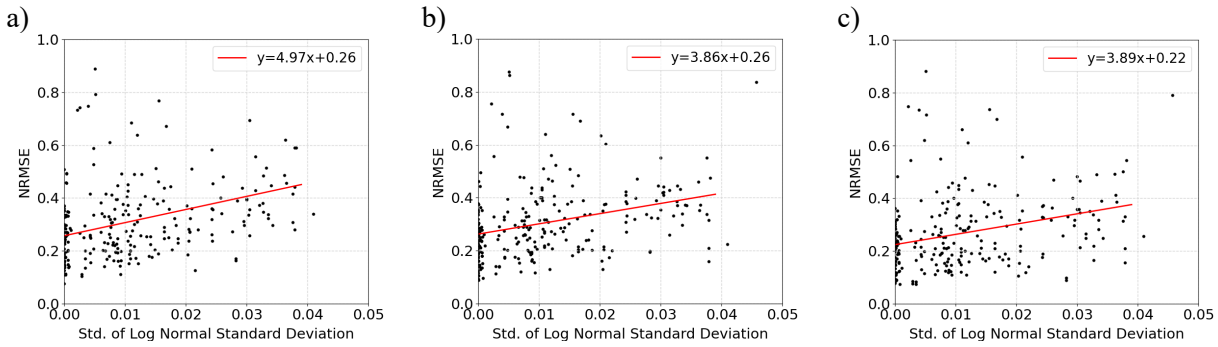


Figure 3.14. Scatter plot of the variation of logarithmic standard deviation of estimated $V_{s_{30}}$ for each target site and its four closest neighbors versus normalized root mean square error of response spectrum for a) EW direction, b) NS direction, and c) RotD50

It is observable from Figure 3.14 that the estimation error for both horizontal directions increase 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 with higher uncertainty, are more prone to have a higher prediction error. Figure 3.15 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 3.15 displays 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.

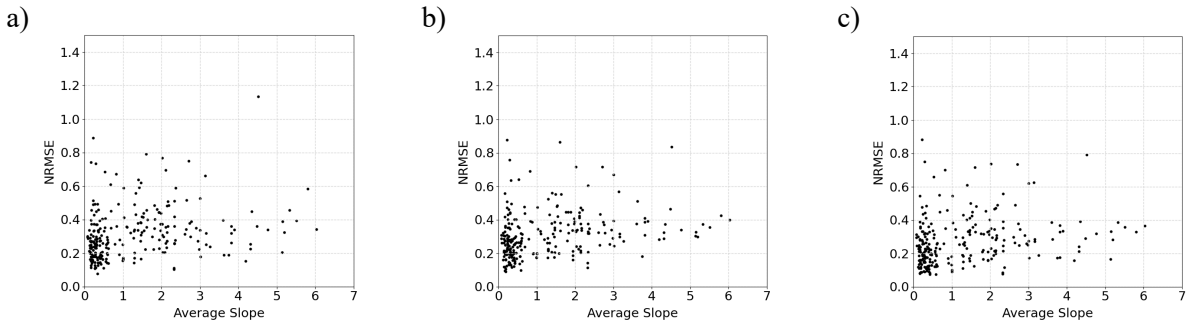


Figure 3.15. Scatter plot of the average estimated slope of each target site and its four closest neighbors versus response spectrum normalized root mean square error for the directions a) East-West, b) North-South, and c) RotD50

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

Herein I study the potential improvement of the ground motion prediction using combined observations from different seismic networks. There are various seismic networks in California, and the combined network is called California Integrated Seismic Network (CISN). First, I execute LOO ground motion prediction at each CISN station as a target site utilizing all other CISN sites (except the target site) as observation. Second, I perform the same procedure to estimate the ground motion time series at each CISN site using all other CISN and CSN sites as observation. Comparing the predicted motions resulting from these two observed sets with the recorded ones reveals the improvement of the GPR model's output. Ground motions recorded in two recent earthquakes are employed for this purpose; 1) 2019 **M**7.1 Ridgecrest and 2) 2020 **M**4.5 South El Monte earthquakes, as elaborated below.

3.3.1 Ground Motion Simulation for M7.1 Ridgecrest Earthquake

I selected 121 ground-level stations from the CISN network that recorded the 2019 **M**7.1 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 over a 460 km^2 region. The distribution of the CISN and CSN sites over Los Angeles is shown in Figure 3.16a. The $\hat{\lambda}$ depends on the observation density for each target site, as illustrated in Table 2.3. Thus, I separated the Main domain into three subdomains: 1) Inner, 2) Middle, and 3) Exterior Domains (Figure 3.16b). The observation density existing in each domain and, subsequently, the required $\hat{\lambda}$ are different. On the other hand, there is one observation density when the observation includes only the CISN stations. Table 3.2 shows the density of observation for each domain and

implemented $\hat{\lambda}$ for the target sites at the corresponding region. In Table 3.2, "Target Domain" refers to the region containing the target sites for which $\hat{\lambda}$ is suggested. It should be noted that observation density for the sites within the Inner, Middle, and Exterior target domains is derived by dividing the number of sites available inside the Inner, Inner plus Middle, and Main domains by their respective areas. Although the observation density for the Inner and Middle domains is approximately uniform, the Exterior domain's density varies from one region to another, similar to most existing seismic networks.

Table 3.2 demonstrates how the sites within the Exterior domain (Figure 3.16b) require $\hat{\lambda}$=0.2 for having more observations (CISN and CSN). In contrast, the same sites with fewer observations (Figure 3.16a) need a larger $\hat{\lambda}$=0.4. In addition, it is shown that the inner region of the CISN where the added CSN sites exist requires the smallest $\hat{\lambda}$=0.05.

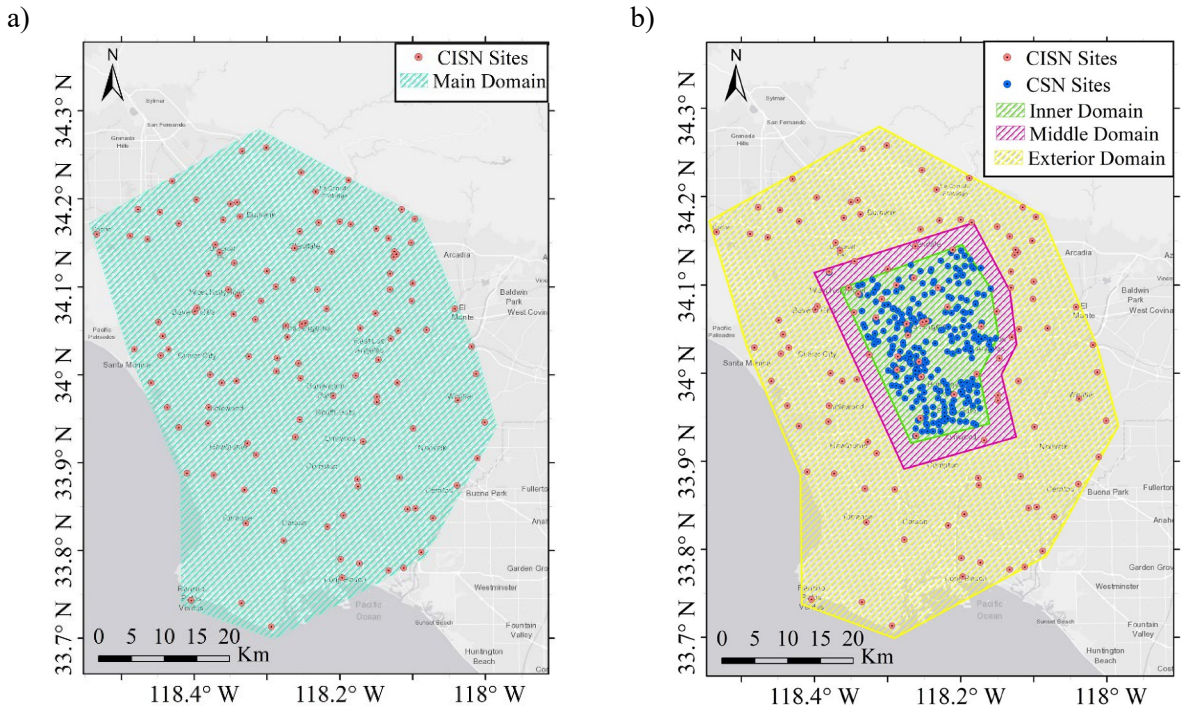


Figure 3.16. a) Distribution of the CSN and CISN sites recorded 2019 **M**7.1 Ridgecrest earthquake within Los Angeles, and b) Division of CISN network in Los Angeles into three different sub domains with various density of observations

Table 3.2. The implemented $\hat{\lambda}$ for 2019 **M**7.1 Ridgecrest earthquake

Observations	Target Domain	Area (km2)	Observation Density (site/km2)	$\hat{\lambda}$
CISN and CSN	Inner	464	0.57	0.05
CISN and CSN	Middle	764	0.36	0.10
CISN and CSN	Exterior	3103	0.12	0.20
CISN	Main	3103	0.04	0.40

I need to make the recorded ground motions at the CSN and CISN sites consistent with each other. First, all CISN and CSN motions are rotated to line up with the EW and NS directions. In addition, zero padding at the records' beginning and the end is implemented to ensure that all motions start and finish at the same Coordinated Universal Time (UTC). Lastly, the lowest sampling rate among all recorded motions is chosen as the target site's generated motion's sampling rate. Figure 3.17a and b show the distribution of the mean estimated motions' RotD50 NRMSE (Tamhidi et al., 2022c).

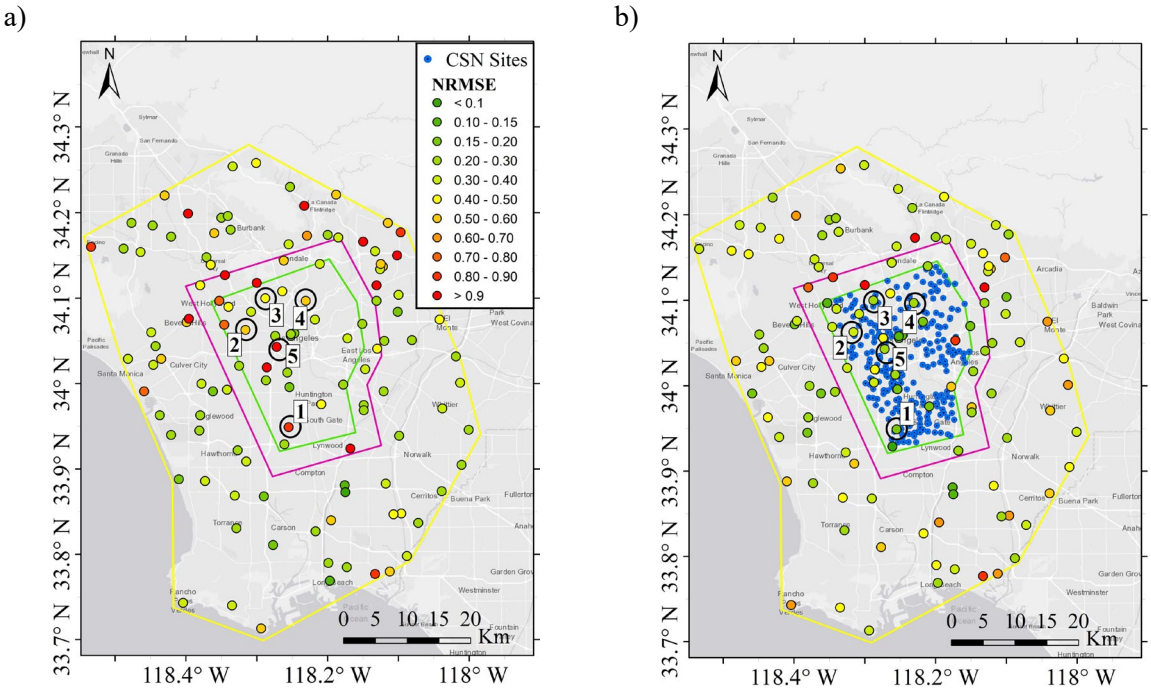


Figure 3.17. Distribution of RotD50 NRMSE for having a) CISN and b) all CISN and CSN sites as observations and the five chosen CISN test sites within the **M**7.1 Ridgecrest earthquake dataset

The average RotD50 NRMSE for all target sites utilizing just CISN as observation and both CISN and CSN as observation is 0.48 and 0.39, respectively. This means that the average RotD50 NRMSE is reduced by 19% due to the added CSN sites. In general, the NRMSE below a judgmental value of 0.3 indicates a reasonably precise estimation in terms of both time series and response spectrum. On the other hand, an NRMSE larger than 0.4 demonstrates a poor estimation. Interested readers are referred to Appendix A of this manuscript.

About 80% of the target sites inside the inner domain had mean estimated motions' RotD50 NRMSE lower than 0.34. However, there are two sites within the inner domain with RotD50 NRMSE values of 0.52 and 0.9 (orange and red points in Figure 3.17b). Figure 3.16b indicates that the majority of added CSN observations are positioned on almost one side of these two target points, resulting in a non-uniform observation distribution around them, which might lead to inaccurate ground motion estimations as evidenced by Tamhidi et al. (2021). Table 3.3 compares the average of the mean estimated ground motions' NRMSE along each horizontal component and RotD50 spectra for various domains.

Table 3.3. The prediction error along EW, NS, and RotD50 response spectra in different domains for the 2019 **M**7.1 Ridgecrest earthquake dataset

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

* Error Reduction shows the reduction in the average NRMSE among all CISN target sites due to the added CSN sites

Table 3.3 demonstrates that additional CSN sites generally improve the accuracy of the generated motions along both horizontal components for Inner domain target sites. Furthermore,

Table 3.3 reveals that the added CSN sites had the least impact on the predictions for the target sites in the Exterior domain. Therefore, the prediction for the target sites inside the added network's borders is improved as more observations become available. However, this effect is less substantial for the target sites outside the added network's domain.

I also investigate the CSN sites added observations' effect on the prediction errors at both short and long-period content of the motions. To do so, I plotted the distribution of the relative error at two periods, T = 0.4 (s) and T = 2.0 (s) for RotD50 spectrum in Figures 3.18 and 3.19, respectively. Tables 3.4 and 3.5 summarizes 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 3.4 and 3.5, 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 exterior domain.

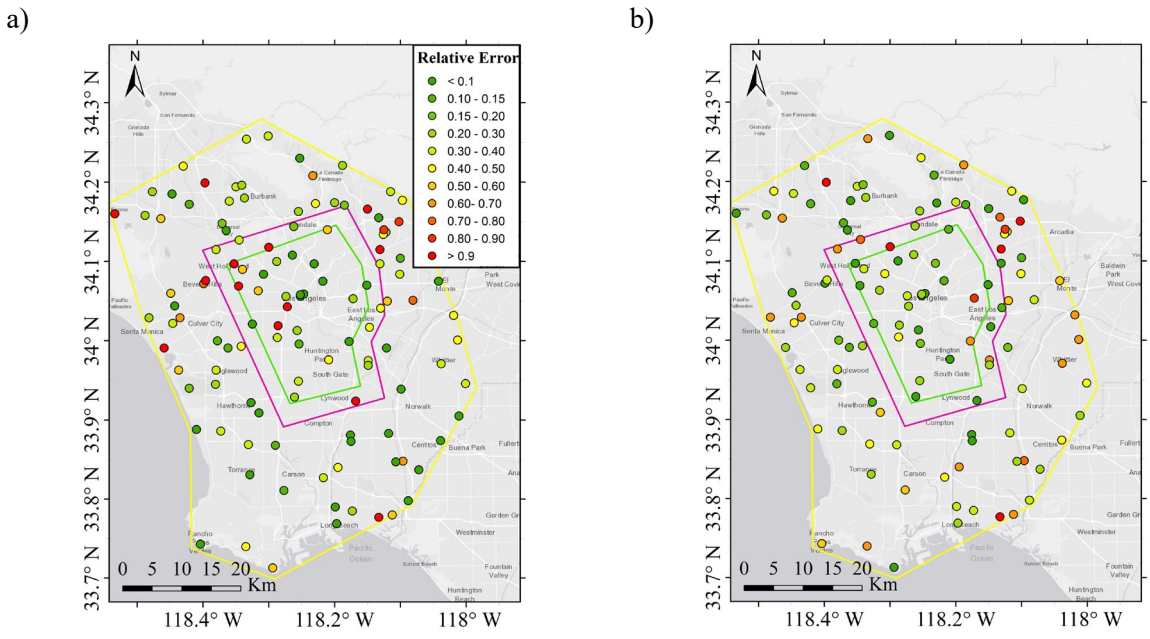
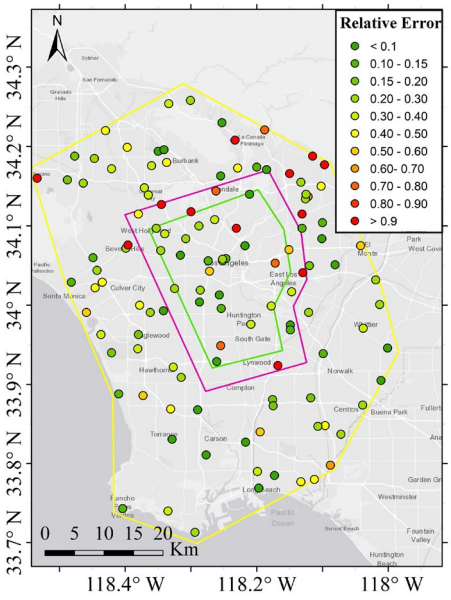


Figure 3.18 Distribution of RotD50 relative error at T = 0.4 (s) for having a) CISN and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

a)



b)

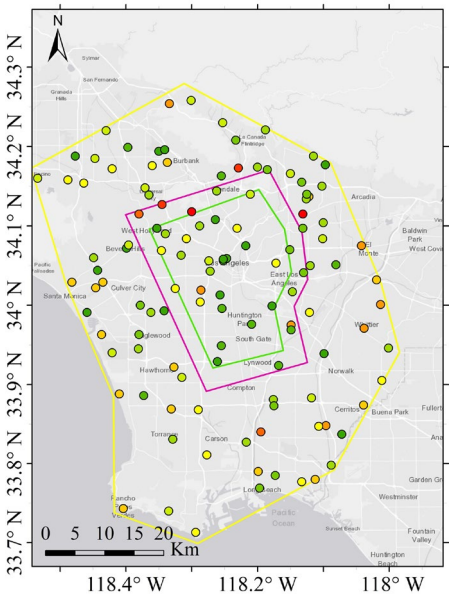


Figure 3.19 Distribution of RotD50 relative error at $T = 2.0$ (s) for having a) CISN and b) all CISN and CSN sites as observations for the **M**7.1 Ridgecrest earthquake

Table 3.4 . 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 3.5 . 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	

I chose five sites within the inner domain to show the improvement of the predicted results after adding CSN sites as observation. Figure 3.17a and b display 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 3.20 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 3.20 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. Readers can refer to Appendix A to observe more results for other test sites within Figure 3.17.

I also demonstrate the logarithmic standard deviation of one hundred generated ground motions' PSA at two periods, $T = 0.4$ (s) and $T = 2.0$ (s), before and after added CSN observations in Table 3.6 and Table 3.7, respectively. As is shown in Table 3.6, 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). On the other hand, the uncertainty of the longer period content of the motions decreases inside the added observation CSN network (inner domain) (cf. Table 3.7). 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 many informative inputs for the exterior domain's predictions.

Table 3.6. 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 3.7. 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	

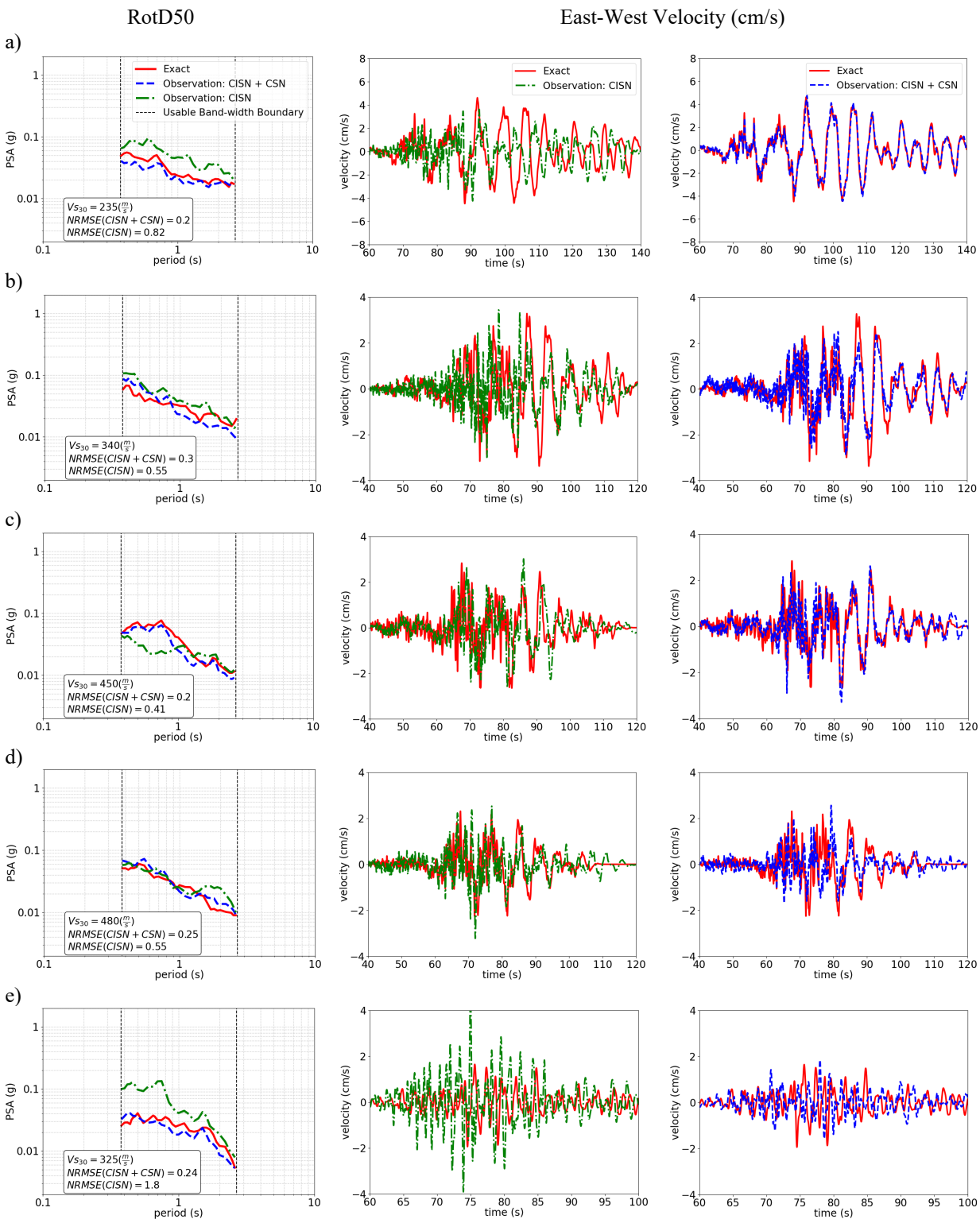


Figure 3.20. The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation as well as the exact motions along East-West direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within the CISN for **M**7.1 Ridgecrest earthquake

3.3.2 Ground Motion Simulation for M4.5 South El Monte Earthquake

In addition, I evaluate the influence of added observations for the recently recorded ground motions of the 2020 **M**4.5 South El Monte earthquake. Table 3.8 outlines the **M**4.5 South El Monte earthquake characteristics (USGS, 2020). I utilized 95 and 215 ground-level recording sites for CISN and CSN in Los Angeles, respectively (see Figure 3.21a and b). The number of sites is obtained by eliminating those with a too narrow usable bandwidth.

Table 3.8. 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

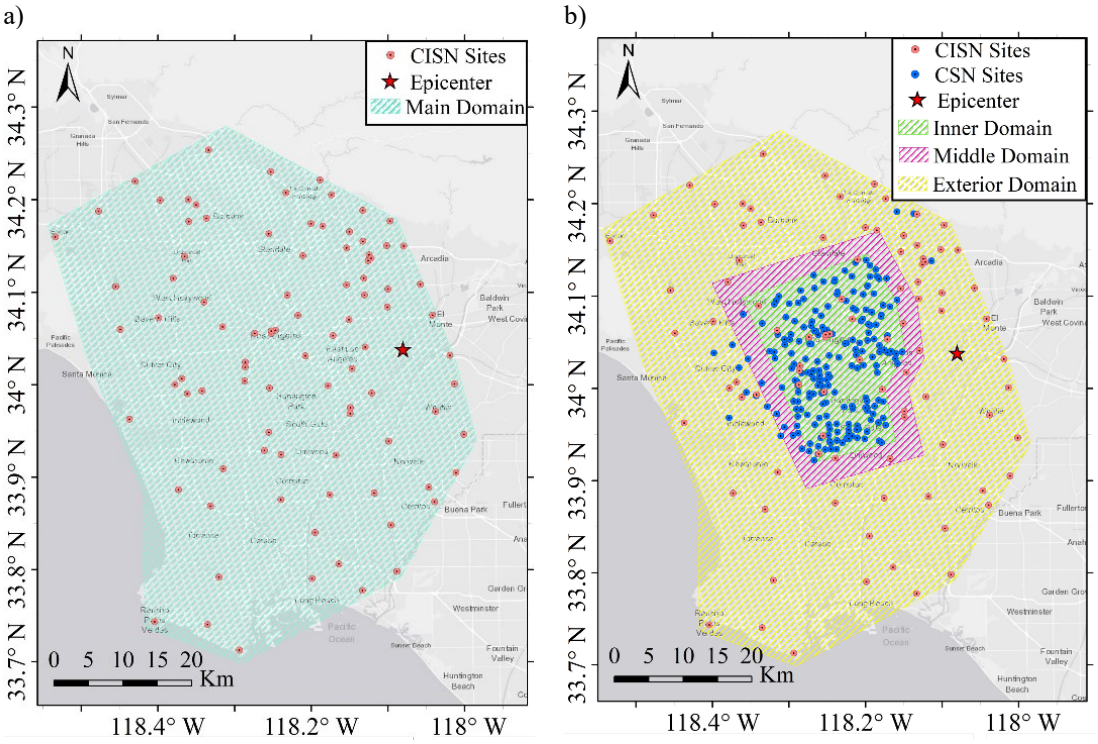


Figure 3.21. Distribution of a) CISN sites, b) CISN and CSN sites, c) RotD50 spectrum NRMSE for having CISN sites as observation, and d) RotD50 spectrum NRMSE for having both CISN and CSN sites as observation for 2020 South El Monte earthquake

Table 3.9 summarizes the observation density and the corresponding employed $\hat{\lambda}$. The $\hat{\lambda}$ for Inner domain having both CISN and CSN sites and $\hat{\lambda}$ for the Main domain having just CISN

sites as observations are obtained using logarithmic interpolation and extrapolation over the $\hat{\lambda}$ values presented in Table 2.3, respectively.

Table 3.9. The implemented $\hat{\lambda}$ for 2020 **M**4.5 South El Monte earthquake

Observations	Target Domain	Area (km2)	Observation Density (site/km2)	$\hat{\lambda}$
CISN and CSN	Inner	464	0.46	0.08
CISN and CSN	Middle	764	0.30	0.10
CISN and CSN	Exterior	3103	0.10	0.20
CISN	Main	3103	0.03	0.50

Figure 3.22a and b demonstrate the distribution of the mean estimated motions' RotD50 NRMSE at each CISN site. The average RotD50 NRMSE among all target sites for CISN-only and CISN-plus-CSN observed sites is 0.80 and 0.75, respectively. Approximately 67% (twelve sites) of the target sites inside the Inner domain had an NRMSE smaller than 0.32 (Figure 3.22b). There are three target sites inside the Inner domain with an NRMSE larger than 0.5, indicating that their estimates worsened after adding more ground motions from CSN sites.

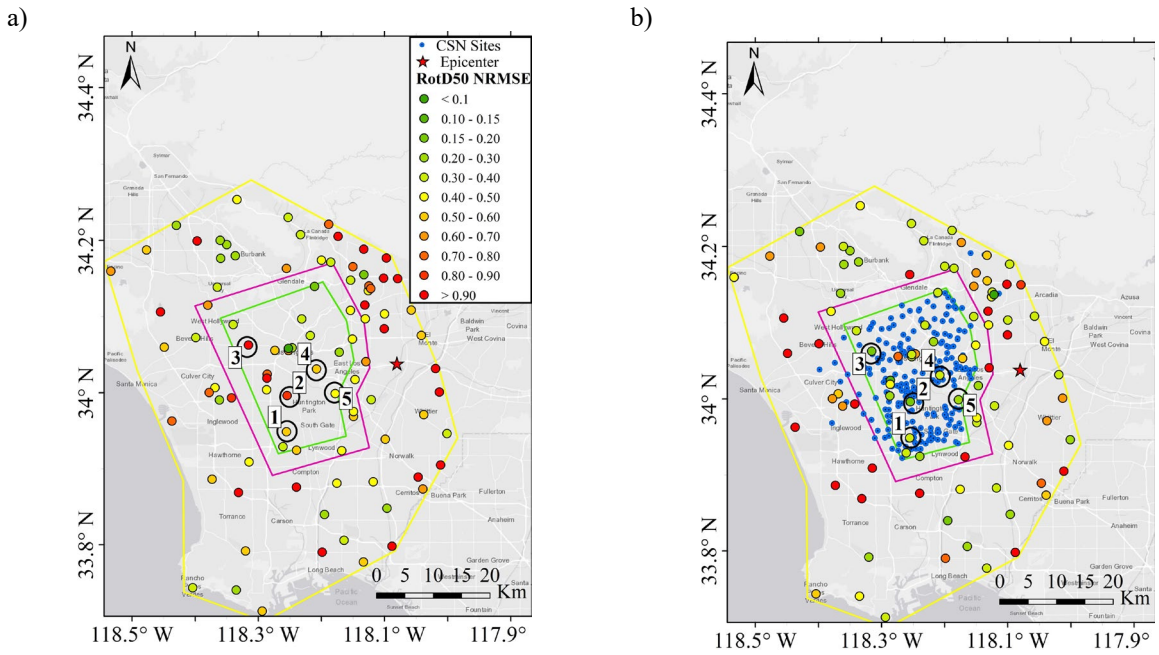


Figure 3.22. Distribution of RotD50 NRMSE for having a) CISN and b) all CISN and CSN sites as observations and the five chosen CISN test sites within the **M**4.5 South El Monte earthquake dataset

Table 3.10 compares the NRMSE of the mean estimated ground motions for each target domain. Table 3.10 and Figure 3.22 indicate that the addition of observed sites from CSN, generally improved the prediction of the ground motions inside the Inner domain (30% reduction in RotD50 NRMSE); yet, there are a few sites within the Inner domain where the estimation deteriorated after observing more sites from CSN (orange sites in Figure 3.22b). Comparing Table 3.10 and Table 3.3 reveals that the influence of added CSN sites for the **M**4.5 South El Monte earthquake is less than that for the **M**7.1 Ridgecrest earthquake. There are two reasons for the latter. First, the mutual usable bandwidth of the estimated motions for the South El Monte earthquake (0.11 s – 0.55 s) is narrower and shorter than that for the Ridgecrest earthquake (0.38 s – 2.8 s), and it is discussed that the effect of the additional observations on the precision of the generated motions is higher at long-periods. Second, the isotropic covariance functions deployed in the GPR model may provide somewhat inaccurate estimates in the epicentral area (Tamhidi et al., 2021). Thus, the added CSN observations might have a negligible effect on improving the estimations for the 2020 **M**4.5 South El Monte earthquake dataset.

Table 3.10. The prediction error along EW, NS, and RotD50 response spectra in different domains for the 2020 **M**4.5 South El Monte earthquake dataset

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

The influence of added observations is negligible for the sites within the Middle or Exterior domains and, in some cases, can worsen the estimations. It should be noted that the number of

available CISN sites within the Middle domain is sparse (9 sites), which can affect the statistical inference of the added CSN observations' effect in that region.

I chose five stations within the inner domain of the CISN (Figure 3.22) 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 ones. Although there are 95 sites within all inner, middle, and outer regions, I chose five of them within the inner domain where the added CSN sites exist. Figure 3.23 demonstrates the predicted motions' PSA and velocity time series along the NS direction for both scenarios. It is observable from Figure 3.23 that how the amplitude of the velocity-time series becomes closer to the recorded one for having more observations from CSN. The readers can refer to Appendix A to see the prediction results for other CISN test sites.

Table 3.11 also shows the logarithmic standard deviation of one hundred generated realizations' 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 is possible to investigate its uncertainty is T=0.5 (s). Table 3.11 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. 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.

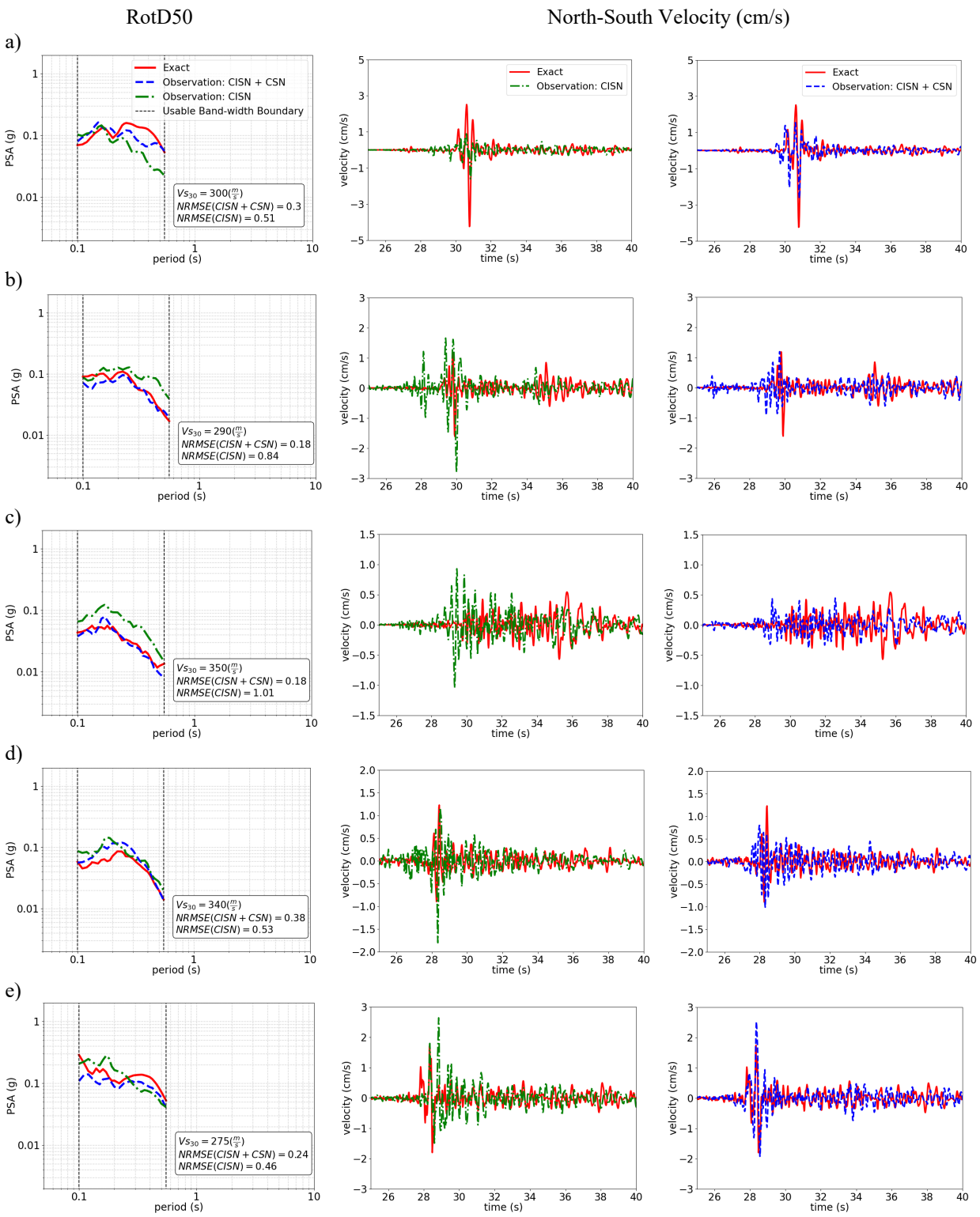


Figure 3.23. The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation as well as the exact motions along North-South direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, and e) No. 5 within the CISN for **M**4.5 South El Monte earthquake

Table 3.11. 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

CHAPTER 4. Rapid Structural Model Classification for Soft-Story Wood-Frame Buildings

4.1 Introduction and Overview

Wood-frame residential buildings are among the most common types of buildings in the United States cities. There are wood-frame buildings that the first story is widely open for parking space, wide doors, or retails that impose an opening on the wall layout of the buildings. This opening within the first-story wall layout causes a considerable reduction in the lateral stiffness and strength of the first story compared to the immediate above story, resulting in a structural system known as a soft- or weak -story (SS) (e.g., FEMA P-807, 2012). A building is officially classified as a SS building if one of the stories has 70% or less stiffness than the immediate above floor (ASCE 31-03, 2003). In addition, the weak story is referred to one that has strength equal or less than 80% of that of above ones. The soft-story buildings are vulnerable to damage and collapse during moderate to severe earthquake shakings (e.g., Holmes and Sommers, 1996). As an illustration, about two-thirds of 49,000 collapsed or damaged buildings in Los Angeles due to the 1994 **M**6.7 Northridge earthquake were soft-story structures (Public Policy Institute of California, 2006). Moreover, collapses of such buildings have been recorded extensively during the 1971 San Fernando and 1989 Loma Prieta (e.g., Harris and Egan, 1992) earthquakes.

Reliable computer structural models approximating the existing buildings are required to complete the damage assessment methodology. Such building models can then be excited using the generated ground motion time series to estimate the structural responses for existing SS buildings. There are about 13,500 soft, weak, and open front (SWOF) buildings identified in Los Angeles metropolitan area in 2016 (Los Angeles Times, 2016). The Los Angeles City office signed the law mandating the seismic retrofit of SWOF wood-frame buildings in January 2016 (Ordinance

No. 184081). In order to characterize the SWOF structural models, a total of thirty-two archetypical nonlinear building models were developed by Burton et al. (2019) after surveying 3,000 buildings out of identified SWOF structures in Los Angeles. The nonlinearity of the computer models embedded in the utilized Pinching4 model to construct wood shear walls (Burton et al., 2019). These thirty-two un-retrofitted models are developed in the *OpenSees* platform (Mazzoni et al., 2006). The development of these archetype structural models requires four key parameters: 1) the first-story wall layout, 2) the number of stories, 3) floor plan dimensions, and 4) the material of exterior/interior panels. One can rapidly assign the archetype structural model to an existing SS building, given the four critical features of the structure.

This chapter presents an automated method to identify the main key components required for selecting the structural model for a given SS building. The first two components are obtained using image recognition through the trained Convolutional Neural Network (CNN) models. I provide the trained CNN models to classify the first-story wall layout and number of stories of the target buildings using visual recognition of street view imagery. To create a scalable system for this task, this study leverages the Google Street View (GSV) service as a large-scale yet inexpensive source of urban imagery. The GSV images of the target buildings are automatically harvested from the Application Programming Interface (API) given the buildings' addresses. These images are then fed into the CNN models to train them for detecting the first story wall layout and number of stories out of street view images. The Floor plan dimensions (third component) are assigned after harvesting the building footprints from *OpenStreetMap* (OSM) platform. The last fourth key element, the material of the interior/exterior walls of the buildings, remains unknown, yet, as a future study, one can obtain such information out of the tax registry databases, which might provide the detailed features of the residential buildings.

4.2 Developed Soft-Story Structural Models for Los Angeles

Burton et al. (2019) considered four first-story wall layouts for the SWOF buildings in Los Angeles. These four wall layouts are called L1, L2, L3, and L4, which are shown in Figure 4.1. The L1 has one completely open wall line in the shorter direction and two partially open wall lines in the longer direction. In contrast, the L2 has a completely open wall line in the longer direction and two partially open ones in the shorter direction. L3 has one single open wall line surrounded by three walls, and L4 has two partially open wall lines. The collapse performance of the SWOF buildings is highly dependent on the first-story wall layout. Burton et al. (2019) identified 17%, 2%, 61%, and 20% of their surveyed SWOF buildings in Los Angeles as L1, L2, L3, and L4, respectively (Burton et al., 2019).

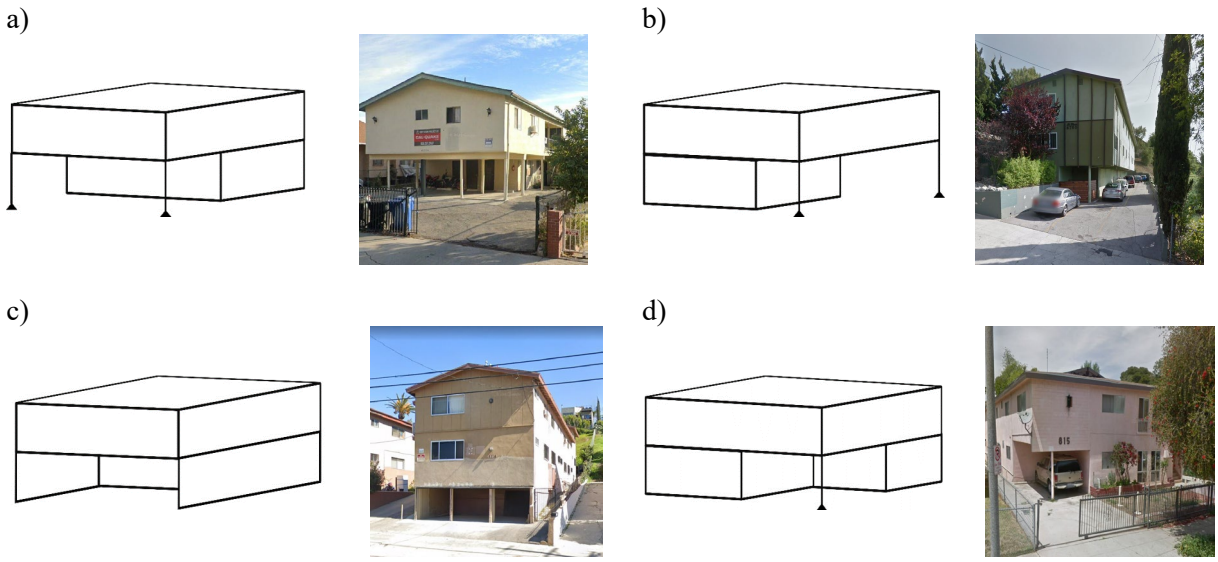


Figure 4.1. Schematic models for first-story wall layouts and their examples within Los Angeles SWOF woodframe buildings inventory for type a) L1, b) L2, c) L3, and d) L4

The 3000 surveyed SWOF buildings by Burton et al. have revealed that 72% and 23% are two and three-story buildings, respectively, and about 5% have more than three stories. Thus, the constructed *OpenSees* models included only two and three-story buildings. In addition, two variations of floor plan dimensions based on the ratio of the length (L) to width (B) of the buildings

are considered. The higher L/B and lower L/B are called large and small aspect ratios, respectively. Lastly, the exterior wall material is assumed to be stucco on the outside and either gypsum wallboard (GWB) or horizontal wood siding (HWS) inside. The interior wall materials are also considered to be constructed with either GWB or HWS on both sides. The features of the thirty-two SWOF building models developed by Burton et al. (2019) are summarized in Table 4.1. The wood-frame wall distribution considered on the first floor and upper floors of the developed models are found in Burton et al. (2019).

Table 4.1. The summary of SWOF building models features constructed by Burton et al (2019)

Building ID	No. of Stories	Plan dimensions	Interior wall material	Exterior wall material
L1-2S-60X30-GWB	2	L = 60′, B = 30′	2 Layers GWB	Stucco+GWB
L1-3S-60X30-GWB	3	L = 60′, B = 30′	2 Layers GWB	Stucco+GWB
L1-2S-100X30-GWB	2	L = 100′, B = 30′	2 Layers GWB	Stucco+GWB
L1-3S-100X30-GWB	3	L = 100′, B = 30′	2 Layers GWB	Stucco+GWB
L1-2S-60X30-HWS	2	L = 60′, B = 30′	2 Layers HWS	Stucco+HWS
L1-3S-60X30-HWS	3	L = 60′, B = 30′	2 Layers HWS	Stucco+ HWS
L1-2S-100X30-HWS	2	L = 100′, B = 30′	2 Layers HWS	Stucco+ HWS
L1-3S-100X30-HWS	3	L = 100′, B = 30′	2 Layers HWS	Stucco+ HWS
L2-2S-60X50-GWB	2	L = 60′, B = 50′	2 Layers GWB	Stucco+GWB
L2-3S-60X50-GWB	3	L = 60′, B = 50′	2 Layers GWB	Stucco+GWB
L2-2S-100X50-GWB	2	L = 100′, B = 50′	2 Layers GWB	Stucco+GWB
L2-3S-100X50-GWB	3	L = 100′, B = 50′	2 Layers GWB	Stucco+GWB
L2-2S-60X50-HWS	2	L = 60′, B = 50′	2 Layers HWS	Stucco+ HWS
L2-3S-60X50-HWS	3	L = 60′, B = 50′	2 Layers HWS	Stucco+ HWS
L2-2S-100X50-HWS	2	L = 100′, B = 50′	2 Layers HWS	Stucco+ HWS
L2-3S-100X50-HWS	3	L = 100′, B = 50′	2 Layers HWS	Stucco+ HWS
L3-2S-50X30-GWB	2	L = 50′, B = 30′	2 Layers GWB	Stucco+GWB
L3-3S-50X30-GWB	3	L = 50′, B = 30′	2 Layers GWB	Stucco+GWB
L3-2S-80X30-GWB	2	L = 80′, B = 30′	2 Layers GWB	Stucco+GWB
L3-3S-80X30-GWB	3	L = 80′, B = 30′	2 Layers GWB	Stucco+GWB
L3-2S-50X30-HWS	2	L = 50′, B = 30′	2 Layers HWS	Stucco+ HWS
L3-3S-50X30-HWS	3	L = 50′, B = 30′	2 Layers HWS	Stucco+ HWS
L3-2S-80X30-HWS	2	L = 80′, B = 30′	2 Layers HWS	Stucco+ HWS
L3-3S-80X30-HWS	3	L = 80′, B = 30′	2 Layers HWS	Stucco+ HWS
L4-2S-60X50-GWB	2	L = 60′, B = 50′	2 Layers GWB	Stucco+GWB
L4-3S-60X50-GWB	3	L = 60′, B = 50′	2 Layers GWB	Stucco+GWB
L4-2S-100X50-GWB	2	L = 100′, B = 50′	2 Layers GWB	Stucco+GWB
L4-3S-100X50-GWB	3	L = 100′, B = 50′	2 Layers GWB	Stucco+GWB
L4-2S-60X50-HWS	2	L = 60′, B = 50′	2 Layers HWS	Stucco+ HWS
L4-3S-60X50-HWS	3	L = 60′, B = 50′	2 Layers HWS	Stucco+ HWS
L4-2S-100X50-HWS	2	L = 100′, B = 50′	2 Layers HWS	Stucco+ HWS
L4-3S-100X50-HWS	3	L = 100′, B = 50′	2 Layers HWS	Stucco+ HWS

4.3 Proposed Methodology for Building's Features Classification

The GSV available dataset visual information received increasing attraction in a wide variety of fields of research, such as housing price estimation (Bency et al., 2017; Law et al., 2019), detecting the soft-story structures (Yu et al., 2020; Chen et al., 2021; Kalfarisi et al., 2022, Tamhidi et al., 2022d), and facilitating the traffic infrastructures (Campbell et al., 2019; Alipour and Harris, 2020). I chose 2,681 buildings randomly out of 13,641 identified SS buildings in Los Angeles to harvest the images out of GSV API to train CNN models for building features classifications. The distribution of the SS buildings and the training sets' buildings within Los Angeles city is shown in Figure 4.2a.

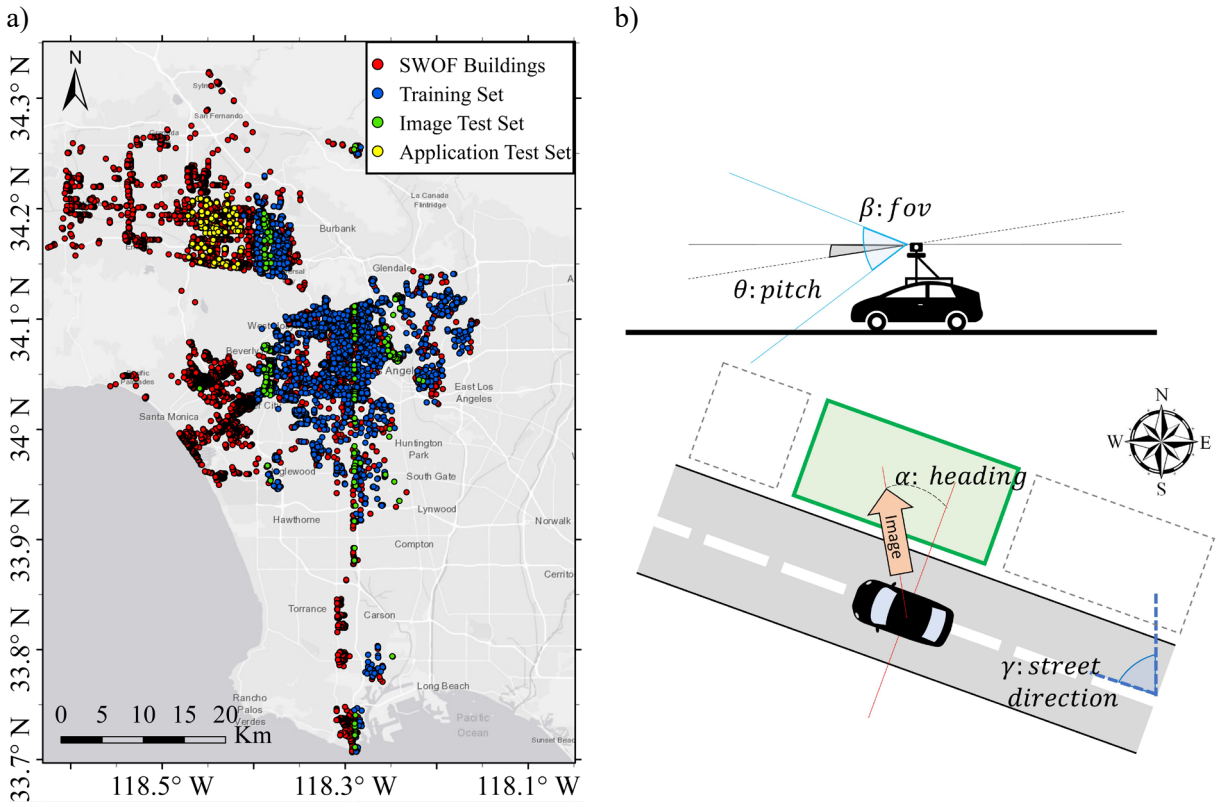


Figure 4.2. a) The distribution of the SWOF building inventory, training and test sets and b) Google Street View imaging parameters for building image collection

4.3.1 Image Datasets Preparation

The GSV API returns one image from the closest available camera location on the street, given the coordinates of the building's center. This one collected image might not be sufficient to detect the opening of the first story due to various reasons, such as the occlusions from trees or trucks. In addition, there are many cases in which the opening of the first story is observable from side (back) streets or another viewpoint on the main street corresponding to their address. Therefore, I collected all available images within an 80m-radius of the building's center from different camera locations around the perimeter of the buildings using their corresponding calculated headings (see Figure 4.2b) to construct the image sets of the building. Figure 4.2b depicts the imaging parameters specified within the Google API for image harvesting. The camera locations are obtained using interpolation from the centerlines of the streets adjacent to the target building (from where the building is observable). The street centerlines are provided by Geographic Information Systems (GIS) layers created by Los Angeles City (L.A. City). The implementation of such methodology provides a rich imagery dataset of the buildings assuring us to collect the images showing the opening of the target SS buildings. Figure 4.3 illustrates an example of a SWOF building in Los Angeles, the obtained camera locations, and their images from GSV API.

Figure 4.3 shows the photos supplied by the GSV API using all interpolated camera positions, including the defaults (green circles) along each street line. The default camera location is the one that corresponds to the building's center coordinates, longitude, and latitude (images 4 and 3 along East-West and North-South streets, respectively). Figure 4.3 demonstrates that although the building is identified as a SS building, the default images cannot demonstrate the first story wall opening. On the other hand, images 5 and 6 along North-South Street display the wall opening in the first story.

Figure 4.3. The default and interpolated camera locations and their harvested images within 80 m-radius of the target building's center along a) EW street, and b) NS street centerlines adjacent to the target building

I defined the fifth type of soft-story wall layout, called "L0", in which the first-story wall layout of the target building does not fall into any of those L1 through L4 (Figure 4.1) categories or the images harvested from the building do not show any opening in the first story wall layout. In about 70% of the 2,681 surveyed buildings (training set in Figure 4.2a), the first-story wall layout was observable within the harvested GSV API images surrounding the building. Among those, about 63% fall within the L1 through L4 classes, and the remaining 7% have shown openings categorized as L0. The 29%, 6%, 45%, and 20% of the buildings categorized as non-L0 are classified as L1, L2, L3, and L4 classes, respectively, which are comparable with those ratios resulting from Burton et al. (2019) study. A total of 5,586 images were selected and manually

annotated as L0 through L4 in order to make the training set for the CNN classifier to identify the first-story wall layout pattern.

76% and 21% of the surveyed buildings are either two or three-story buildings, respectively, whereas 3% of them are either four or five-story buildings. These fractions also are comparable to those identified by Burton et al. (2019). Therefore, I consider two categories of 2S (two-story) and 3S (three-story) for the number of stories classification task. I chose and manually labeled 2,494 images in order to construct the training set for the number of stories classification task. I randomly selected 2% of each training dataset as a validation set to evaluate and choose the best CNN models for the classification tasks.

In addition, 357 test buildings (green locations in Figure 4.2a) other than those chosen for the training set are picked to examine the performance of the trained CNN models. One image is obtained from each of the test buildings, which makes 357 and 350 images for soft-story wall layout and number of stories classification tasks, respectively. Table 4.2 summarizes the number of images available for each classification task within their training, validation, and test subsets. The training and validation datasets are then used to train and choose the best CNN model for the final application. Figure 4.4 and Figure 4.5 show the percentage of each class in the training, validation, and test sets for the first-story wall layout and the number of stories classification tasks, respectively.

Table 4.2. The number of images used for SS wall layout and number of stories classification tasks

	First Story Wall Layout Classification					No. of Story Classification	
	L0	L1	L2	L3	L4	2S	3S
Training Set	973	1346	667	1403	1086	1261	1184
Validation Set	18	30	11	35	17	21	28
Test Set	83	93	21	94	66	273	77
Total	1074	1469	699	1532	1169	1555	1289

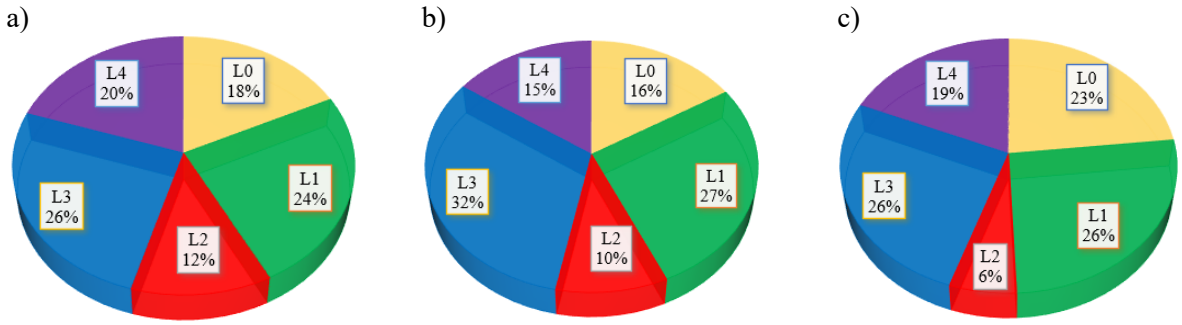


Figure 4.4. Portion of each class in the a) training, b) validation, and c) test sets for the wall layout classification task

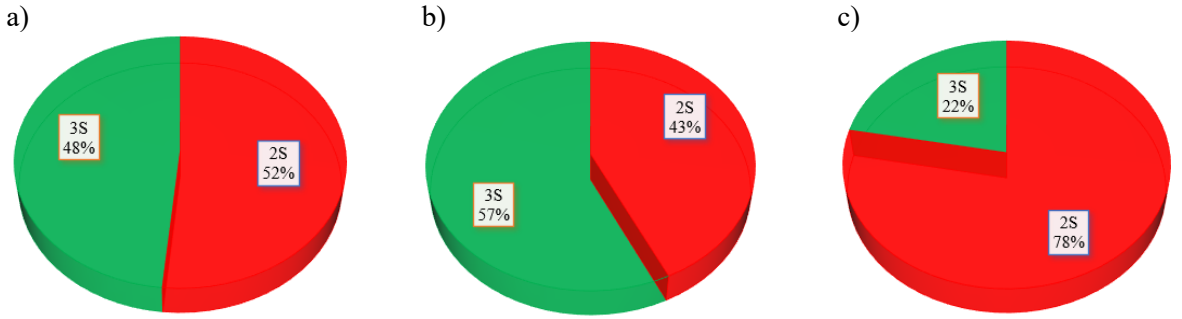


Figure 4.5. Portion of each class in the a) training, b) validation, and c) test sets for the number of stories classification task

4.3.2 Model Training

There are numerous deep CNN architectures developed and used in image recognition tasks such as Residual networks (ResNet50, ResNet101, and ResNet152) (He et al., 2016), Inception V3 (Szegedy et al., 2016), VGG 16 (Simonyan and Zisserman, 2015), Inception-ResNet (Szegedy et al., 2017), and MobileNet (Howard et al., 2017). The Residual and Inception networks were utilized previously by Yu et al. (2020) and Kalfarisi et al. (2021) to classify buildings as soft-story or non-soft-story using the GSV images. In this study I have trained four CNN models: 1) InceptionResNet V2, 2) Inception V3, 3) ResNet50, and 4) ResNet152.

As is shown in Table 4.2, the SS wall layout L2 is the minority class among the five classes in the training set. Thus, two augmentation layers containing random horizontal flip and random zoom in/out are added to each of the networks to increase the number of images at the training level. A transfer learning methodology is used to train each of the networks mentioned above (Torrey and Shavlik, 2010). To do so, previously trained networks' weights using the general-purpose ImageNet dataset (Fei-Fei et al., 2010) are initially used to start the training. The ImageNet dataset includes over a million images annotated for 1000 object classes, which have been used previously to optimize the model parameters of deep neural networks. The training procedure is conducted in two steps: 1) training the top classifier and 2) fine-tuning the initial parameters. In the first step, I train the top classifier, a fully connected neural network (FCNN) ending with five and two neurons for soft-story layout and number of stories classification tasks, respectively. In this step, all convolutional layers are kept "frozen" from training. Then, a portion of the last convolutional layers is trained, called "fine-tuning," based on the provided training set. I used 1e-5 and 1e-6 learning rates and Adam (Kingma and Ba, 2015) and RMSprop optimizers for steps 1 and 2 of the training process, respectively. In addition, a dropout layer with a probability of 20% is used in the FCNN to prevent overfitting the model. One hundred twenty epochs over the training set are implemented using the TensorFlow and high-performance Graphics processing units (GPUs) of NVIDIA Tesla V100.

Table 4.3 and Table 4.4 summarize the validation set's performance evaluation for the four trained CNN models for soft-story wall layout and the number of stories classifications, respectively. In Table 4.3 and Table 4.4, the F1-score is calculated by Eq. (4.1)

$$F1 = \frac{2 \times P \times R}{(P + R)} \tag{4.1}$$

where P and R denote the precision and recall. The precision equals the fraction of predicted instances as i^{th} class that truly belongs to the i^{th} class. On the other hand, recall is the fraction of i^{th} class instances that are truly estimated as i^{th} class. The accuracy in Table 4.3 and Table 4.4 refers to the percentage of correct estimations among all classes.

Table 4.3. F1-score and accuracy of validation set for various CNN architectures for soft-story wall layout classification

Model	Accuracy	F1-Score				
		L0	L1	L2	L3	L4
InceptionResNet-V2	0.86	0.80	0.90	0.91	0.84	0.91
Inception-V3	0.84	0.79	0.87	0.80	0.86	0.88
ResNet50	0.85	0.77	0.93	0.87	0.82	0.88
ResNet152	0.86	0.86	0.88	0.90	0.86	0.89

Table 4.4. F1-score and accuracy of validation set for verious CNN architectures for number of stories classification

Model	Accuracy	F1-Score	
		2S	3S
InceptionResNet-V2	0.95	0.95	0.96
Inception-V3	0.95	0.95	0.96
ResNet50	0.91	0.91	0.93
ResNet152	0.95	0.95	0.96

As is shown in Table 4.3 and Table 4.4, the ResNet152 and InceptionResNet-V2 CNN models performed best on the Validation sets for both classification tasks. In this study, I selected the trained ResNet152 model to perform on the test set. Figure 4.6 demonstrates the ResNet152 architecture utilized in this study. In Figure 4.6, the top classifier ends with two sets of estimating neurons. The first five neurons output the probability of the image belonging to each of SS wall layout's classes after applying the SoftMax function. Similarly, the second two neurons provide the probability of being either two or three-story buildings for the given image.

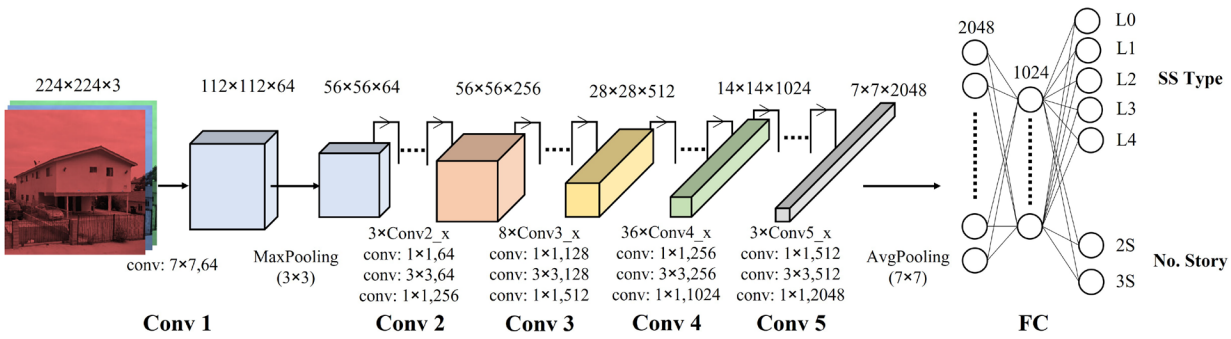


Figure 4.6. ResNet152 architecture utilized for soft-story wall layout and number of stories classification tasks

4.4 Performance Evaluation

I evaluate the performance of the reference ResNet152 model on the independent test datasets. I examine the accuracy of the predictions in two levels: 1) image level and 2) building level. In the first level, I utilize the trained reference model to estimate the soft-story wall layout and the number of stories classes using the 357 images harvested from the test set's buildings (green points in Figure 4.2a). Then, I utilize the reference model to estimate the target buildings labels regarding the first-story wall layout and number of stories. To do so, I randomly selected 120 buildings (yellow points in Figure 4.2a) as an application test set to examine our proposed model and methodology in predicting the features of the SS buildings (see Building Level Performance Evaluation Section).

4.4.1 Image Level Performance Evaluation

Table 4.5 illustrates the estimation results of the trained ResNet152 model on the first-story wall layout and the number of stories classification tasks. It is observed that the F1-score for the wall layout classification task is almost more than 0.8 for the three classes of L0, L1, and L3. The class L2, which was the minority class in both training and test sets, was predicted with lower accuracy (F1-score of 0.68). The average accuracy for the wall-layout classification task, the

number of correct estimations divided by the total number of test images, is 80.1%. On the other hand, the overall accuracy for the number of stories classification task is 91.4%. The ResNet152 model demonstrated better performance on the prediction of two-story buildings, yet its estimation accuracy for the three-story buildings is still acceptable.

Table 4.5. Performance of ResNet152 model on first story wall layout and number of stories classifications on image test set

	First Story Wall Layout Classification					No. of Story Classification	
	L0	L1	L2	L3	L4	2S	3S
Precision	0.91	0.80	0.70	0.78	0.75	0.98	0.74
Recall	0.77	0.94	0.67	0.81	0.68	0.91	0.94
F1-Score	0.84	0.86	0.68	0.79	0.71	0.94	0.83
Accuracy	80.1%					91.4%	

Figure 4.7 displays the confusion matrix, C, of the test set's predictions for the first-story wall layout classification. The confusion matrix element C_{ij} shows the recall for the i^{th} class, i.e., the fraction of i^{th} class images in the test set that is correctly labeled by the model. In Figure 4.7, the number of images constructed in each cell is shown in parentheses. As Figure 4.7 depicts, the confusion matrix includes the maximum recall on its diagonal elements, which confirms the applicability of the trained model in detecting the first-story wall layout. A sample of correctly predicted classes within the image test set is provided in Figure 4.8 with the probability of the true estimated class (after applying the SoftMax function on the last layer of the FCNN). Figure 4.7 shows that the top four most probable model's misclassification is classifying L0, L2, L3, and L4 mistakenly by L3, L4, L1, and L1, respectively.

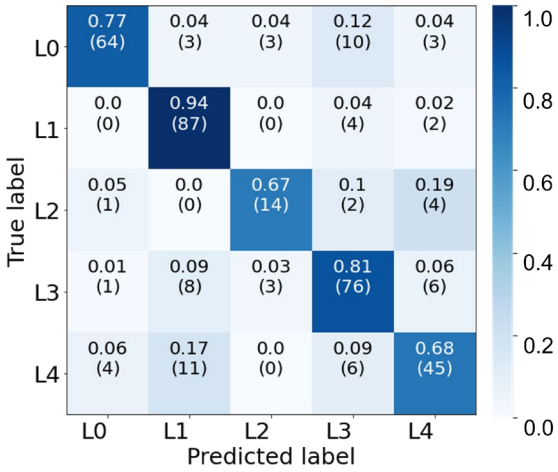


Figure 4.7. Confusion matrix for image test set of first story wall layout classification

Figure 4.8. Samples of correctly classified test set's images regarding their first-story wall layout for class a) L1, b) L2, c) L3, and 4) L4

In order to better acquire the reason for the classification and interpretability of the proposed model, I generated the Gradient-weighted Class Activation Maps (Grad-CAM) (Selvaraju et al., 2020) for a sample of images correctly classified in the image test set. Grad-CAM

is a well-established methodology that utilizes the gradient information of a class activation with respect to different particles (pixels) of an image in order to produce a heatmap highlighting the regions essential for the classifier's decision about the estimated label. Figure 4.9a displays the correctly classified images Grad-CAM corresponding to the estimated (correct) class, which confirms that the CNN model is successfully utilizing the visual features of the first-story openings to classify the SS building layout. In addition, Figure 4.9a shows that when an image allocates to L0 class, the CNN classifier looks for objects other than buildings, such as trees and trash cans. Figure 4.9b through Figure 4.9e illustrates an example of each top four most probable model's mistakes mentioned above.

Figure 4.9b shows how an L0 picture might be misclassified as L3 if the entire first-story wall arrangement is not visible. The reason for that is L3 has a single open wall line with walls on three sides. Figure 4.9b demonstrates that if the entire first-story wall arrangement is not supplied into the model, it may capture an opening and incorrectly identify it as L3, even if the image belongs to an L0 structure. Figure 4.9c shows a sample of actual L2 that has been approximated to be L4. The camera perspective cannot witness the whole depth of the structure, which is required to identify the wall-line opening along the longer dimension of the building, as stated for class L2. Figure 4.9d and Figure 4.9e show L3 and L4 misclassified as L1, respectively. The L1 has been mislabeled in both of these examples because the side walls are not appropriately identified due to the camera's heading and zoom level (FOV in Figure 4.2b).

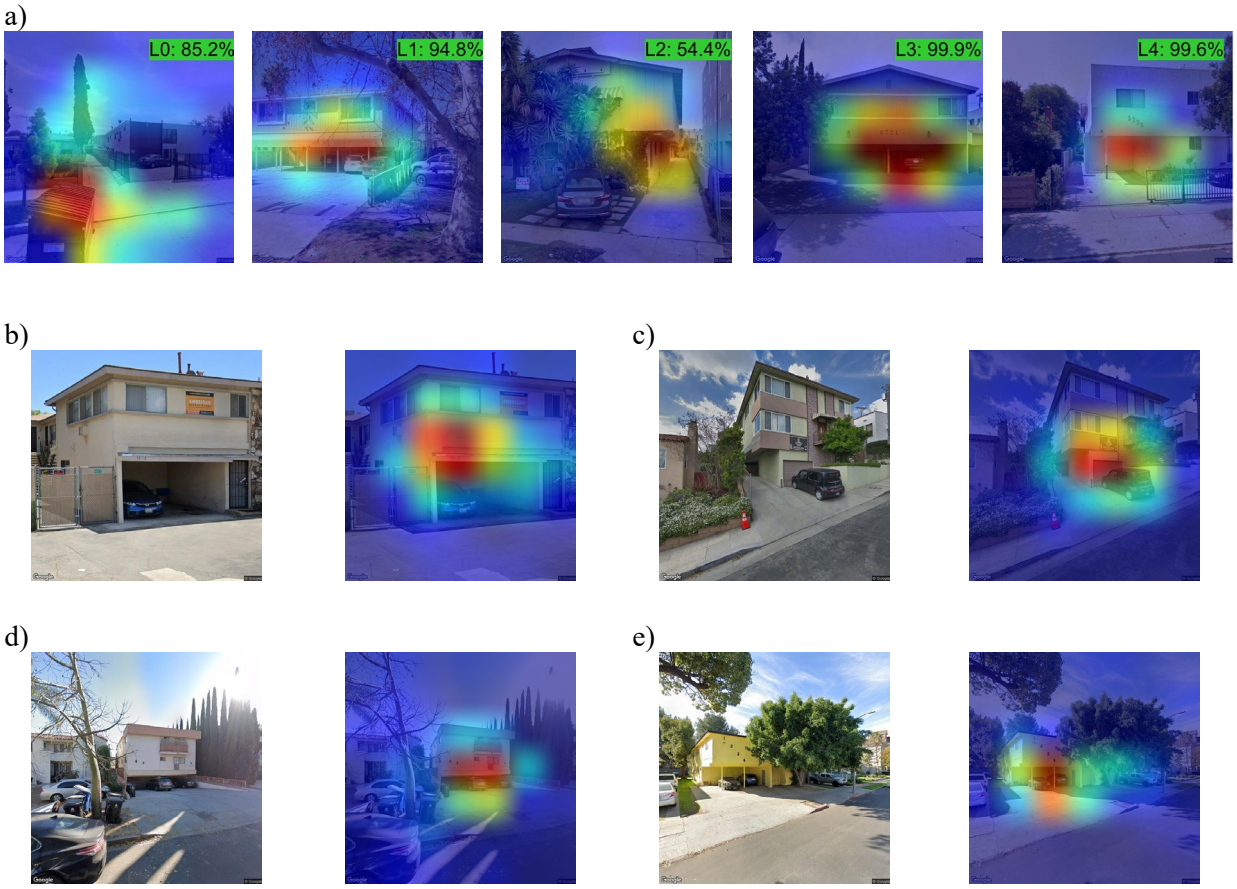


Figure 4.9. a) Correctly classified images and their Grad-CAM. Misclassified images of class b) L0 as L3, c) L2 as L4, d) L3 as L1, and d) L4 as L1 Grad-CAM

Figure 4.10a displays the confusion matrix for the number of stories classification task's test set. Figure 4.10a shows that the trained CNN has a descent performance on classifying two-story building images. In addition, Figure 4.10b shows two samples of correctly classified images and their estimated probability by the trained ResNet152 model. As is shown in the confusion matrix of the test set, there are 25 and 5 two-story and three-story misclassified buildings. Figures 4.11a and 4.11b demonstrate two-story and three-story buildings that are misclassified as three-story and two-story buildings, respectively.

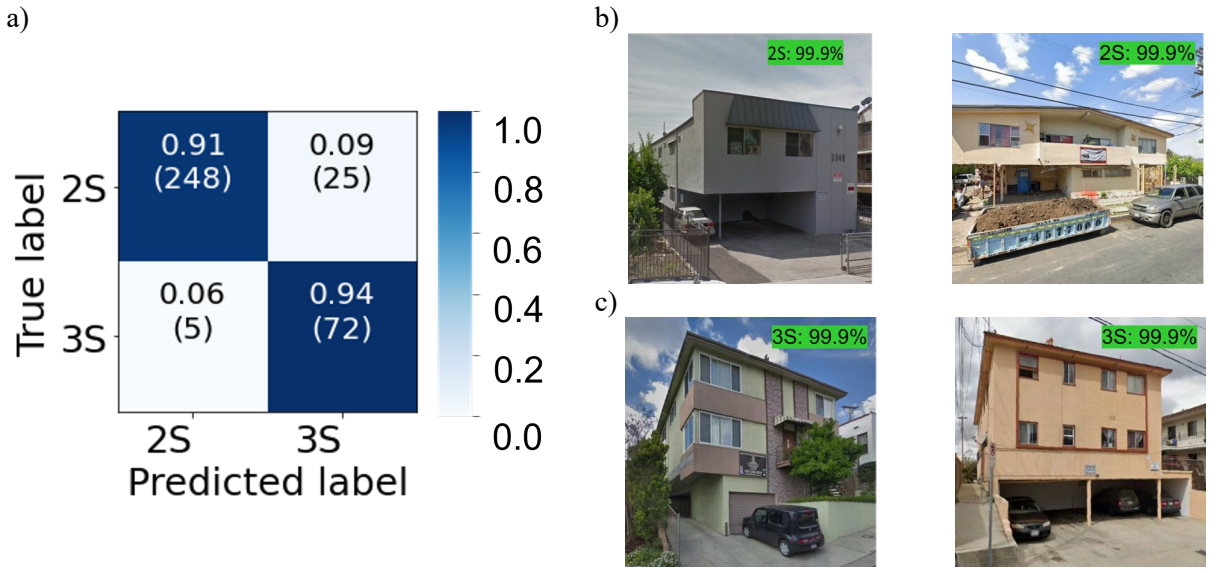


Figure 4.10. a) Test set's confusion matrix for number of stories classification. b) True two-story and c) True three-story classified building images

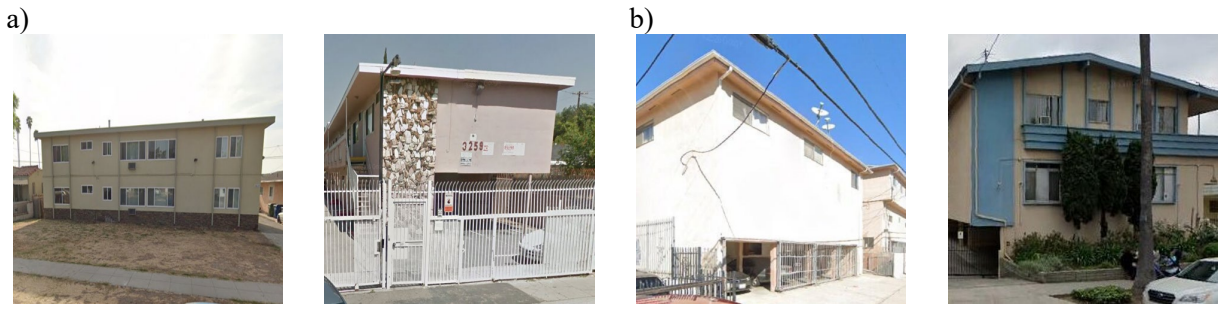


Figure 4.11. Samples of a) true two-story misclassified as three-story and b) tru three-story misclassified as two-story buildings

Figure 4.11a shows that the main reason for detecting a two-story building as a three-story one is to misunderstand the first floor. As shown by Figure 4.11a, the brick stripe façade or the front door fence is recognized mistakenly as a separate story by the model. In addition, Figure 4.11b shows that if the building story's windows' row is not observable or if the first story is not completely seen (especially if it is below the ground level), it might make the CNN model miss a story.

In order to detect a building's first-story wall layout and the number of stories, the CNN model must receive and classify an ensemble of photos harvested surrounding the building (cf. Figure 4.3), which reduces the misclassification probability for the target buildings.

4.4.2 Methodology to Extract Building Features

In this section, I describe how the first three critical features of the *OpenSees* models for each target building are classified using the trained models. As shown previously, images within an 80 m radius of the neighborhood from the center of each target building are extracted. First, all camera locations at every seven meters steps along the two adjacent streets are determined. Then, six images from each of those locations are downloaded. These six images are obtained by rotating the camera's heading with ±7º at two zoom levels (FOV) of 30 and 50. A subset of the test set's buildings (green points in Figure 4.2a) is used to evaluate various algorithms and procedures for labeling each building, which resulted in the chosen following algorithm to determine each building's features.

For the first-story wall layout classification purpose:

1- Each extracted image is fed into the trained CNN model, and the probability of each class is obtained (the last SoftMax layer's output).
    1.1- If the estimated probability of class L0 for an image is more than 60%, then the image is labeled as L0; otherwise, it is labeled as a non-L0 image.
    1.2- If more than 85% of a building's extracted images are labeled as L0, then consider the building as L0; otherwise, move on to step 2.
2- The average probability of each L1 through L4 class is obtained for all non-L0 images. The building class is considered the one with the higher average probability among all images.

Figure 4.12 demonstrates the flowchart of the procedure described above for the first-story wall layout classification.

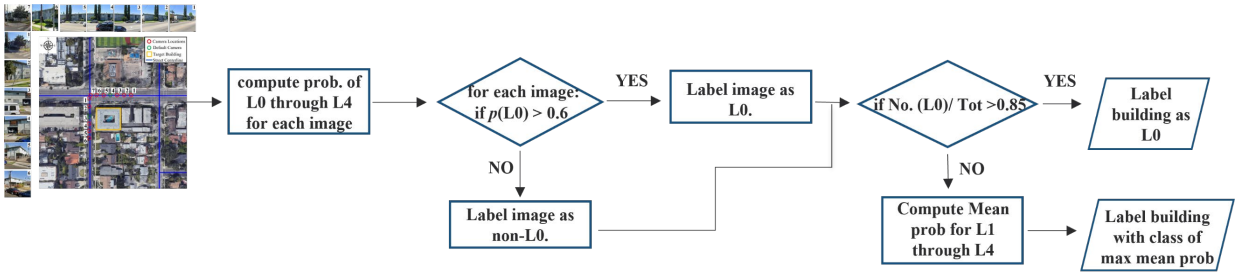


Figure 4.12. The flowchart for the buildings first-story wall layout classification

For the number of stories classification purpose:

1- The images from the default and its two adjacent camera locations along the surrounding streets are obtained (a total of six locations).

2- The probability of being 2S (two-story) or 3S (three-story) for each image is calculated by the trained model.

3- The building's number of stories is labeled as the one with the higher average probability among all provided images in step 1.

The building's third feature, the floor plan dimension, must be chosen as either longer or shorter aspect ratio, as provided in Table 4.1. To do so, I obtain the building's footprint coordinates from *OpenStreetMap* (OSM), providing geographical coordinates of urban facilities, including residential buildings. Figure 4.13 demonstrates a sample of the training set's buildings' footprints which are extracted via OSM. The buildings' floor plan area can then be calculated by having the coordinates of the footprints' vertices.

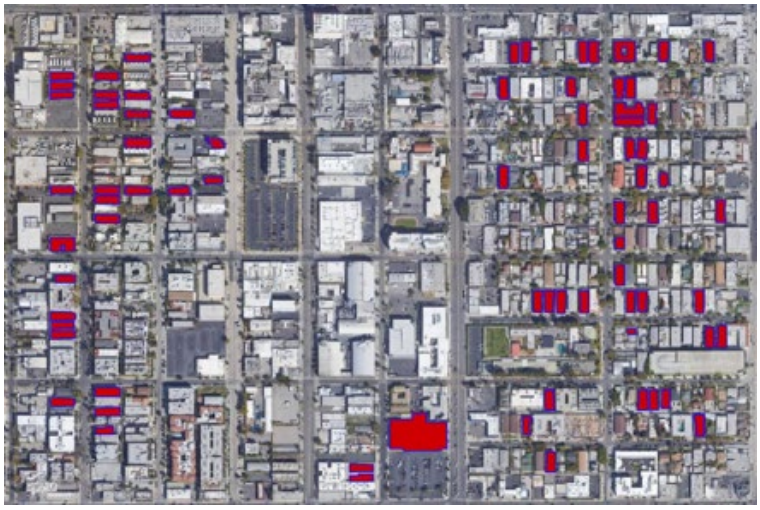


Figure 4.13. The buildings footprint extracted from *OpenStreetMap* for a subset of training set's buildings

The key elements that play the main role in the lateral seismic resistance of wood-frame residential buildings are the wood shear walls. A range of 0.12 ft/ft^2 to 0.20 ft/ft^2 wall densities for the developed *OpenSees* models are considered, which is comparable to the actual existing buildings (Burton et al., 2019). I choose the plan dimension (cf. Table 4.1) that constructs the closest area to the existing building's area. Thus, the total shear wall length would be close between the structural model and the actual building. However, it is worth mentioning that the effect of the floor plan aspect ratio in the actual building cannot be captured by choosing the dimensions through the mentioned method.

Eventually, the last required component, the shear wall materials, remains unknown. There are several resources that can provide such detailed information, such as Zillow's Transaction and Assessment Database (ZTRAX) (Zillow, 2021) or the California Department of Tax and Fee Administration; however, our investigation shows that such information can be missed for a majority target residential buildings. Thus, I will analyze the closest structural models (considering the first three features) with both shear wall materials considered in Burton et al. (2019) study.

4.4.3 Building Level Performance Evaluation

In this section, I evaluate the performance of our methodology described in the previous section on a separate buildings' test set, including 120 buildings (yellow points in Figure 4.2a). To do so, I obtained all images surrounding each target building using the methodology described in the previous section. Then, I fed the harvested images to the trained ResNet152 model to label each building for the first-story wall layout and number of stories classification tasks (cf. section 4.4.2). In another attempt, I passed the same images extracted for each building to a trained intern to classify the buildings for the same tasks. The latter is due to evaluate the performance of a human in detecting the structural features out of the same images. Eventually, I classified the buildings as a reference "ground truth" label. Figures 4.14a and 4.14b display the confusion matrices of the CNN model and human predictions for the 120 buildings (compared with the "ground truth" labels) for first-story wall layout classification.

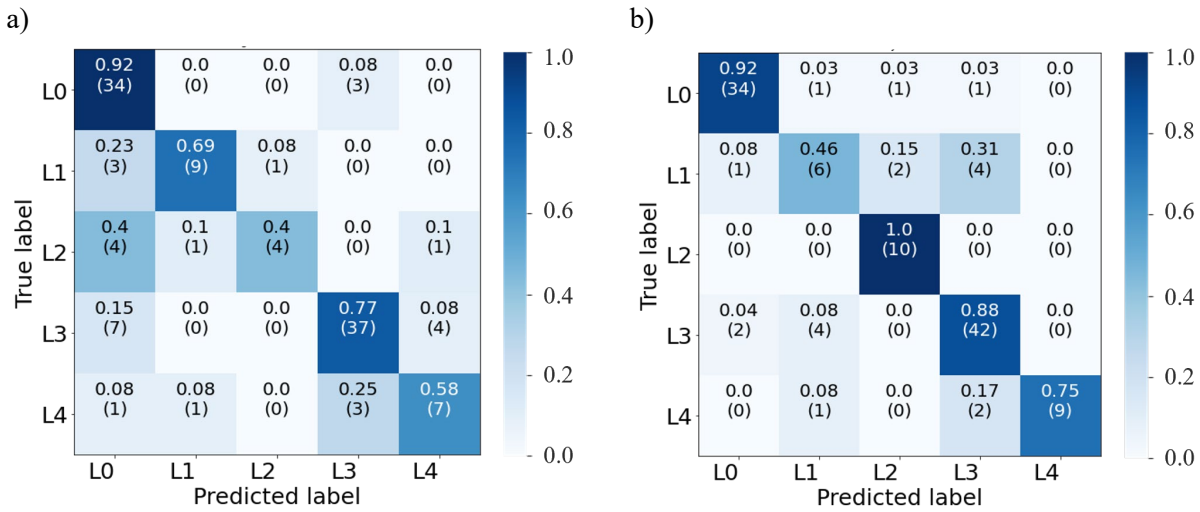


Figure 4.14. Confusion matrix of first-story wall layout classification for 120 building test sets for a) CNN model and b) Human predictions

In addition, Table 4.6 and Table 4.7 summarizes the precision, recall, and F1-score of the CNN model and human performance, respectively. It is observed that the CNN model accuracy, 0.76, is

close to the human's classification accuracy, 0.84, which technically is considered as a performance limit that a machine learning model can gain.

Table 4.6. Building level performance of ResNet152 model on first story wall layout classification over 120 building test sets

	First Story Wall Layout				
	L0	L1	L2	L3	L4
Precision	0.69	0.82	0.80	0.86	0.58
Recall	0.92	0.69	0.40	0.77	0.58
F1-Score	0.79	0.75	0.53	0.81	0.58
Accuracy			0.76		

Table 4.7. Building level performance of Human on first story wall layout classification over 120 building test sets

	First Story Wall Layout				
	L0	L1	L2	L3	L4
Precision	0.92	0.50	0.77	0.86	1.0
Recall	0.92	0.46	1.0	0.88	0.75
F1-Score	0.92	0.48	0.87	0.87	0.86
Accuracy			0.84		

Figure 4.14 demonstrates that both CNN and the human performance model show two considerably off-diagonal recall values. The CNN model mostly mislabeled L2 buildings with L0. The main reason for such misclassification is that L2 buildings must be seen from a viewpoint that the whole depth of the wall opening is present in the image. Thus, as were seen previously, it is possible to misclassify such buildings as either L0 or L4. The human performance confusion matrix also shows that the actual L1 might be misclassified as L3. After investigating the images, I found that for the L1 buildings with a deep opening in their first story (below the hanging part) surrounded by internal walls, it might be confusing to distinguish between classes L3 and L1. An example of such a situation can be seen in image 5 along NS street in Figure 4.3.

Figures 4.15a and 4.15b demonstrate the confusion matrices of the CNN model and human predictions for the 120 buildings for the number of stories classification.

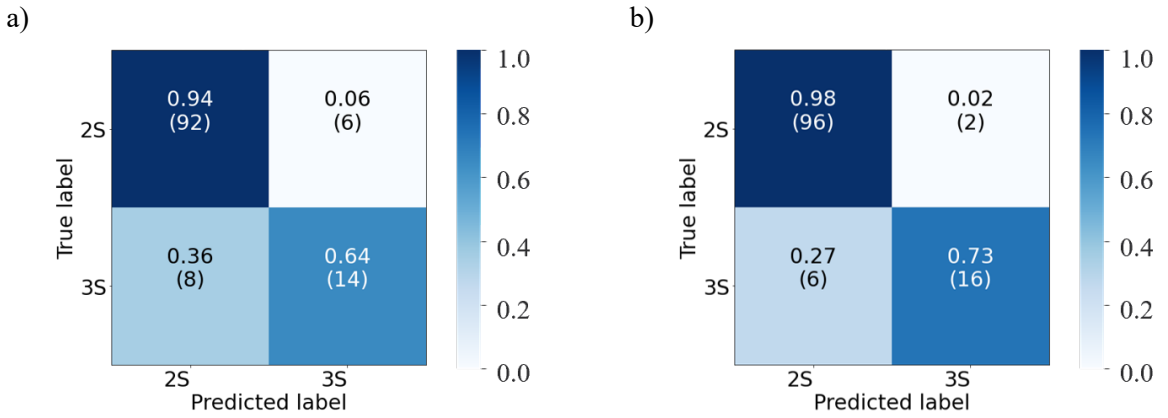


Figure 4.15. Confusion matrix of number of stories classification for 120 building test sets for a) CNN model and b) Human predictions

In addition, Table 4.8 and Table 4.9 summarizes the precision, recall, and F1-score of the CNN model and human performance for the number of stories classification, respectively. It is observed that the CNN model accuracy, 0.88, is close to the human classification accuracy, 0.93. It is also noticed that the estimation of the model, and in some cases of humans, on detecting three-story buildings can be wrong. The main reason for missing a story is that the row of windows might not be observable, or the floor exists below the ground level for parking usage.

Table 4.8. Building level performance of CNN model on number of stories classification over 120 building test sets

	Number of Stories	
	2S	3S
Precision	0.92	0.70
Recall	0.94	0.64
F1-Score	0.93	0.67
Accuracy	0.88	

Table 4.9. Building level performance of Human on number of stories classification over 120 building test sets

	Number of Stories	
	2S	3S
Precision	0.94	0.89
Recall	0.98	0.73
F1-Score	0.96	0.80
Accuracy	0.93	

I developed and evaluated a methodology to classify the first three main features required to allocate a structural *OpenSees* model for the target SS buildings. Next, I will evaluate the whole framework for a subset of target SS buildings as a testbed to estimate their performance for a scenario earthquake in Los Angeles.

CHAPTER 5. Framework for Near-Real-Time Soft Story Structural Damage Assessment

5.1 Scope and Overview

In this section, I aim to apply the developed framework using the established ingredients in previous chapters. In this study, I apply the framework for existing un-retrofitted SS structures as the thirty-two developed *OpenSees* models by Burton et al. (2019) (see Chapter Chapter 4) corresponds to the existing (un-retrofitted) SS buildings. The Los Angeles Department of Building and Safety (LADBS) provides the dates for the retrofits' permit and completion for each SS building. I first filtered the buildings the retrofit completion dates of which are earlier than August 2021. Then, out of approximately 8,100 remained un-retrofitted SS buildings in Los Angeles city, I selected a subset of about 2,000 buildings to apply the methodology on. It is worth mentioning that these randomly selected buildings are chosen outside of the training sets utilized for CNN models (blue points in Figure 4.2a). Figure 5.1b demonstrates the distribution of the chosen SS buildings as the testbed in Los Angeles. The SS locations are selected from the Inner domain illustrated in Figure 3.21b, where the density of the recording instrumentation is highest (see Chapter 3). Thus, the generated ground motions at un-instrumented locations would be more accurate and reliable.

The combination of CSN and CISN recording sites is utilized to generate the ground motion time series at each of the SS buildings' locations. Figure 5.1a illustrates the location of the SS buildings with respect to the seismic network recording sites. In this Section, the following steps are taken to prepare and apply the framework developed in previous chapters:

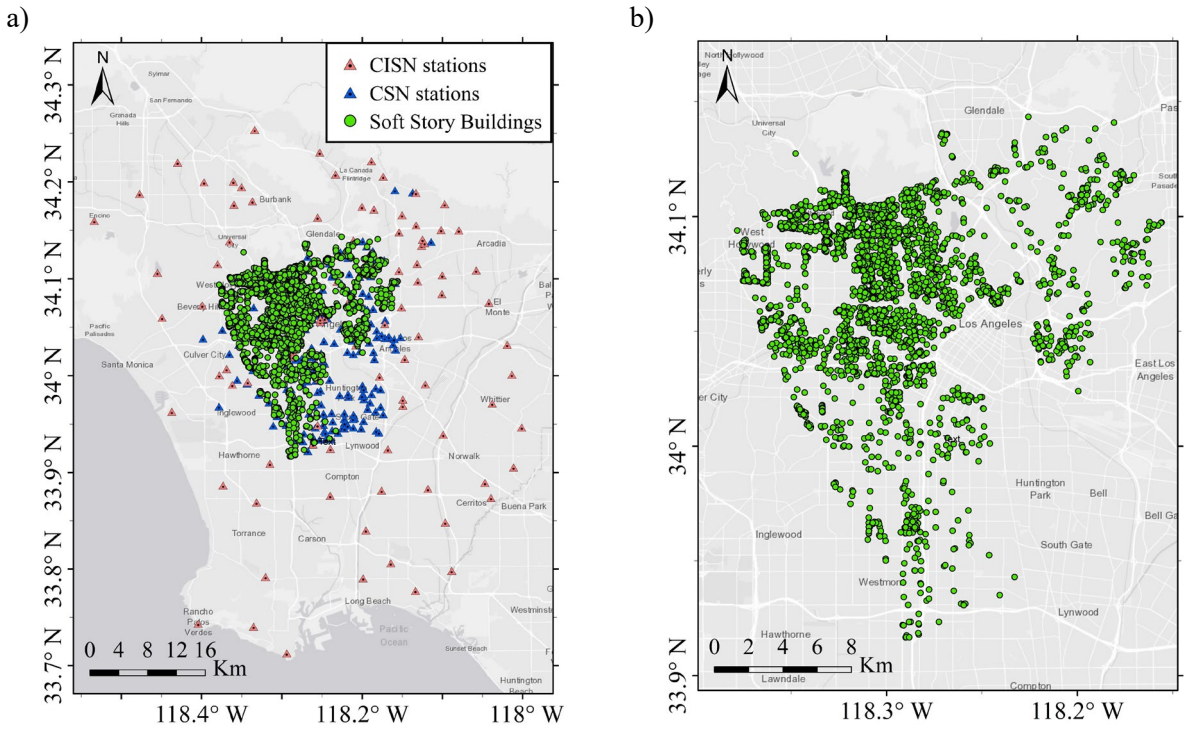


Figure 5.1. a) The distribution of chosen Soft-Story buildings as testbed and CISN and CSN's recording sites in Los Angeles b) a zoomed in view of testbed SS buildings

1. A **M**6.7 scenario earthquake is synthetically generated by amplifying the recorded ground motions of 2020 **M**4.5 South El Monte at CISN and CSN sites. The Bayless and Abrahamson (2019b) Fourier amplitude spectrum prediction model is utilized in order to magnify the recorded motions.

2. The GPR model established in Chapter 2 is utilized with the hyper-parameter, $\hat{\lambda}$, corresponding to the target region's observation density to generate the ground motion time series at each of the SS building's locations, given the CISN and CSN sites' motions as observation.

3. The structural features, including first-story wall layout, number of stories, and floor plan dimension, are obtained using extracted GSV images and the OSM platform as described in Chapter 4. The closest *OpenSees* models (considering both wall materials) out of the thirty-two models is chosen for each SS building in the testbed.

4. Nonlinear response history analyses for all 2,000 SS buildings are conducted using the simulated ground motions and the selected structural models. The damage state of the buildings is estimated using the structural responses.

5. The distribution of the structural performances is plotted as a map where one can find which buildings are probably safe, severely damaged, or potentially collapsed.

Figure 5.2 schematically summarizes the whole framework described above to map the performance estimates of the chosen SS buildings in Los Angeles.

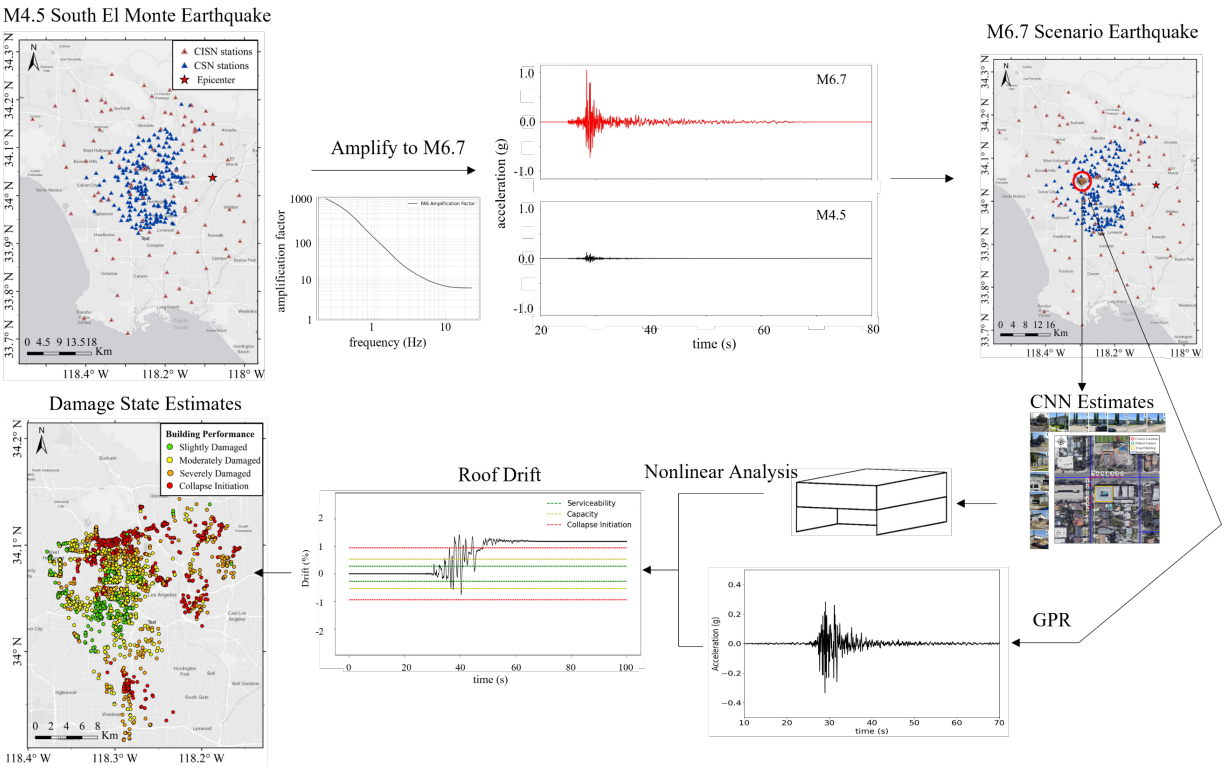


Figure 5.2. The schematic framework for estimating the damage state of the testbed SS buildings in Los Angeles

In the following sections, I will describe each component of the steps above. Next, I will explain how the **M**4.5 2020 South El Monte earthquake records are employed to generate a new synthetic **M**6.7 earthquake in Los Angeles.

5.2 M6.7 Scenario Earthquake in Los Angeles

I constructed a scenario **M**6.7 earthquake in Los Angeles. A local earthquake with such a magnitude is expected to cause some damage to the SS buildings which are the structural focus of this research. It is also recalled that the last "major" earthquake in urban Los Angeles was the 1994 Northridge earthquake with a magnitude 6.7. In order to develop such earthquake ground motions at the locations of the CISN and CSN sites (see Figure 5.1a), I amplify the recorded motions of the 2020 **M**4.5 South El Monte earthquake.

The features of the South El Monte earthquake are summarized in Table 3.8. USGS provides the nodal planes of focal mechanism characteristics for this event. Figure 5.3 and Table 5.1 illustrate the "beachball" of the fault planes and their strike, dip, and rake angles (USGS, 2020). It is observed that the South El Monte earthquake is a reverse faulting earthquake with approximately an EW fault strike. In summary, the scenario **M**6.7 earthquake is expected to have a similar focal mechanism to the **M**4.5 South El Monte earthquake, which directly impacts the seismic energy radiation pattern and directivity (if any). The scenario **M**6.7 earthquake also is constructed based on the point-source assumption, which does not consider finite fault, which is expected for an earthquake with magnitude 6.7.

a) b)

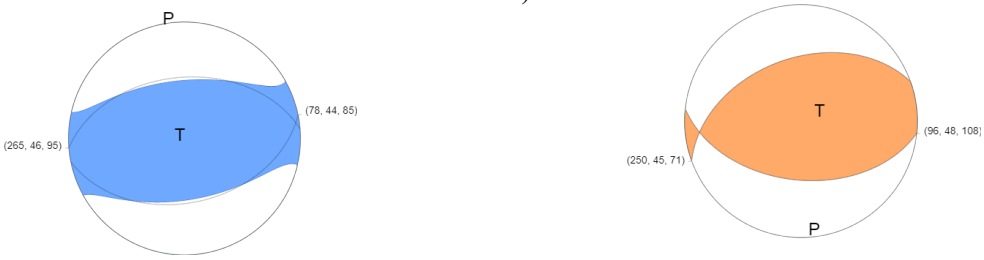


Figure 5.3. a) Moment Tensor and b) Focal Mechanism of the 2020 **M**4.5 South El Monte earthquake (USGS, 2020)

Table 5.1. Focal Mechanism's nodal planes features for South El Monte earthquake (USGS, 2020)

Plane	Strike	Dip	Rake
NP1	250	45	71
NP2	96	48	108

Bayless and Abrahamson (2019b) developed an empirical ground motion model (GMM) of Fourier Amplitude Spectra for shallow crustal earthquakes utilizing the Next Generation Attenuation-West 2 (NGA West 2) (Bozorgnia et al., 2014), known as the BA18 model, to estimate the effective amplitude spectrum (EAS) defined by Goulet et al. (2018). The EAS is defined as

$$EAS(f) = \sqrt{\frac{1}{2}[FAS_{HC1}(f)^2 + FAS_{HC2}(f)^2]} \tag{5.1}$$

where the FAS_{HC1} and FAS_{HC2} are the orthogonal horizontal components' Fourier amplitude spectra of the ground motions. The BA18 model provides the median estimated EAS by Eq. (5.2).

$$\ln(EAS_{med}) = f_M + f_P + f_S + f_{Z_{tor}} + f_{NM} + f_{Z_1} \tag{5.2}$$

In Eq. (5.2), $f_M, f_P, f_S, f_{Ztor}, f_{NM}, f_{Z1}$, are the Magnitude scaling, Path Scaling, Site Response, Depth to Top of Ruptures Scaling, Normal Style-of-Faulting Effect, and Soil Depth Scaling factors. The description of each factor and the algorithm to obtain them is described in detail in Bayless and Abrahamson (2019b).

In this study, I utilized the BA18 model to simulate the EAS of the ground motions at each CISN and CSN station for the **M**4.5 South El Monte earthquake. Then, I predict the EAS of the motions corresponding to a scenario **M**6.7 earthquake at the same sites. Eventually, the ratio between the two estimated EAS proposes the amplification factor. After amplifying the FAS of the motions, the initially recorded phase spectra are used to convert the Fourier Spectra to the time domain associated with the scenario **M**6.7 earthquake.

The Z_{tor} for the South El Monte earthquake is assumed to be 14.0 km (about 3 km shallower than its hypocentral depth), while the Z_{tor} of the scenario earthquake is assumed to be 5 km (similar Z_{tor} for the Northridge earthquake based on USGS, 2013). The epicenter location of the scenario earthquake and faulting type are considered the same as those of the South El Monte earthquake. The $Z_{1.0}$, depth to the horizon with a shear-wave velocity of 1.0 Km/s, is estimated using the SCEC CVMS-4 model (Lee et al., 2014). The Vs_{30} of the sites are also predicted through the proxy-based model, as described in Ahdi et al. (2020). Figure 5.4 displays the distribution of the estimated Vs_{30} and $Z_{1.0}$ at each CISN and CSN site's locations.

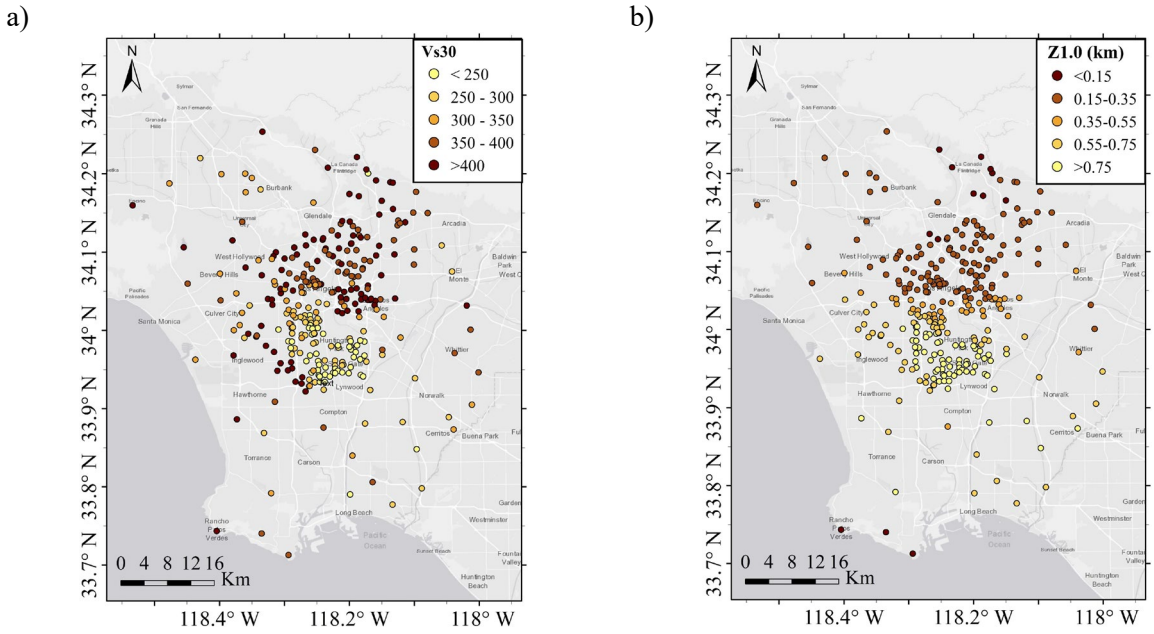


Figure 5.4. The distribution of estimated a)Vs_{30} and b) $Z_{1.0}$ at CISN and CSN sites

Figure 5.5a and b demonstrate the distribution of the RotD50 intensity measure (IM) at $T = 0.2$ s for **M**4.5 South El Monte and the scenario **M**6.7 earthquake within Los Angeles. Figure 5.5 shows that the RotD50 IM has been intensified approximately ten times for most of the sites within the epicentral region after amplifying the magnitude to 6.7. I chose five test sites in Figure 5.5 to demonstrate the amplification factor (ratio of BA18 estimated EAS between **M**6.7 and **M**4.5),

PSA, and FAS along NS directions. Figure 5.6 shows the results for the five chosen test sites in Figure 5.5.

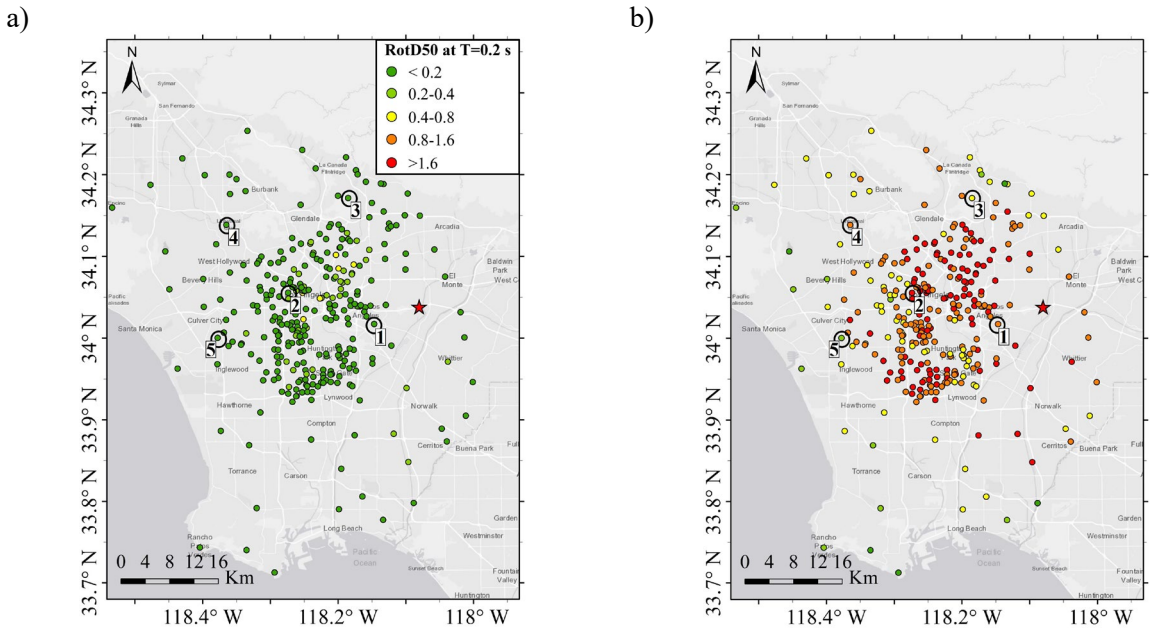


Figure 5.5. The distribution of the RotD50 at $T = 0.2$ s for a) **M**4.5 South El Monte and b) scenario **M**6.7 earthquakes at CISN and CSN sites in Los Angeles

Figure 5.6 illustrates that the long-period contents of the motions (short frequencies) are intensified by about hundreds of times with respect to the initial records. On the other hand, the higher frequency contents of the motion are magnified approximately ten times. Such amplifications are also observable in the pseudo-spectral acceleration response spectrum (PSA) that the long-period ordinates are amplified more than those of short periods. It is worth mentioning that the amplification factors resulted from average of three NGA West 2 GMMs, CB14, BSSA14, and ASK14 are similar and consistent to those resulted from BA18 model in terms of RotD50 response spectral ordinates within usable period bandwidth. The ground motion time series along the NS direction for both earthquakes at the chosen target sites are shown in Figure 5.6. It is observable in Figure 5.6 that how both the amplitude and the significant duration of the time series increased for the higher magnitude scenario earthquake.

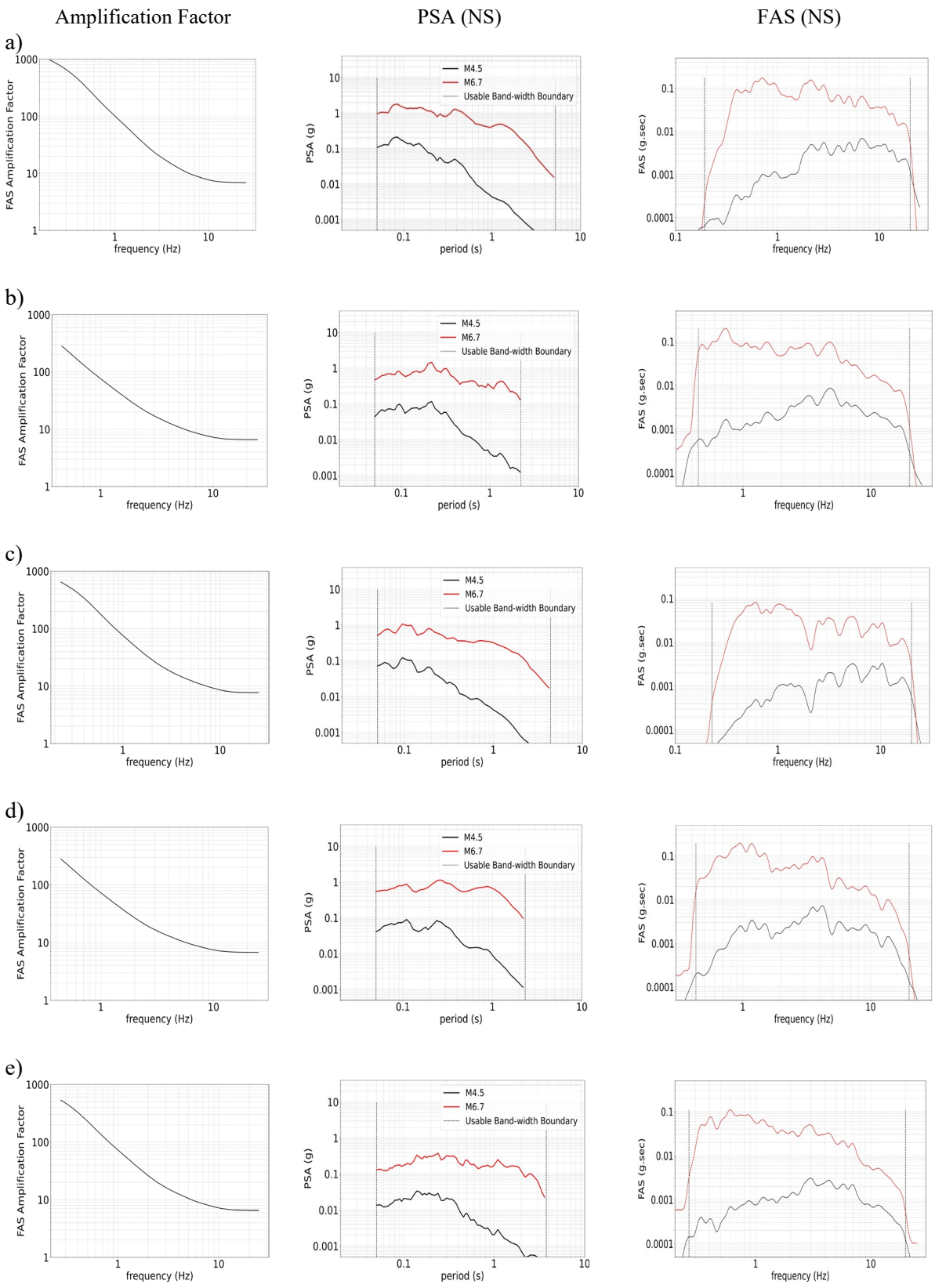


Figure 5.6. Frequency amplification factor, PSA, and FAS at test sites a) 1, b) 2, c) 3, d) 4, and e) 5

Next, the amplified ground motions for scenario **M**6.7 earthquake at CISN and CSN sites are employed to generate ground motion time series at all testbed SS building locations through the trained GPR model. The corresponding $\hat{\lambda}$ using the observation density and Table 2.3 is picked for that purpose.

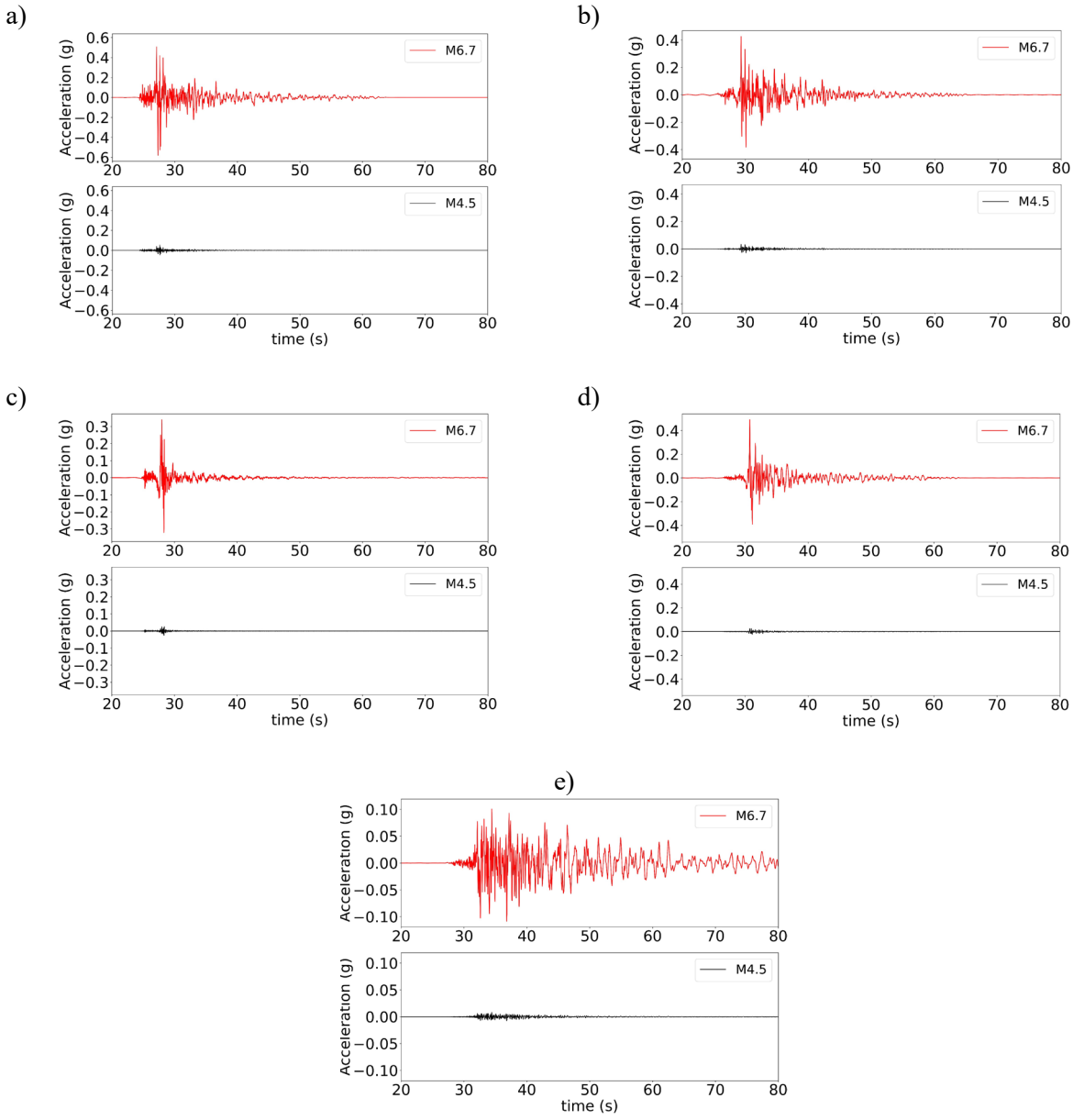


Figure 5.7. The ground motion time seris along North-South direction for both **M**4.5 and **M**6.7 earthquakes at test sites a) 1, b) 2, c) 3, d) 4, and 5) e shown in Figure 5.5

5.2.1 Conditioned Simulation of Ground Motion at SS buildings

In this section, I describe the simulation of ground motion time series at each of the SS buildings' locations. The target SS buildings (Figure 5.1a) are selected from the "inner domain" for **M**4.5 South El Monte earthquake ground motion simulation (cf. Figure 3.19b). All the CISN and CSN sites shown in Figure 5.1a are utilized as the observed ground motions to generate the ground motions at target sites for the scenario **M**6.7 earthquake. Thus, as previously used and shown in Table 3.9, the optimum $\hat{\lambda}$=0.08 is employed based on the observation density within the Inner target domain. The optimum $\hat{\theta}$ (inverse of length-scale) for the fitted GPR over the observed ground motions for both real and imaginary parts of DFT coefficients along both horizontal EW and NS components are demonstrated in Figure 5.8. In Figure 5.8, it is shown that the correlation between the NS component at long periods is higher than those of the EW direction (the blue curve is below the red one at short frequencies).

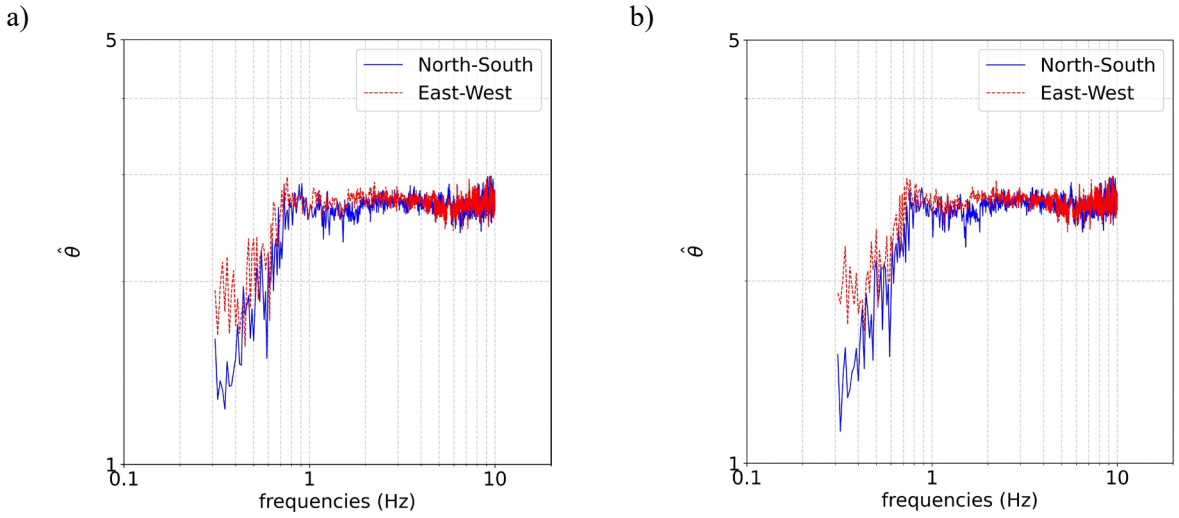


Figure 5.8. The GPR optimum $\hat{\theta}$ for scenario **M**6.7 earthquake's observation set for a) real and b) imaginary parts of the DFT coefficients

The latter might be due to the fault strike, which is almost along the EW direction. It is worth mentioning that the conditioned simulated ground motions are constructed within the usable bandwidth (mutual bandwidth among all observations from CISN and CSN sites).

The simulated ground motions (at SS buildings' locations) RotD50 at T=0.2 s is plotted in Figure 5.9. The T = 0.2 s are chosen as that is close to the majority of the SS buildings' predominant first mode period (in the archetypical structures inventory) (Yi, 2020). Figure 5.9 shows that the closest locations to the epicenter experience severe intensities (RotD50 at T = 0.2 s) as high as 1.6 *g*, and the farther buildings (Northwest side of the chosen testbed) can experience less intensity measures as 0.4 *g* and lower (again for RotD50 at T = 0.2 s).

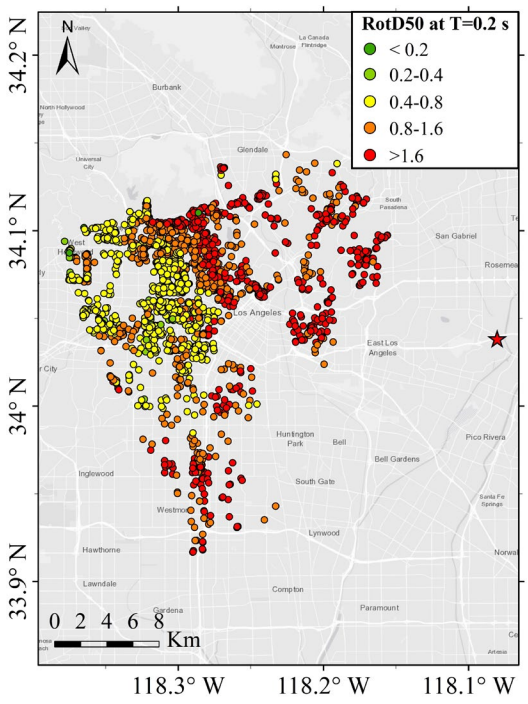


Figure 5.9. The distribution of the RotD50 at T = 0.2 s for the simulated ground motions at SS buildings locations for scenario **M**6.7 earthquake

5.3 Structural Model Assignment

I aim to allocate the "closest" SS structural model to each of the selected SS buildings to be excited with the site-specific simulated ground motions elaborated in the previous section. To do so, the images surrounding each testbed's buildings are extracted and fed to the trained CNN models described in Chapter 4. The algorithm shown in Figure 4.12 is utilized to label the building's first-story wall layout and the number of stories. Figure 5.10 demonstrates the prediction results for the first-story wall layout and number of stories classification for the 1,883 SS buildings within the testbed.

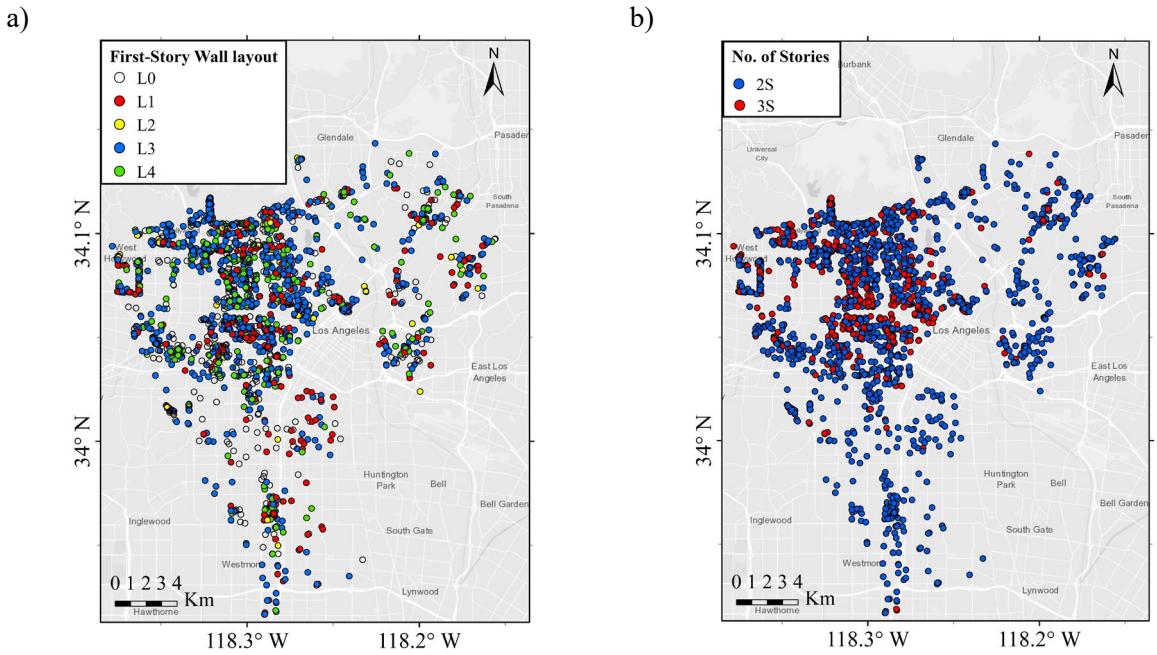


Figure 5.10. The prediction of a) first-story wall layout and b) number of stories for the chosen SS buildings as the testbed in Los Angeles using trained CNN models

In Figure 5.10, 475 buildings are labeled as L0, meaning that their harvested images are not sufficient to classify corresponding buildings as either L1 through L4 classes. In other words, about 25% of the 1,883 testbed SS buildings are classified L0. About 21%, 4%, 56%, and 19% of the non-L0 class buildings are estimated as L1, L2, L3, and L4, respectively. The estimation results of the CNN model for the SS buildings are consistent with the survey results by Burton et al.

(2019) study (17%, 2%, 61%, and 20% for L1, L2, L3, and L4 classes). This verifies that the trained CNN model estimation asserts the distribution of different first-story wall layout types with the in-person surveying of 3,000 SS buildings.

In Figure 5.10, 76% and 24% of the SS buildings were classified as two-story and three-story buildings, respectively. The closest floor plan dimension has been assigned to each non-L0 building using the OSM input features as described in Chapter 4. Both the Gypsum Wall Board and Horizontal Wood Siding material are considered as the material of the buildings, and both will be investigated through nonlinear response history analysis. For the buildings labeled as L0, all four first-story wall layout types with their closest corresponding dimension are used to estimate the structural responses. Section Real-Time Structural Damage State Estimates provides a detailed description of the procedure.

5.3.1 Structural Performance Capacity Quantification

The damage state of the target SS buildings after the earthquake is quantified based on their structural responses. For that purpose, the damage state levels are defined based on the drift limits (e.g., HAZUS, 2015). The pushover response of the structural models is commonly utilized to define the "capacities" for different performance levels of the structures (e.g., Ghobarah, 2004). The drift limit states using pushover analysis are shown by (Yi, 2020) and are employed in this study. For that purpose, each SS structural model in *OpenSees* developed by Burton et al. (2019) (listed in Table 4.1) is analyzed through displacement control pushover along two orthogonal directions, called X and Y. The X direction refers to the longitudinal (longer) dimension of the floor plan, while the Y direction represents the transverse (shorter) one. The detailed distribution of wood-frame shear walls for each category in the first and upper-level stories are adopted from

Yi (2020) and shown in Figure 5.11 and Figure 5.12, respectively. The plots in Figure 5.11 and Figure 5.12 are samples of all buildings in the inventory. In the pushover analysis, the roof's center drift (displacement divided by the height of the building) is obtained concerning the normalized base shear force. In other words, the base shear is divided by the seismic weight of the building. The developed *OpenSees* models include a typical floor dead load of 35 *psf* and 25 *psf* roof dead load. In addition, 10 *psf* and 15 *psf* are considered for the interior and exterior wall's weights per square foot, respectively (Yi, 2020). The seismic weights are calculated considering the dead and live loads.

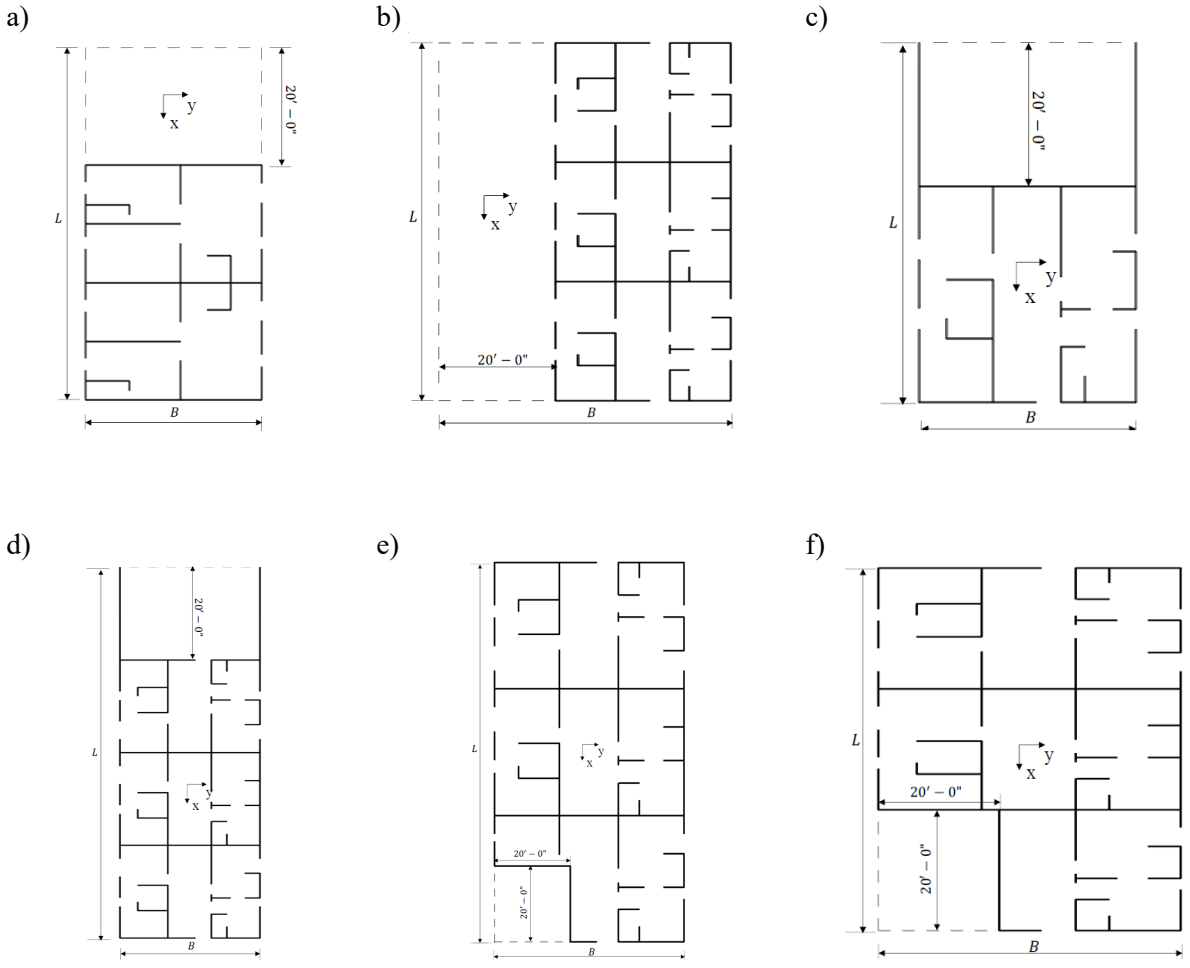


Figure 5.11. The first floor plan including the wood-frame walls for first-story wall layout categories a) L1, b) L2, c) L3 (L=50'), d) L3 (L=80'), e) L4 (L=60'), and f) L4 (L=100') adopted from Yi (2020)

a) b)

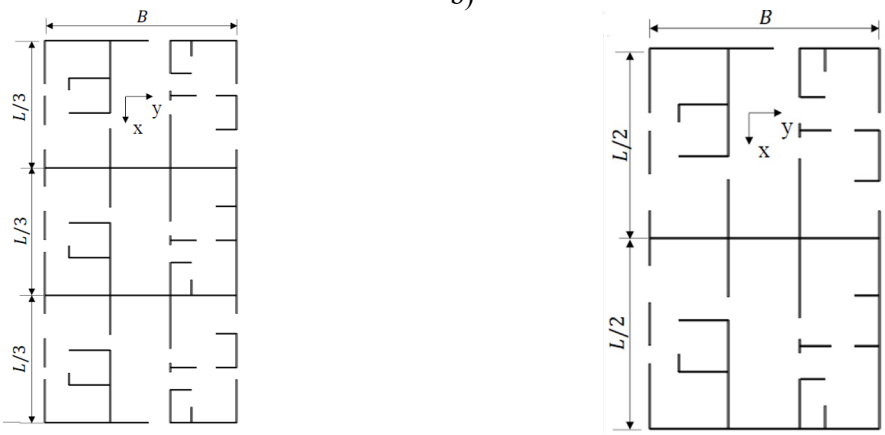


Figure 5.12. The second and third stories' floor plan configuration for the buildings a) L1, L2, L3 (L=80') and L4 and b) L3 (L=50') adopted from Yi (2020)

Figure 5.13 includes the pushover results of the buildings L1-2S-60X30-GWB, L1-2S-60X30-HWS, L2-2S-60X50-GWB, L2-2S-60X50-HWS, L3-2S-50X30-GWB, L3-2S-50X30-HWS, L4-2S-60X50-GWB, and L4-2S-60X50-HWS along both X and Y directions. In Figure 5.13, there are three points shown on each direction's pushover curve. These are the chosen drift limits for the "Serviceability", "Capacity", and "Collapse Initiation" performance states. I denote these roof drift limits with D_e, D_m, and D_p, respectively. The chosen drifts for the serviceability and collapse initiation correspond to the base shear equal to 80% of the strength capacity (ultimate strength). The drift corresponding to the maximum base shear, i.e., the ultimate strength, is picked as the drift associated with the capacity. Then, these drift limits are utilized to decide the damage state of the SS buildings in each direction. Table 5.2 and Table 5.3 summarizes the drift limits corresponding to each performance state for the X and Y directions, respectively.

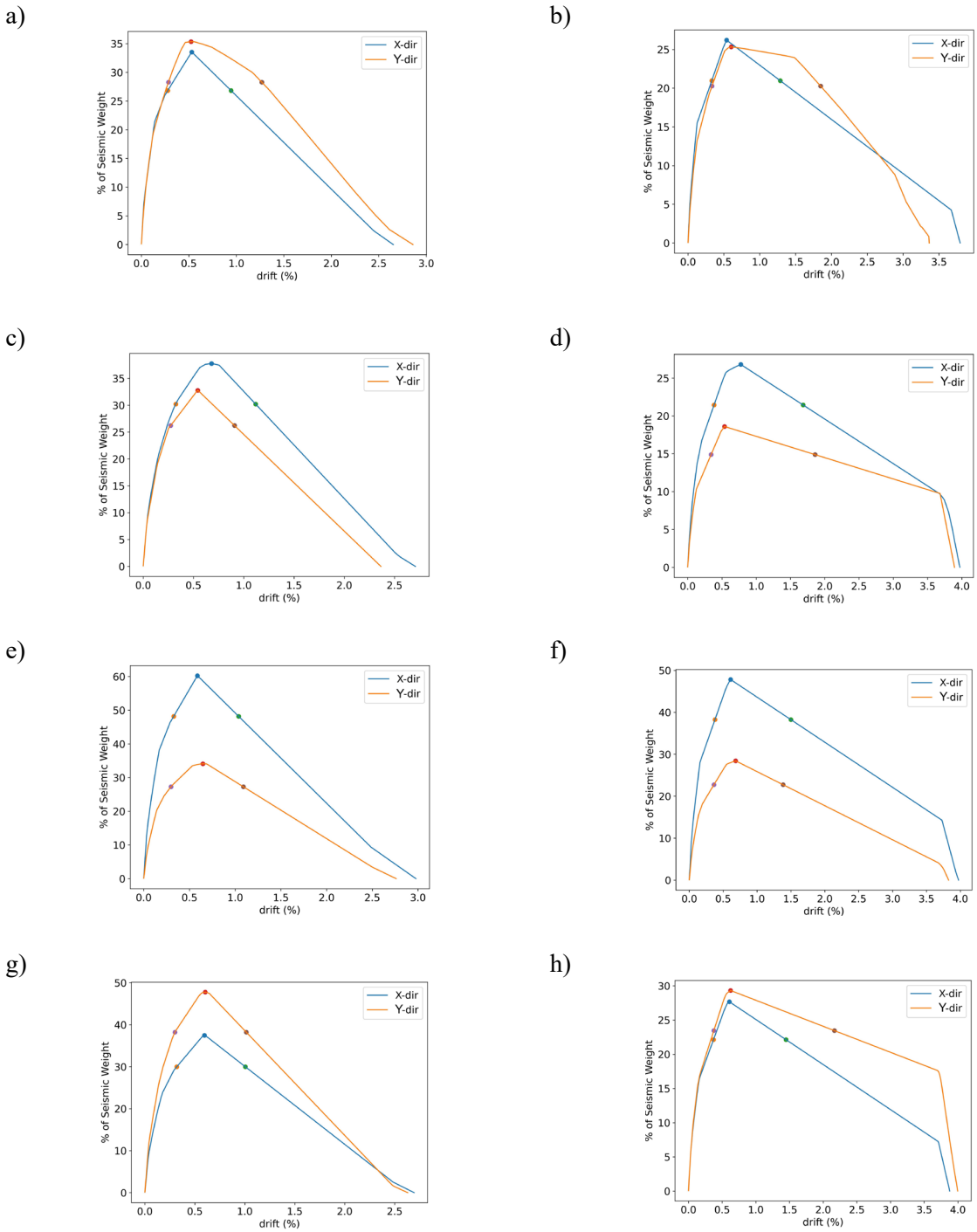


Figure 5.13. The roof drift pushover curves for a) L1-2S-60X30-GWB, b) L1-2S-60X30-HWS, c) L2-2S-60X50-GWB, d) L2-2S-60X50-HWS, e) L3-2S-50X30-GWB, f) L3-2S-50X30-HWS, g) L4-2S-60X50-GWB, and h) L4-2S-60X50-HWS buildings

Table 5.2. Roof drift Limits corresponding to the performance state of the SS buildings along X direction

Building ID	Serviceability Drift	Capacity Drift	Collapse Initiation Drift
L1-2S-60X30-GWB	0.003	0.005	0.009
L1-3S-60X30-GWB	0.002	0.004	0.006
L1-2S-100X30-GWB	0.003	0.005	0.01
L1-3S-100X30-GWB	0.002	0.004	0.007
L1-2S-60X30-HWS	0.003	0.005	0.013
L1-3S-60X30-HWS	0.002	0.004	0.008
L1-2S-100X30-HWS	0.003	0.005	0.014
L1-3S-100X30-HWS	0.003	0.004	0.009
L2-2S-60X50-GWB	0.003	0.007	0.011
L2-3S-60X50-GWB	0.003	0.006	0.008
L2-2S-100X50-GWB	0.003	0.006	0.011
L2-3S-100X50-GWB	0.003	0.005	0.008
L2-2S-60X50-HWS	0.004	0.008	0.017
L2-3S-60X50-HWS	0.003	0.007	0.011
L2-2S-100X50-HWS	0.004	0.007	0.016
L2-3S-100X50-HWS	0.003	0.006	0.011
L3-2S-50X30-GWB	0.003	0.006	0.010
L3-3S-50X30-GWB	0.003	0.005	0.008
L3-2S-80X30-GWB	0.003	0.006	0.010
L3-3S-80X30-GWB	0.004	0.005	0.008
L3-2S-50X30-HWS	0.003	0.006	0.015
L3-3S-50X30-HWS	0.004	0.006	0.011
L3-2S-80X30-HWS	0.003	0.006	0.016
L3-3S-80X30-HWS	0.003	0.006	0.011
L4-2S-60X50-GWB	0.003	0.006	0.010
L4-3S-60X50-GWB	0.003	0.005	0.007
L4-2S-100X50-GWB	0.003	0.006	0.010
L4-3S-100X50-GWB	0.003	0.005	0.007
L4-2S-60X50-HWS	0.004	0.006	0.014
L4-3S-60X50-HWS	0.003	0.006	0.010
L4-2S-100X50-HWS	0.004	0.006	0.015
L4-3S-100X50-HWS	0.003	0.006	0.010

As Table 5.2 and Table 5.3 show, the majority of SS building models have an elastic drift limit, D_e, of around 0.3% in both horizontal directions. In both directions, the average collapse initiation (incipient) drift, D_p, is around 1.1%. It is noted that these building models are for un-retrofitted SS buildings. Retrofitted buildings, which are out of scope of the current study, are expected to have higher capacities.

Table 5.3. Roof drift Limits corresponding to the performance state of the SS buildings along Y direction

Building ID	Serviceability Drift	Capacity Drift	Collapse Initiation Drift
L1-2S-60X30-GWB	0.003	0.005	0.013
L1-3S-60X30-GWB	0.002	0.004	0.008
L1-2S-100X30-GWB	0.003	0.005	0.011
L1-3S-100X30-GWB	0.002	0.004	0.007
L1-2S-60X30-HWS	0.003	0.006	0.018
L1-3S-60X30-HWS	0.003	0.005	0.011
L1-2S-100X30-HWS	0.003	0.009	0.016
L1-3S-100X30-HWS	0.003	0.005	0.010
L2-2S-60X50-GWB	0.003	0.005	0.009
L2-3S-60X50-GWB	0.002	0.004	0.006
L2-2S-100X50-GWB	0.003	0.005	0.009
L2-3S-100X50-GWB	0.002	0.004	0.006
L2-2S-60X50-HWS	0.003	0.005	0.009
L2-3S-60X50-HWS	0.002	0.004	0.006
L2-2S-100X50-HWS	0.003	0.005	0.018
L2-3S-100X50-HWS	0.002	0.004	0.009
L3-2S-50X30-GWB	0.003	0.006	0.014
L3-3S-50X30-GWB	0.002	0.005	0.007
L3-2S-80X30-GWB	0.003	0.007	0.010
L3-3S-80X30-GWB	0.003	0.005	0.008
L3-2S-50X30-HWS	0.004	0.007	0.014
L3-3S-50X30-HWS	0.003	0.006	0.009
L3-2S-80X30-HWS	0.004	0.008	0.023
L3-3S-80X30-HWS	0.003	0.007	0.012
L4-2S-60X50-GWB	0.003	0.006	0.010
L4-3S-60X50-GWB	0.002	0.005	0.007
L4-2S-100X50-GWB	0.003	0.006	0.010
L4-3S-100X50-GWB	0.003	0.005	0.007
L4-2S-60X50-HWS	0.004	0.006	0.021
L4-3S-60X50-HWS	0.003	0.005	0.012
L4-2S-100X50-HWS	0.004	0.006	0.018
L4-3S-100X50-HWS	0.003	0.005	0.010

5.4 Structural Damage State Estimates

The two components required for the framework for performance characterization of chosen SS buildings are established. I determine the damage state of each SS building along the horizontal directions through the following steps:

1. The first-story wall layout category of the target building is obtained from the CNN model's estimation (Figure 5.10a).

2. The number of stories is predicted by employing the CNN model (Figure 5.10b).
3. The building area is obtained from the *OSM*, given the coordinates of the building. Then, the closest floor plan dimension out of the two plan aspect ratio choices is chosen such that the model's area become as close as possible to the actual building's area.
4. The two *OpenSees* models corresponding to GWB and HWS wood-frame wall materials, given the three determined features in steps 1 through 3, are used to conduct the nonlinear response history analysis. The 3D *OpenSees* models are simultaneously excited, employing two horizontal components of ground motion time series. The structural responses, specifically the roof drift in two horizontal directions are recorded.
5. The maximum roof drift during the excitation in each direction is recorded. Then, those maxima are compared to their corresponding drift limits in Table 5.2 and Table 5.3. The damage state of the corresponding direction is considered as slight, moderate, severe, and collapse initiation if the maximum drift in that direction falls below D_e, between D_e and D_m, between D_m and D_p, and beyond D_p, respectively. Eventually, the conservative condition (worst damage state) among two horizontal directions is chosen as the predicted damage state of the SS building.

It is worth mentioning that the input excitations are "mean" generated ground motions at the target sites. A potential future study would be to investigate the uncertainty of the generated damage states using various ground motion realizations.

In Figure 5.10a, there are 475 SS buildings that are labeled as L0. For these buildings, each L1 through L4 model using the corresponding features (number of stories and floor plan dimension) is employed to perform the nonlinear response history analysis through the generated

ground motions. Then, the damage state of each category is determined through the defined procedure above. Finally, the weighted average of damage states using the weights 0.29, 0.06, 0.45, and 0.20 for the categories L1, L2, L3, and L4, respectively, is calculated and rounded up to the closest worst condition. Therefore, all the 1,883 SS buildings' performance considering two types of GWB and HWS wood-frame wall material is determined. Figure 5.14 demonstrates the distribution of the damage states considering GWB or HWS as the wood-frame shear wall materials.

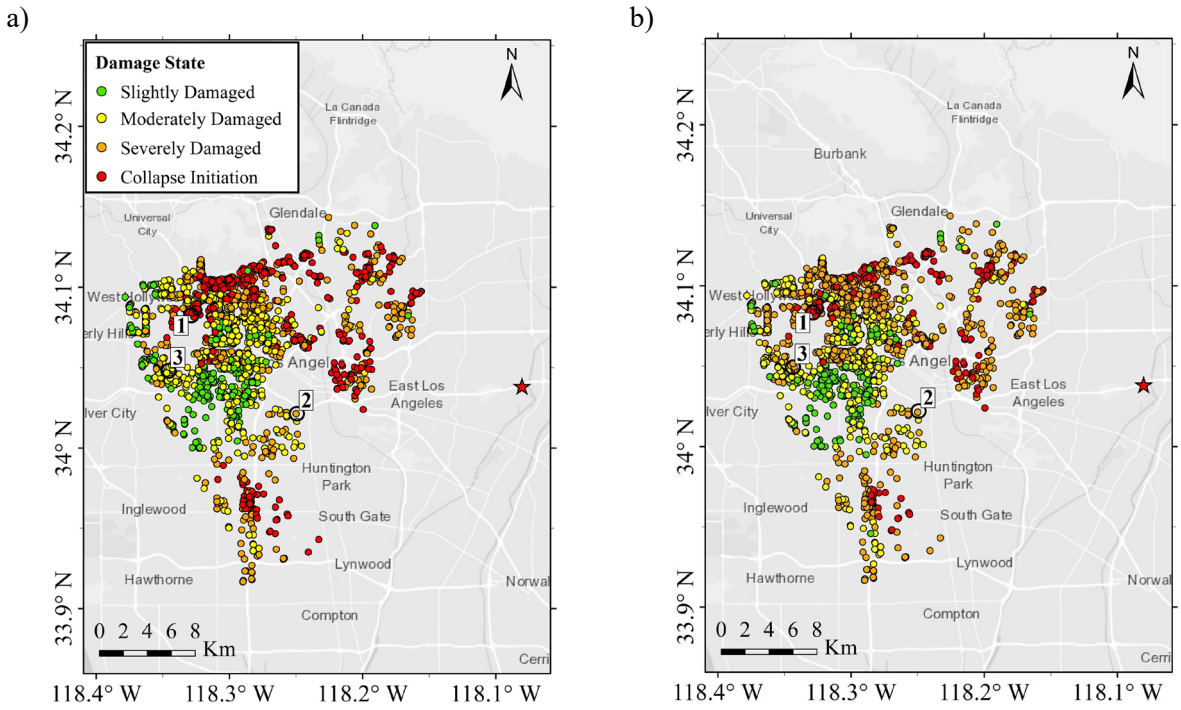


Figure 5.14. The distribution of the estimated damage states for the testbed SS buildings with a) Gypsum Wall Board and b) Horizontal Wood Siding wall material for scenario **M**6.7 earthquake

Tables 5.4 and 5.5 summarize the damage state of each building category for GWB and HWS considered wall materials, respectively. In Tables 5.4 and 5.5, the values within parentheses in front of the percentages demonstrate the number of buildings that falls in that partition. The percentage in each cell shows the portion of buildings inside each layout category.

Table 5.4. Damage state distribution among various categories with Gypsum Wall Board material

	Damage State			
	Slightly Damaged	Moderately Damaged	Severely Damaged	Collapse Initiation
L0	17.9%(85)	33.7%(160)	23.15%(110)	25.3%(120)
L1	9.3%(27)	29.3%(85)	27.6%(80)	33.8%(98)
L2	8.9%(5)	19.6%(11)	32.2%(18)	39.3%(22)
L3	17.4%(138)	33.7%(268)	25.3%(201)	23.6%(187)
L4	17.2%(46)	34.3%(92)	16.4%(44)	32.1%(86)
Total	**16%(301)**	**32.7%(616)**	**24%(453)**	**27.3%(513)**

Table 5.5. Damage state distribution among various categories with Horizontal Wood Siding material

	Damage State			
	Slightly Damaged	Moderately Damaged	Severely Damaged	Collapse Initiation
L0	15.6%(74)	30.7%(146)	42.7%(203)	11%(52)
L1	8.9%(26)	36.5%(106)	32.1%(93)	22.5%(65)
L2	5.3%(3)	12.5%(7)	58.9%(33)	23.3%(13)
L3	14.4%(114)	33.6%(267)	42.1%(334)	9.9%(79)
L4	17.1%(46)	23.5%(63)	43.6%(117)	15.7%(42)
Total	**14%(263)**	**31.3%(589)**	**41.4%(780)**	**13.3%(251)**

Tables 5.4 and 5.5 show that generally about 53% of the SS buildings experienced either severely damaged or collapse initiation damage states after the scenario **M**6.7 local earthquake (average of GWB and HWS materials). In addition, it is concluded that just around 15% of the SS buildings experienced slight damage after the earthquake and, thus, approximately 85% of SS buildings experienced considerable damage.

It is shown from both Tables 5.4 and 5.5 that L2 has the worst performance as it has the lowest percentage of slightly damaged buildings and the highest percentage of collapse initiations. On the other hand, it is shown that L3 had the best performance as it has a generally higher percentage of slightly damaged and lowest percentage of the collapse initiation damage states. Moreover, comparing the results from Table 5.5 with those of Table 5.4 reveals that the performance of the buildings with HWS wall material is generally better in collapse prevention, as it is seen there are 27% versus 13% of collapses for buildings with GWB and HWS wall

materials, respectively. In other words, the buildings with HWS materials experienced fewer collapses but more severe damages compared to those with GWB materials. In Figure 5.14, there are three chosen test buildings. Their roof drift time history for both materials is shown in Figure 5.15. In Figure 5.15, the response history analysis of the central roof of the buildings is shown for the direction (either X or Y) that the damage state category of the building is chosen from that. In other words, the worst-case direction in terms of the damage state is picked to display the time history. It is clearly shown in Figure 5.15 that the buildings experiencing nonlinear behavior by entering beyond the elastic drift limit show a permanent residual drift following the end of the input excitations. It is worth mentioning that the computed roof drift time series includes the free vibration of the structure at the end of the excitation.

In summary, I demonstrate the whole framework's application for a large set of SS buildings in Los Angeles. The procedure including the ground motion time series simulation and nonlinear response history analyses of the SS buildings, is conducted on the University of California Los Angeles computational and storage services associated with the Hoffman2 Shared Cluster for Digital Research and Education's Research Technology Group. Using such computational power makes the whole framework rapid, as the nonlinear dynamic analyses process for about 2,000 SS buildings took about 30 minutes to be completed using 200 CPU cores.

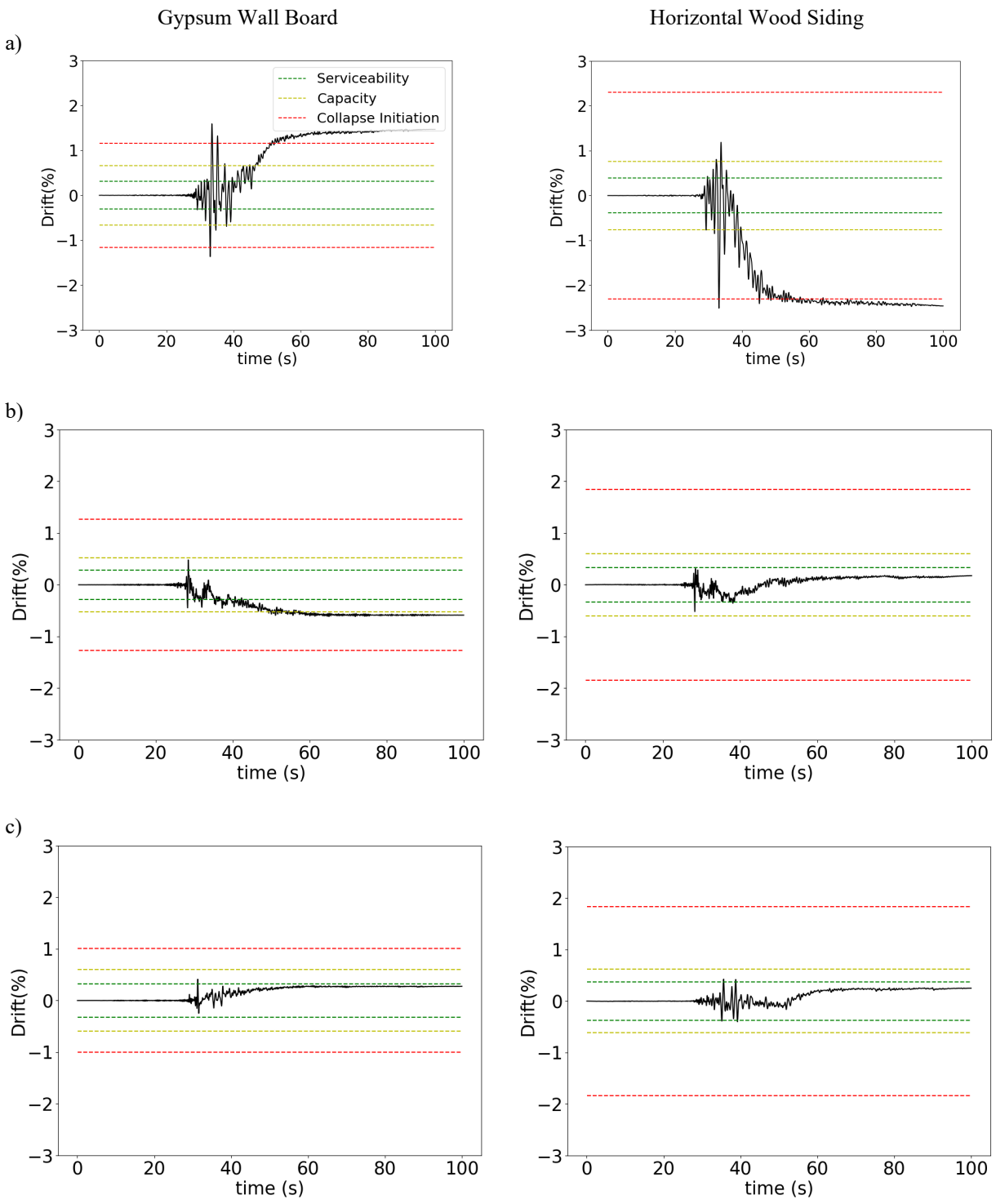


Figure 5.15. The central roof drift time history for the test buildings a) 1, b) 2, and c) 3 shown in Figure 5.14

CHAPTER 6. Summary, Conclusions, and Future Work

6.1 Overview of the Research

The traditional post-earthquake damage assessment of structures is a labor-intensive and time-consuming process. Thus, many organizations, including government agencies, can benefit from a near-real-time seismic performance assessment. As the price of instrumentations becomes cheaper by passing the time, the future number of recording instruments will be much higher than those of today and cities will become densely instrumented, i.e., gradually becoming "smart" cities where highly dense granular recording instruments are installed outside and inside of structures. For mega-cities, such as Los Angeles, becoming a smart city is more challenging even though the number of earthquake instruments is exponentially increasing. Therefore, following an earthquake, and given a set of densely distributed ground-motion instruments, there is a need to generate ground motions at un-instrumented sites to quantify site-specific near-real-time seismic performance of structures.

The main objective of this study is to provide the ingredients required for establishing a framework which can be utilized for the post-earthquake near-real-time seismic performance assessment of a large portfolio of structures, in this study for soft-story (SS) buildings. More precisely, the issues addressed in this study are listed below.

- Providing a methodology and model to rapidly generate the ground motion time series at the locations of interest where there are no recording instruments (Chapter 2).
- Evaluate the generated ground motions' reliability, uncertainty, and validity through the introduced model for various conditions (Chapter 3).

- Presenting an automated methodology using image recognition techniques to establish the “closest” structural computer models for the existing SS buildings (Chapter 4).
- Evaluate the application of established ingredients as a framework to estimate the damage states of the SS buildings for a major earthquake in Los Angeles (Chapter 5).

6.2 Summary of the Research

6.2.1 Chapter 2: Conditioned Simulation of Ground Motion Time Series

This Chapter mainly focused on the development of a methodology to construct the ground motion time series at un-instrumented sites. A novel Gaussian Process Regression (GPR) model that can efficiently generate time series, utilizing the surrounding recorded ground motions, was presented. The optimum parameters of the model at each frequency were computed based on the given observed (training) sets of recorded motions. The hyperparameter of the model was recognized to be sensitive to the instrumentation density at the target region. Thus, the hyperparameter of the model was fine-tuned for the 2019 **M**7.1 Ridgecrest earthquake recorded in Southern California. The key set of data used in this study were the recorded motions at the Community Seismic Network (CSN) which is a highly dense network in Southern California. In addition, the performance of the GPR model was evaluated for two physics-based earthquakes simulated for Northern California as well as an actual recorded earthquake in Southern California. The performance of the GPR model is fairly reliable for the predictions at long-period contents.

6.2.2 Chapter 3: Uncertainty Quantification and Sensitivity Analysis of Simulated Ground Motions

The generated ground motion time series in Chapter 2 had their own uncertainty. Chapter 3 presented a methodology to generate multiple realizations of the ground motion time series at

the target sites. The inter-frequency correlation of the realizations of the motions was constructed through the BA19 (Bayless and Abrahamson, 2019a) model. The uncertainty of the generated ground motions was then quantified. The results illustrated that the instrumentation density closer to the target site is critical to generate reliable and accurate frequency contents of the motions, especially at long periods. Also, it was concluded that close-by instrumentation could not necessarily decrease the uncertainty of the generated motions at short periods. Moreover, the combination of various seismic networks' recorded motions is used to estimate the ground motion time series for **M**7.1 Ridgecrest and the 2020 **M**4.5 South El Monte earthquakes. The results demonstrated that the error and uncertainty of the generated motions (especially at long periods) decrease with additional observations from multiple seismic networks.

6.2.3 Chapter 4: Rapid Structural Model Classification for Soft-Story Wood-Frame Buildings

Chapter 4 aimed to develop an automated methodology for allocating the "closest" computer structural model for the target SS building. This Chapter presented an automated methodology for Google Street View (GSV) image harvesting of target SS buildings. Then, these images were used to classify the two key features required to allocate the "closest" computer structural model utilizing image recognition methodologies, Convolutional Neural Networks (CNNs) models. The finite element computer structural models developed by Burton et al. (2019) to represent the existing SS buildings were employed for this purpose. The evaluation of the trained CNN models, in order to classify first-story wall layout and the number of stories, illustrated 80.1% and 91.4% accuracies, respectively. In addition, a methodology to select the floor plan dimensions among the computer structural models by matching the area between the actual building and the computer structural model is presented. For that purpose, the building's footprint coordinates were

obtained from OpenStreetMap (OSM). Ultimately, an algorithm to label the buildings using multiple images harvested from their perimeter was introduced. The performance of the methodology was compared with the trained human's estimation accuracy. The results showed that the trained CNN models' performance was close to those of humans.

6.2.4 Chapter 5: Framework for Near-Real-Time Soft Story Structural Damage Assessment

Chapter 5 demonstrated a testbed including approximately 2,000 SS buildings in Los Angeles. These buildings were used to apply the whole framework for assessing their damage states for a scenario **M**6.7 local earthquake in Los Angeles. The **M**6.7 earthquake was generated by scaling up the recorded motions of the 2020 **M**4.5 South El Monte earthquake recorded at the CSN and California Integrated Seismic Network (CISN) sites. Then, the ground motion time series at each specific SS building location was constructed using the GPR model, while the buildings' "closest" computer structural models were identified utilizing the trained CNN models. Eventually, comprehensive nonlinear response history analyses for about 2,000 testbed SS buildings were performed. The results demonstrated that Horizontal Wood Siding (HWS) material performed better than Gypsum Wall Board (GWB) ones in preventing the buildings from collapse initiation damage state. Also, approximately 85% of the SS buildings experienced considerable damages (moderately, severely damaged, or collapse initiation), whereas only about 15% of the SS buildings were in a slightly damaged state. A damage state distribution map for the 2,000 targets SS buildings were generated for both types of materials, which took approximately 30 minutes using the Hoffman2 shared Cluster of the University of California, Los Angeles.

In summary, the framework presented in this research can be implemented in post-earthquake near-real-time damage assessment of a large portfolio of structures.

The framework's components established in this study, e.g., the ground motion simulation methodology, can later be utilized by government organizations to generate the shaking time series at any desired locations in a near-real-time manner and more accurately as the number of instrumentations increases. In addition, the image recognition methodologies presented in this study can be utilized to detect any further geometrical features, e.g., accurately detect the building's height or plan's shape and dimensions using both street and aerial images, to develop computer structural models more rapidly.

6.3 Future Work

This study mainly provided the two ingredients, exciting ground motions and "closest" computer structural model, of the framework for near-real-time damage state assessment of structures. There are several potential future studies for each of these components, which will be elaborated next.

The GPR model, which is utilized to construct the ground motion time series, employed four-component input vectors for the sites; three geographical coordinates and local site condition. These input vectors considered Vs_{30} to represent the local site condition. In addition, the covariance kernel function deployed in the presented GPR model was isotropic, meaning that all elements of the input feature vector were normalized with the same length-scale, which might result in erroneous ground motions' generation at the epicentral region. Thus, a potential future study can consider extending the input feature vectors, e.g., $Z_{1.0}$, $Z_{2.5}$, or R_{rup}, as well as the combination of covariance functions utilizing anisotropic kernels to tackle the issue of ground motions generated closer to the epicenter. In addition, the GPR model's performance can be further evaluated by

comparing the structural responses excited by conditioned simulated ground motions with the one excited by actual recorded ground motions.

The uncertainty of the generated ground motions was quantified using the 2019 **M**7.1 Ridgecrest earthquake. However, more recorded ground motions by various earthquakes in the regions with a considerably high instrumentation density are required to extend the current uncertainty quantification. Such a potential future study can present an established model to predict the uncertainty of the estimated ground motions. As the number of recording instruments increases by passing the time, future recorded ground motions can be utilized to extend the current uncertainty quantifications.

The trained CNN models for classifying the first-story wall layout and number of stories of the SS buildings were trained using around 5K obtained images. Such training dataset can be extended by employing semi-supervised learning methodologies where the semi-trained CNN model can itself be utilized to annotate and add more labeled images to the current existing training sets. Thus, the accuracy of the current models could increase. Moreover, object detection methodologies that can identify the target SS building within a given image and outcrop it to provide more informative input to the CNN models can be utilized. Such an addition can improve the performance of the current CNN models by removing the noises (useless objects) within the images. In addition, tax registry or ZTRAX datasets can be deployed as additional sets to be incorporated for detecting the wood-frame shear wall materials for choosing the "closest" computer structural models. Another supervised learning model might also be trained to estimate the wall material using the buildings' features harvested from the aforementioned datasets.

The presented framework can be extended utilizing automated methodologies to produce more accurate near-real-time *OpenSees* scripts for generating computer structural models. For

example, the footprint of the buildings can be obtained to create precise floor plan dimensions in computer structural models. In addition, other types of residential buildings, such as cripple wall residentials, can be utilized to broaden the map of damage states for a city (currently for SS buildings). As mentioned, the number of instrumentations increases, and such a methodology can be studied for other cities as well as other types of structures. In an ultimate manner, the resolution of the estimated damage states can be increased to the building-by-building level.

The generated damage state distribution maps considered the mean estimated ground motions as well as one most probable representing computer structural model. Such maps have uncertainty intrinsically in the estimated damage states. Thus, a potential study can utilize multiple nonlinear response history analyses by utilizing realizations of ground motions as well as various computer structural models based on the Convolutional Neural Network models' output probabilities. Such a process can produce a map of uncertainty for each building's predicted performance state. Also, further investigation in the local regions of the map can be conducted to study the local statistics of the structural damage distribution with respect to the epicentral distance, buildings' category, and the intensity measures.

The efficiency of the computational algorithms developed in this study can be increased for near-real-time post-earthquake implementation for a large set of structural systems in a large urban area. Such a framework also can be employed to estimate the financial loss of the large portfolio in a near-real-time manner.

Finally, although a large portfolio of soft-story buildings in Los Angeles was used as a testbed of the methodology, the framework developed here can be implemented for other types of structures in other cities.

Appendix A. Predicted Motions for CISN stations for M7.1 Ridgecrest and M4.5 South El Monte earthquakes

The estimated ground motion results for 2019 **M**7.1 Ridgecrest earthquake dataset at CISN sites using 1) just all CISN sites and 2) CSN-plus- CISN sites as observation are shown in Figure A2.

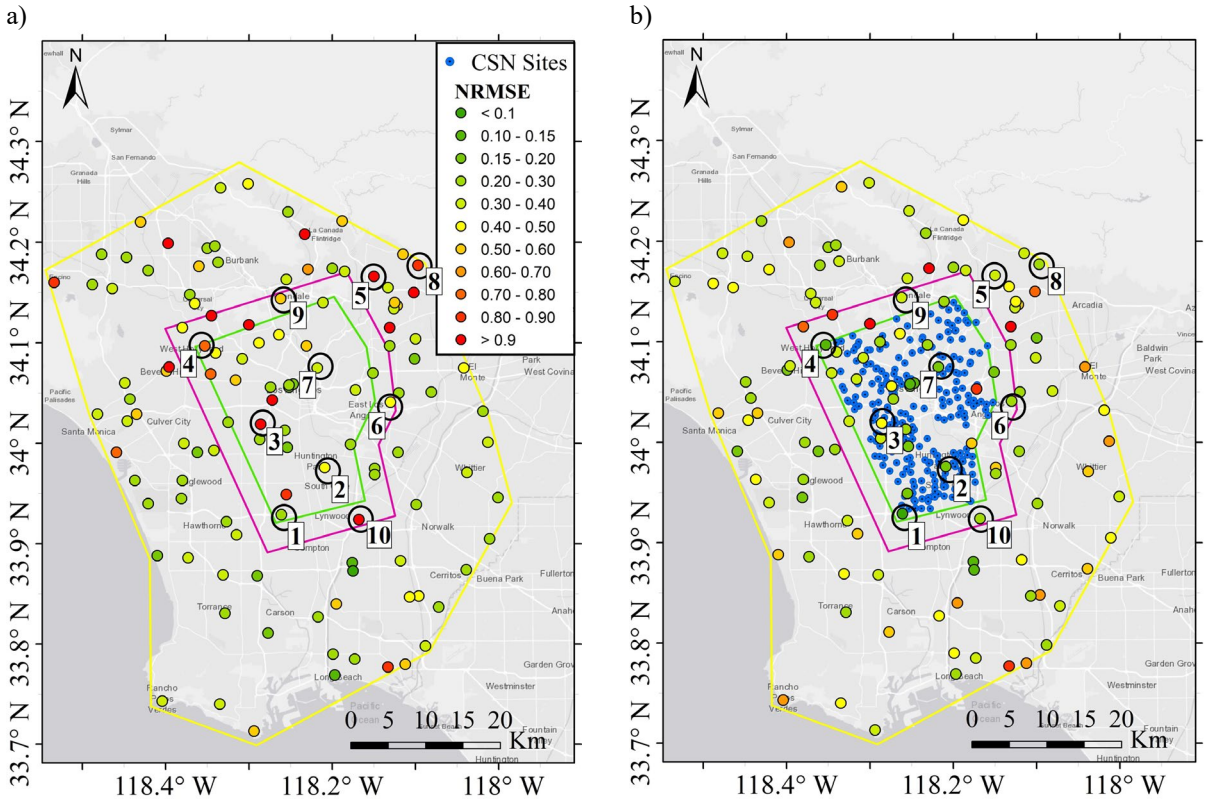
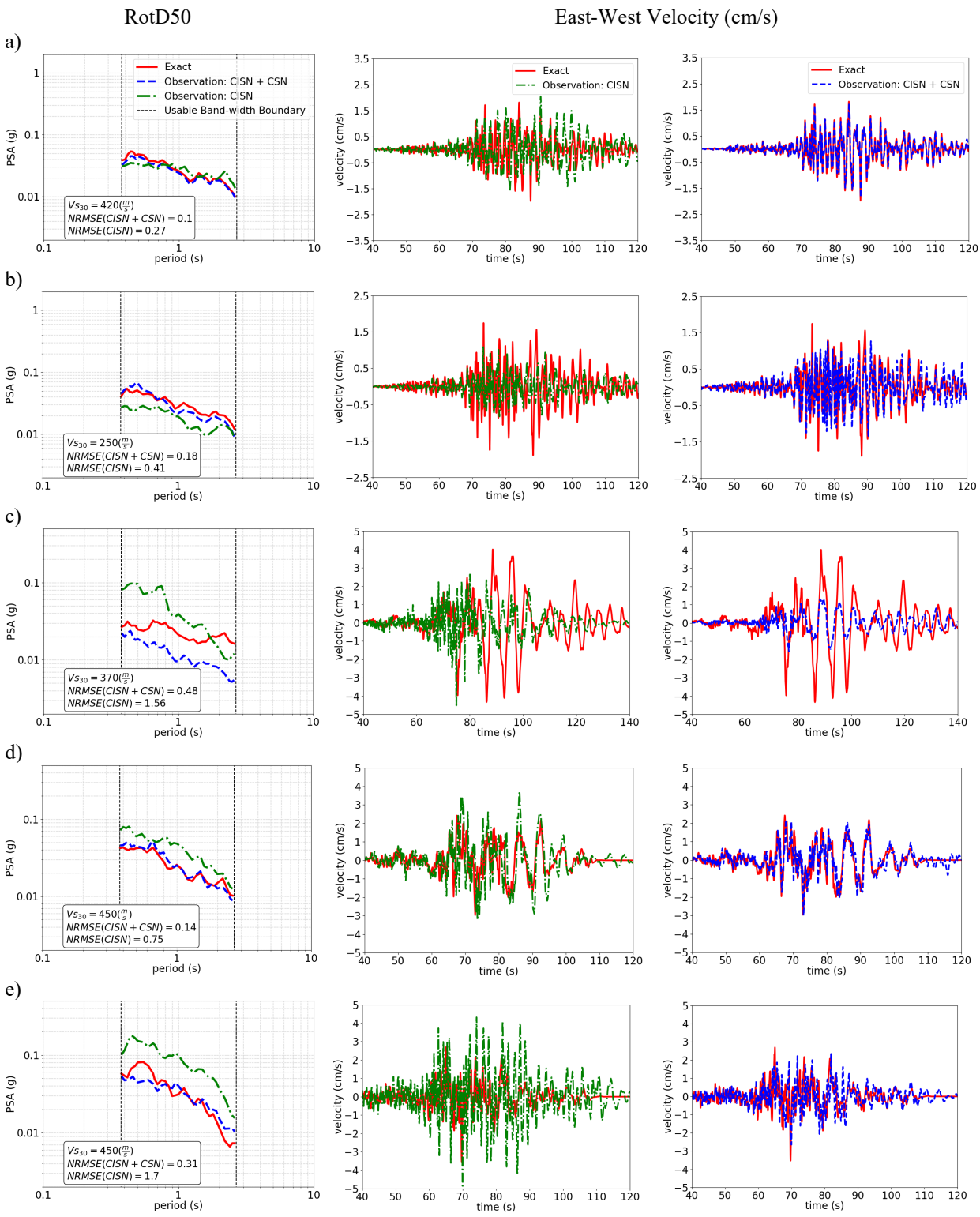


Figure A1. Distribution of RotD50 NRMSE for having a) just CISN sites and b) all CISN and CSN sites as observations and ten chosen CISN test sites within the M7.1 Ridgecrest earthquake dataset


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

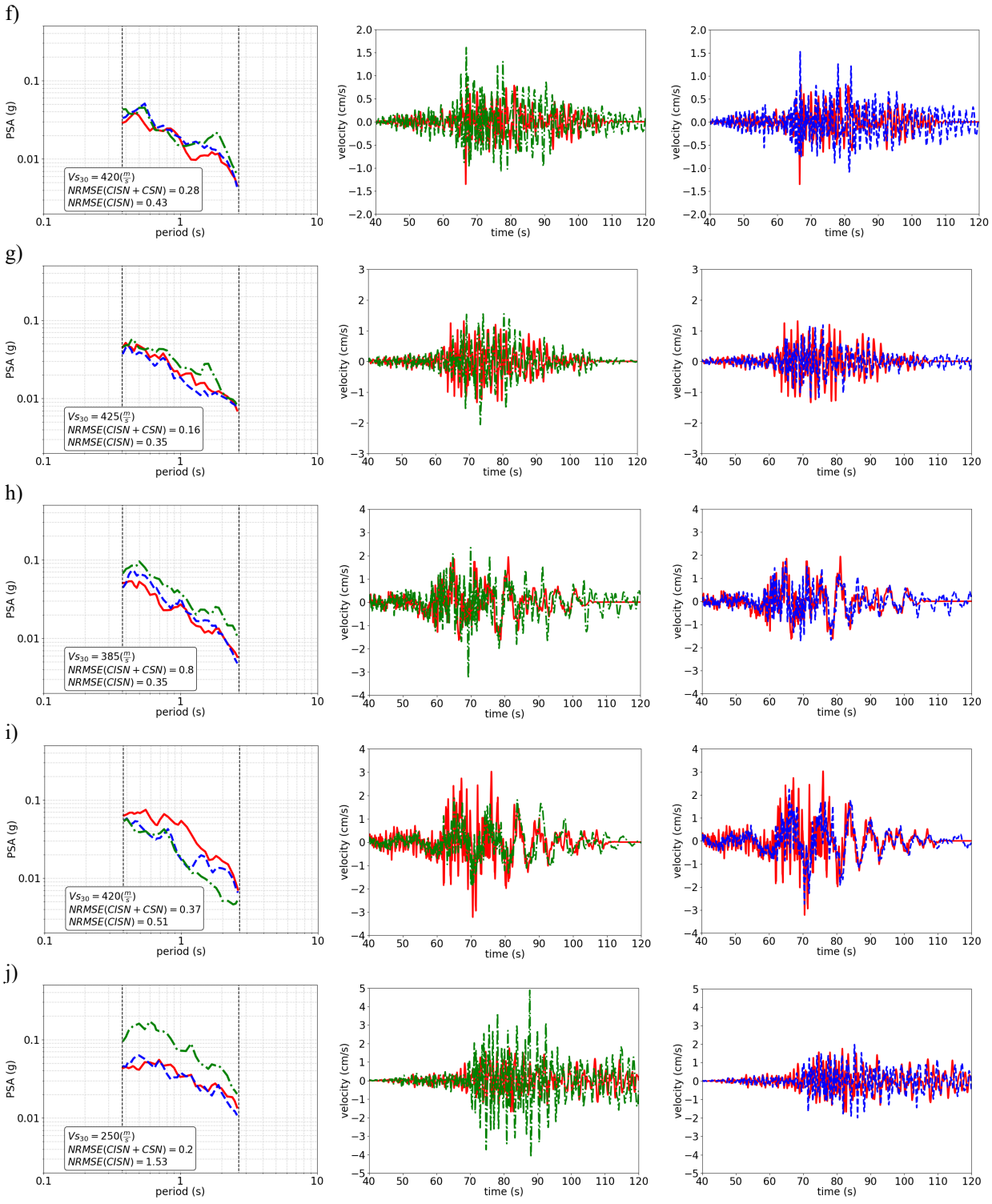


Figure A2. The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation and exact motions along East-West direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, e) No. 5, f) No. 6, g) No. 7, h) No. 8, i) No. 9, and j) No. 10 within CISN network for M7.1 Ridgecrest earthquake

The estimated ground motion results for 2020 **M**4.5 South El Monte earthquake dataset at CISN sites using 1) just all CISN sites and 2) CSN-plus- CISN sites as observation are shown in Figure A4.

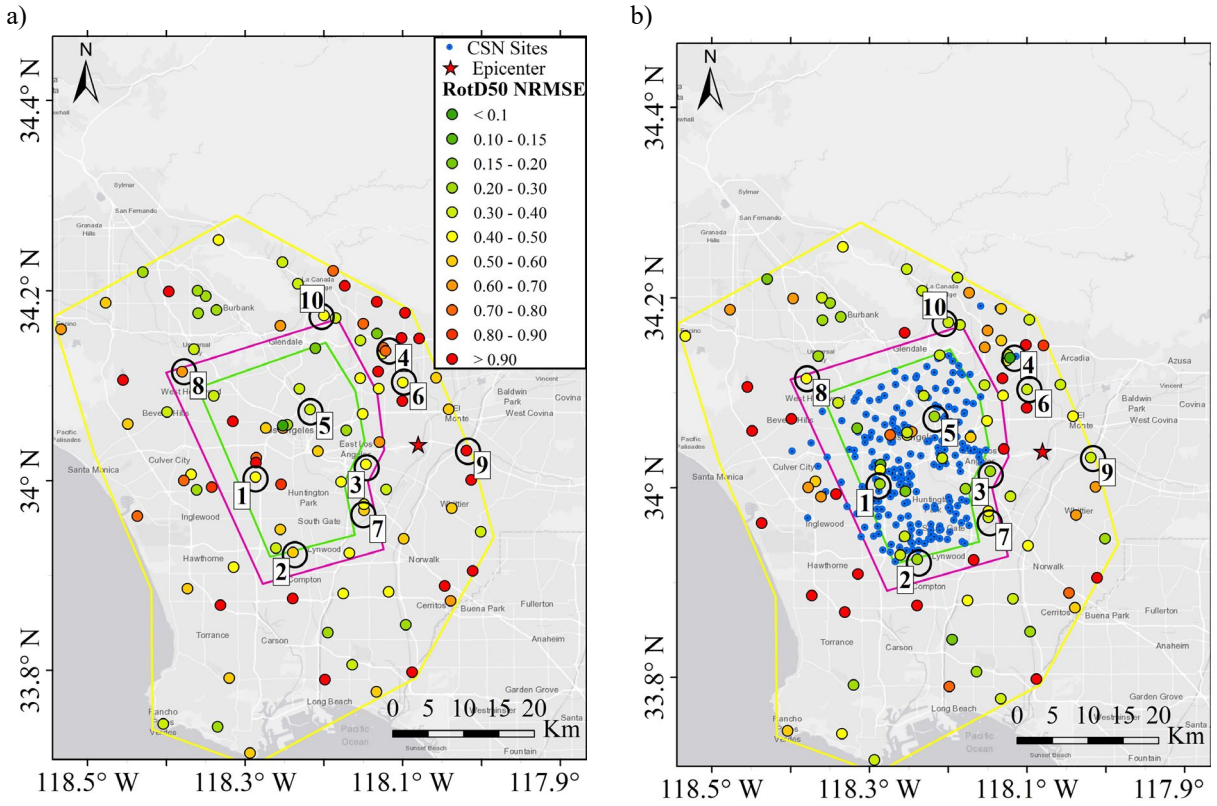
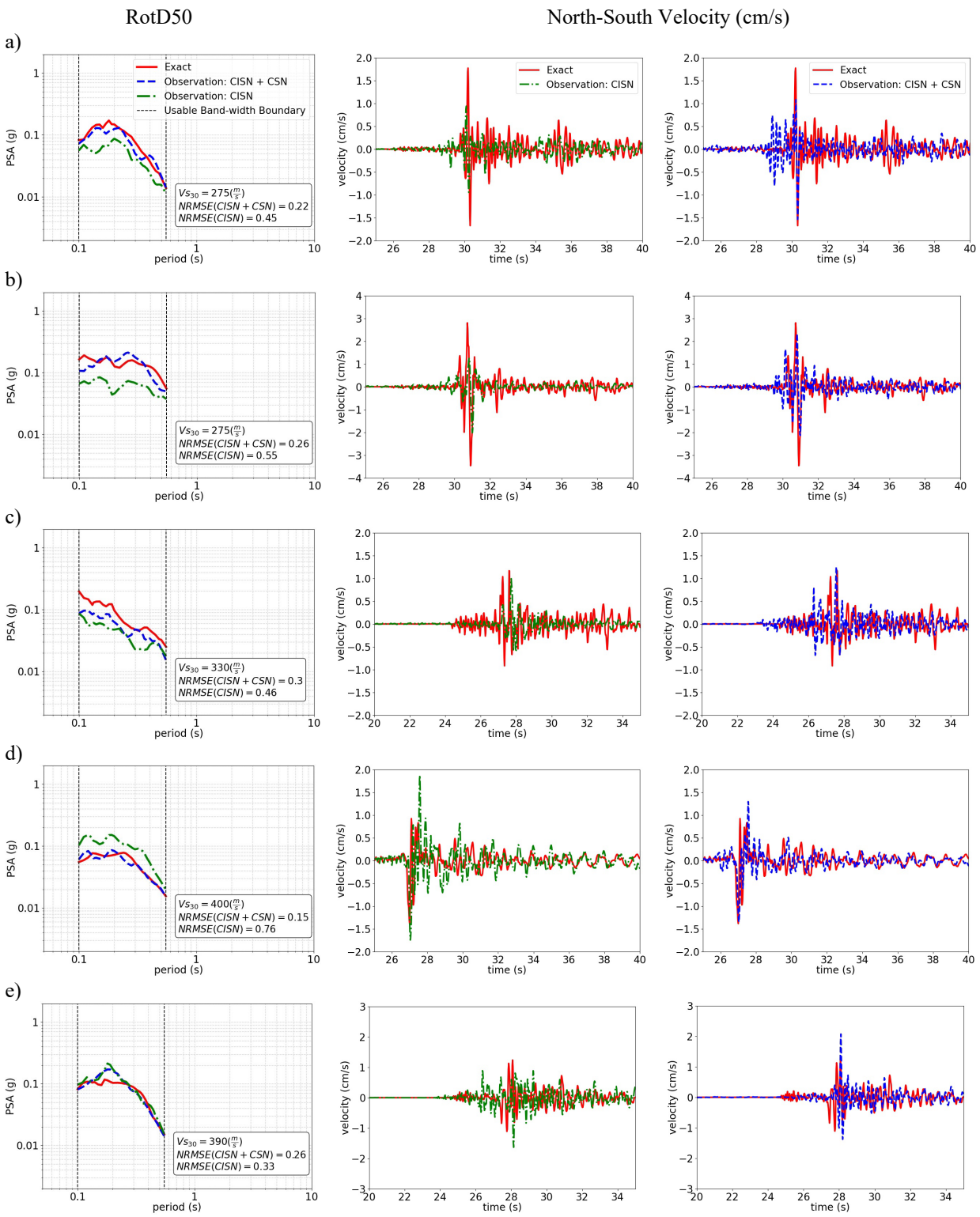


Figure A3. Distribution of RotD50 NRMSE for having a) just CISN sites and b) all CISN and CSN sites as observations and ten chosen CISN test sites within the M4.5 South El Monte earthquake dataset


RotD50
North-South Velocity (cm/s)
a)
b)
c)
d)
e)
Exact
Observation: CISN + CSN
Observation: CISN
Usable Band-width Boundary
$Vs_{30} = 275(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.22$
$NRMSE(CISN) = 0.45$
$Vs_{30} = 275(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.26$
$NRMSE(CISN) = 0.55$
$Vs_{30} = 330(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.3$
$NRMSE(CISN) = 0.46$
$Vs_{30} = 400(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.15$
$NRMSE(CISN) = 0.76$
$Vs_{30} = 390(\frac{m}{s})$
$NRMSE(CISN + CSN) = 0.26$
$NRMSE(CISN) = 0.33$
PSA (g)
period (s)
velocity (cm/s)
time (s)

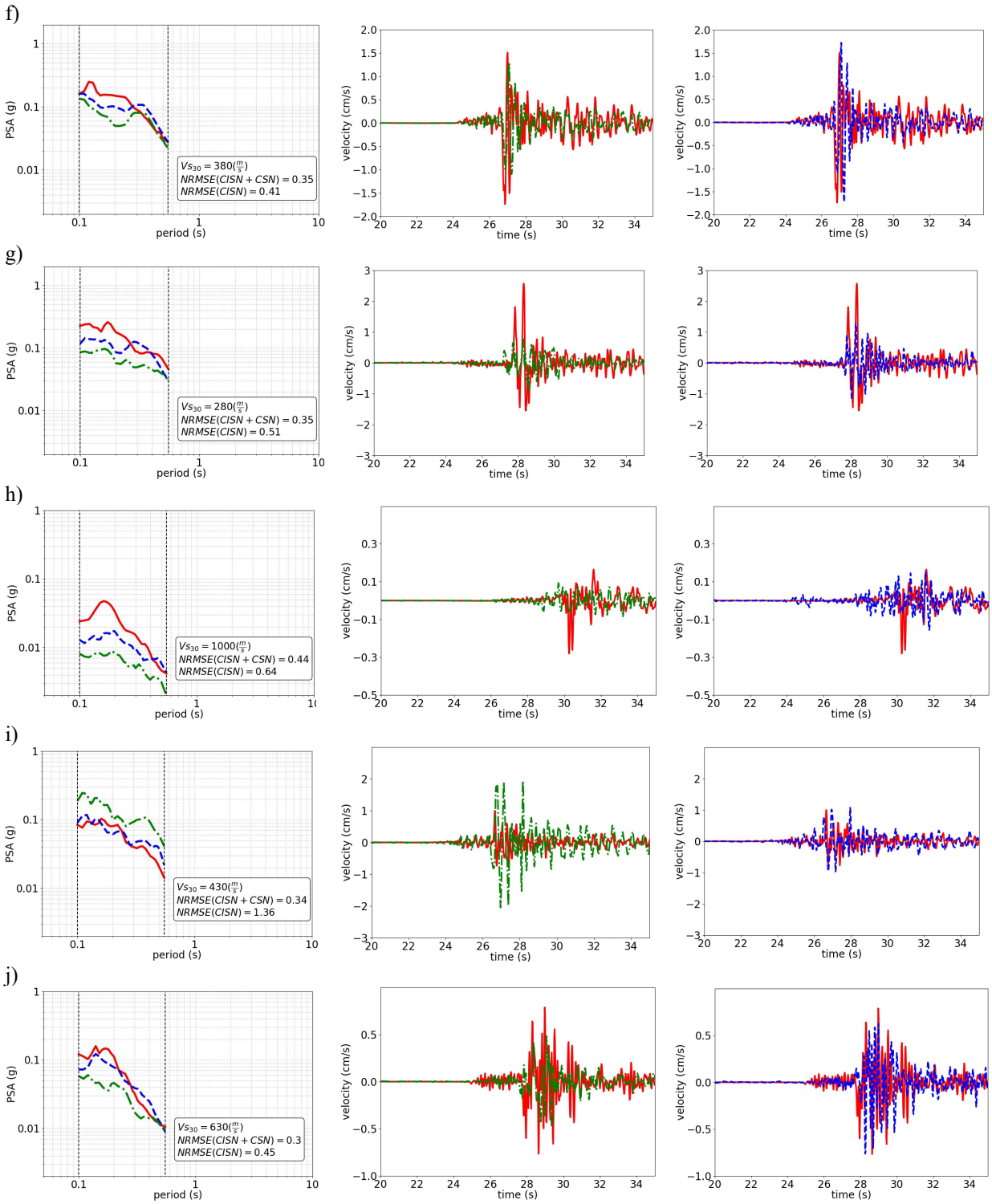


Figure A4. The RotD50 and velocity time series of the predicted using just CISN and CISN plus CSN observation and exact motions along North-South direction for the chosen test sites a) No. 1, b) No. 2, c) No. 3, d) No. 4, e) No. 5, f) No. 6, g) No. 7, h) No. 8, i) No. 9, and j) No. 10 within CISN network for M4.5 South El Monte earthquake

REFERENCES

Aagaard, B.T., Brocher, T.M., Dolenc, D., Dreger, D., Graves, R.W., Harmsen, S., Hartzell, S., Larsen, S., McCandless, K., Nilsson, S. and Petersson, N.A., 2008a. Ground-Motion Modeling of the 1906 San Francisco Earthquake, Part II: Ground-Motion Estimates for the 1906 Earthquake and Scenario EventsGround-Motion Modeling of the 1906 San Francisco Earthquake, Part II. Bulletin of the Seismological Society of America, 98(2), pp.1012-1046.

Aagaard, B.T., Brocher, T.M., Dolenc, D., Dreger, D., Graves, R.W., Harmsen, S., Hartzell, S., Larsen, S. and Zoback, M.L., 2008. 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.

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.

Abrahamson, N.A., Silva, W.J. and Kamai, R., 2014. Summary of the ASK14 ground motion relation for active crustal regions. Earthquake Spectra, 30(3), pp.1025-1055.

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.

Alipour, A., M. and Harris, D.K., 2020. A big data analytics strategy for scalable urban

infrastructure condition assessment using semi-supervised multi-transform self-training. Journal of Civil Structural Health Monitoring, 10(2), pp.313-332.

Ancheta T., Robert, B.D., Jonathan, P.S., Emel, S., Walter, J.S., Brian, S., Chiou, J., Wooddell, K.E., Graves, R.W., Kottke, A.R. and Boore, D.M., 2014. NGA-West2 database. Earthquake Spectra, 30(3), pp.989-1005.

ATC, A., 1995. 20-2: Addendum to the ATC-20 Post earthquake Building Safety Evaluation Procedures. Applied Technology Council, San Francisco.

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., 2019a. An Empirical Model for the Interfrequency Correlation of Epsilon for Fourier Amplitude SpectraAn Empirical Model for the Interfrequency Correlation of Epsilon for Fourier Amplitude Spectra. Bulletin of the Seismological Society of America, 109(3), pp.1058-1070.

Bayless, J. and Abrahamson, N.A., 2019b. Summary of the BA18 Ground-Motion Model for Fourier Amplitude Spectra for Crustal Earthquakes in California. Bulletin of the Seismological Society of America, 109(5), pp.2088-2105.

Bency, A.J., Rallapalli, S., Ganti, R.K., Srivatsa, M. and Manjunath, B.S., 2017, March. Beyond spatial auto-regressive models: Predicting housing prices with satellite imagery. In 2017 IEEE winter conference on applications of computer vision (WACV) (pp. 320-329). IEEE.

Boatwright, J., Blair, J.L., Aagaard, B.T. and Wallis, K., 2015. The distribution of red and yellow

tags in the City of Napa. Seismological Research Letters, 86(2A), pp.361-368.

Boore, D.M., Stewart, J.P., Seyhan, E. and Atkinson, G.M., 2014. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. Earthquake Spectra, 30(3), pp.1057-1085.

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.

Bozorgnia, Y., Abrahamson, N.A., Atik, L.A., Ancheta, T.D., Atkinson, G.M., Baker, J.W., Baltay, A., Boore, D.M., Campbell, K.W., Chiou, B.S.J. and Darragh, R., 2014. NGA-West2 research project. Earthquake Spectra, 30(3), pp.973-987.

Burton, H., Rad, A.R., Yi, Z., Gutierrez, D. and Ojuri, K., 2019. Seismic collapse performance of Los Angeles soft, weak, and open-front wall line woodframe structures retrofitted using different procedures. Bulletin of earthquake engineering, 17(4), pp.2059-2091.

Campbell, A., Both, A. and Sun, Q.C., 2019. Detecting and mapping traffic signs from Google Street View images using deep learning and GIS. Computers, Environment and Urban Systems, 77, p.101350.

Campbell, K.W. and Bozorgnia, Y., 2014. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. Earthquake Spectra, 30(3), pp.1087-1115.

Chen, P.Y., Wu, Z.Y. and Taciroglu, E., 2021. Classification of Soft-Story Buildings Using Deep Learning with Density Features Extracted from 3D Point Clouds. Journal of Computing in Civil Engineering, 35(3), p.04021005.

Chen, Y. and Baker, J.W., 2019. Spatial Correlations in CyberShake Physics-Based Ground-

Motion SimulationsSpatial Correlations in CyberShake Physics-Based Ground-Motion Simulations. Bulletin of the Seismological Society of America, 109(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.

Earle, P.S., Wald, D.J., Jaiswal, K.S., Allen, T.I., Hearne, M.G., Marano, K.D., Hotovec, A.J. and Fee, J.M., 2009. Prompt Assessment of Global Earthquakes for Response (PAGER): A system for rapidly determining the impact of earthquakes worldwide. US Geological Survey Open-File Report, 1131(2009), p.15.

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.

Fei-Fei, L., Deng, J. and Li, K., 2009. ImageNet: Constructing a large-scale image database. Journal of vision, 9(8), pp.1037-1037.

FEMA. 2015. "Technical and User's Manual of Advanced Engineering Building Module (AEBM) 'Hazus MH 2.1.'" Federal Emergency Management Agency.

FEMA P-807. 2012. "Seismic Evaluation and Retrofit of Multi-Unit Wood-Frame Buildings With Weak First Stories, FEMA P-807." Federal Emergency Management Agency, Washington, *DC.*, no. May: 336.

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).

Ghobarah, A., 2004, June. On drift limits associated with different damage levels. In International workshop on performance-based seismic design (Vol. 28). Ontario, Canada: Department of Civil Engineering, McMaster University.

Goulet, C.A., Kottke, A., Boore, D.M., Bozorgnia, Y., Hollenback, J., Kishida, T., Der Kiureghian, A., Ktenidou, O.J., Kuehn, N.M., Rathje, E.M. and Thompson, E.M., 2018, May. Effective amplitude spectrum (EAS) as a metric for ground motion modeling using Fourier amplitudes. In 2018 Seismology of the Americas Meeting.

Harris, S.K. and Egan, J.A., 1992. The Loma Prieta, California, Earthquake of October 17, 1989. The Loma Prieta, California, Earthquake of October 17, 1989--Marina District: Strong Ground Motion and Ground Failure, (1551), p.181.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Holmes, W.T. and Somers, P., 1996. Steel Buildings. Northridge Earthquake of January 17, 1994, Reconnaissance Report, 2.

Hom, D.B. and Poland, C.D., 2004. ASCE 31-03: Seismic Evaluation of Existing Buildings. In Structures 2004: Building on the Past, Securing the Future (pp. 1-9).

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

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

Kalfarisi, R., Hmosze, M. and Wu, Z.Y., 2022. Detecting and Geolocating City-Scale Soft-Story Buildings by Deep Machine Learning for Urban Seismic Resilience. Natural Hazards Review, 23(1), p.04021062.

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

Kingma, D.P. and Ba, J.L., 2015, May. Adam: A method for stochastic gradient descent. In ICLR: International Conference on Learning Representations (pp. 1-15).

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.

Law, S., Paige, B. and Russell, C., 2019. Take a look around: using street view and satellite images to estimate house prices. ACM Transactions on Intelligent Systems and Technology (TIST), 10(5), pp.1-19.

Lee, E.J., Chen, P., Jordan, T.H., Maechling, P.B., Denolle, M.A. and Beroza, G.C., 2014. Full-3-D tomography for crustal structure in southern California based on the scattering-integral and the adjoint-wavefield methods. Journal of Geophysical Research: Solid Earth, 119(8), pp.6421-6451.

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 Proceedings of the 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.

Los Angeles City, Street Centerlines and Street Intersections in the City of Los Angeles. https://data.lacity.org/City-Infrastructure-Service-Requests/Street-Centerline/7j4e-nn4z.

Los Angeles Department of Building and Safety, 2022. online, https://www.ladbs.org/. (last accessed November 2022)

Lu, X., Xu, Y., Tian, Y., Cetiner, B. and Taciroglu, E., 2021. A deep learning approach to rapid regional post-event seismic damage assessment using time-frequency distributions of ground motions. Earthquake Engineering & Structural Dynamics, 50(6), pp.1612-1627.

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), p.04020230.

Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org (last accessed November 2022)

Mazzoni, S., McKenna, F., Scott, M.H. and Fenves, G.L., 2006. The open system for earthquake engineering simulation (OpenSEES) user command-language manual.

Open Street Map, online, https://www.openstreetmap.org. (last accessed November, 2022)

Oppenheim, A.V., Buck, J., Daniel, M., Willsky, A.S., Nawab, S.H. and Singer, A., 1997. Signals & systems. Pearson Educación.

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.

Public Policy Institute of California, 2006, available at https://www.ppic.org/ (last accessed November 2022)

Ranf, R.T., Eberhard, M.O. and Malone, S., 2007. Post-earthquake prioritization of bridge inspections. Earthquake Spectra, 23(1), pp.131-146.

Williams, C.K. and Rasmussen, C.E., 2006. Gaussian processes for machine learning, vol. 2, no. 3.

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., Petersson, N.A., 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.

Sajedi, S.O. and Liang, X., 2020. A data-driven framework for near real-time and robust damage diagnosis of building structures. Structural Control and Health Monitoring, 27(3), p.e2488.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

Shao, J., 1993. Linear model selection by cross-validation. Journal of the American statistical

Association, 88(422), pp.486-494.

Sheibani, M., Ou, G. and Zhe, S., 2020. Rapid seismic risk assessment of structures with Gaussian process regression. In Dynamic Substructures, Volume 4 (pp. 159-165). Springer, Cham.

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Southern California Earthquake Data Center, 2021. available at https://service.scedc.caltech.edu/SCSNStationMap/station.html (last accessed November 2022).

Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A.A., 2017, February. Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

Tamhidi, A., Kuehn, N., Ghahari, S.F., Rodgers, A.J., Taciroglu, E. and Bozorgnia, Y., 2022a. Earthquake Ground Motion Conditioned Simulation Using Sparsely Distributed Observed Motions for Analysis and Design of Lifeline Structures. In Lifelines 2022 (pp. 167-174). https://doi.org/10.1061/9780784484449.015

Tamhidi A, Kuehn NM and Bozorgnia Y, 2022b. Uncertainty and Sensitivity Analysis of Conditioned Simulation of Ground Motion using Gaussian Process Regression. University of California, Los Angeles. Natural Hazards Risk & Resiliency Research Center. DOI: 10.34948/N39G6W

Tamhidi A, Kuehn NM, Bozorgnia Y., 2022c. Uncertainty quantification of ground motion time

series generated at uninstrumented sites. Earthquake Spectra. doi:10.1177/87552930221135286

Tamhidi, A., Alipour, M., Kuehn, N., Yi, Z., Bozorgnia, Y. and Taciroglu, E., 2022d. Rapid Structural Model Development for Soft-Story Buildings using Deep Learning Image Classification. In 12th National Conference on Earthquake Engineering, Earthquake Engineering Research Institute, Salt Lake City, UT.

Tamhidi, A., Kuehn, N.M., Kohler, M.D., Ghahari, F., Taciroglu, E. and Bozorgnia, Y., 2020, September. Ground-motion time-series interpolation within the community seismic network using Gaussian process regression: Application to the 2019 Ridgecrest earthquake. In Poster Presentation at 2020 SCEC Annual Meeting.

Tamhidi, A., Kuehn, N., Bozorgnia, Y., Taciroglu, E. and Kishida, T., 2019. Prediction of ground-motion time-series at an arbitrary location using Gaussian process interpolation: Application to the Ridgecrest earthquake. In Poster Presentation at 2019 SCEC Annual Meeting.

Tamhidi, A., Kuehn, N., Ghahari, S.F., Rodgers, A.J., Kohler, M.D., Taciroglu, E. and Bozorgnia, Y., 2021. Conditioned Simulation of Ground-Motion Time Series at Uninstrumented Sites Using Gaussian Process Regression. Bulletin of the Seismological Society of America, 112(1), pp.331-347.

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.

Torrey, L. and Shavlik, J., 2010. Transfer learning. In Handbook of research on machine learning

applications and trends: algorithms, methods, and techniques (pp. 242-264). IGI global.

Trifunac, M. D., and M. I. Todorovska. 1997. "Northridge, California, Earthquake of 1994: Density of Red-Tagged Buildings versus Peak Horizontal Velocity and Intensity of Shaking." *Soil Dynamics and Earthquake Engineering* 16 (3): 209–22. https://doi.org/10.1016/S0267-7261(96)00043-7.

U.S. Geological Survey, 2013. USGS Response to an Urban Earthquake-Northridge 94. Available at: https://pubs.usgs.gov/of/1996/ofr-96-0263/execsum.htm (Last accessed November 2022.)U.S. Geological 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 November 2022).

U.S. Geological Survey, 2019. M7.1-2019 Ridgecrest earthquake sequence, available at https://earthquake.usgs.gov/earthquakes/ eventpage/ci38457511/executive (last accessed November 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 November 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.

Xia R, Schleuss J. LA releases addresses of 13,500 apartments and condos likely to need earthquake retrofitting. Los Angel Times, 2016.

Yi, Z., 2020. Performance-based analytics-driven seismic retrofit of woodframe buildings. University of California, Los Angeles.

Yu, Q., Wang, C., McKenna, F., Yu, S.X., Taciroglu, E., Cetiner, B. and Law, K.H., 2020. Rapid visual screening of soft-story buildings from street view images using deep learning classification. Earthquake Engineering and Engineering Vibration, 19(4), pp.827-838.

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.

Zillow. 2021. “ZTRAX: Zillow Transaction and Assessor Dataset.” 2021. https://www.zillow.com/research/ztrax/ (last accessed November 2022).