

# Uncertainty quantification of ground motion time series generated at uninstrumented sites

Earthquake Spectra

1–26

© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/87552930221135286

[journals.sagepub.com/home/eqs](https://journals.sagepub.com/home/eqs)

Aidin Tamhidi, M.EERI<sup>1,2</sup> , Nicolas M Kuehn<sup>1,2</sup> , and  
Yousef Bozorgnia, M.EERI<sup>1,2</sup>

## Abstract

Following an earthquake, ground motion time series are needed to carry out site-specific nonlinear response history analysis. However, the number of currently available recording instruments is sparse; thus, the ground motion time series at uninstrumented sites must be estimated. Tamhidi et al. developed a Gaussian process regression (GPR) model to generate ground motion time series given a set of recorded ground motions surrounding the target site. This GPR model interpolates the observed ground motions' Fourier Transform coefficients to generate the target site's Fourier spectrum and the corresponding time series. The robustness of the optimized hyperparameter of the model depends on the surrounding observation density. In this study, we carried out sensitivity analysis and tuned the hyperparameter of the GPR model for various observation densities. The 2019 **M7.1** Ridgecrest and 2020 **M4.5** South El Monte earthquake data sets recorded by the Community Seismic Network and California Integrated Seismic Network in Southern California are used to demonstrate the process. To provide a tool to quantify the uncertainty of the generated motions, a methodology to develop realizations of ground motion time series is also incorporated. The results illustrate that the uncertainty of the generated motions is lower at longer periods. It is shown that the observation density in the proximity of the target site plays a vital role in both error and uncertainty reduction of the generated time series. To demonstrate the concept, the effect of additional observations from combined recording networks is investigated.

## Keywords

Ground motion simulation, Gaussian process regression, uncertainty quantification, conditioned ground motions

Date received: 12 May 2022; accepted: 10 October 2022

<sup>1</sup>Civil and Environmental Engineering Department, University of California, Los Angeles, Los Angeles, CA, USA

<sup>2</sup>B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles, Los Angeles, CA, USA

## Corresponding author:

Aidin Tamhidi, Civil and Environmental Engineering Department, University of California, Los Angeles, 580 Portola Plaza, Boelter Hall, Los Angeles, CA 90095, USA.

Email: [aidintamhidi@ucla.edu](mailto:aidintamhidi@ucla.edu)

## Introduction

The current number of ground-level recording instruments is approximately 2000 in California over multiple recording networks (Southern California Earthquake Data Center, 2022). Consequently, following a major event, ground motion time series at locations devoid of recording instruments are needed to carry out site-specific nonlinear response history analysis. Currently, important products “ShakeCast” and “ShakeMap” established by US Geological Survey (Fraser et al., 2008; Lin et al., 2018; Wald et al., 2008; Worden et al., 2018) provide ground motion intensity measures (GMIM), such as peak ground acceleration and response spectral ordinates. However, as stated above, ground motion time series at uninstrumented sites are also required for seismic performance evaluation of specific structures.

Two commonly used methodologies for generating ground motion time series at uninstrumented sites can be mentioned here. First, the physics-based simulations employing the finite-fault and seismic velocity models consider the source, path, and site effects (e.g. Aagaard et al., 2008; Atkinson and Assatourians, 2015) and the topography of the Earth’s surface (Rodgers et al., 2019; Thomson et al., 2020). Second, the coherency function-based simulations employing cross-spectral density (CSD) and auto-spectral density (ASD) functions (e.g. Kameda and Morikawa, 1992; Konakli and Der Kiureghian, 2012; Zentner, 2013; Rodda and Basu, 2018). Several research studies have also been conducted to simulate unconditional spatially varying ground motions. Deodatis (1996) and Shinozuka and Deodatis (1996) employed the spectral representation method (SRM) to simulate non-stationary stochastic ground motions. Furthermore, conditional simulation of non-stationary random fields was extensively investigated (Cui and Hong, 2020; Heredia-Zavoni and Santa-Cruz, 2000; Hu et al., 2012; Vanmarcke and Fenton, 1991; Wu et al., 2015). The physics-based approaches require detailed information regarding the subsurface properties and fault features. However, The CSDs are driven using empirical or semi-empirical coherency functions, whose coefficients are often determined via data-driven methods (Abrahamson et al., 1991). The CSD functions might need some detailed site properties and wave propagation characteristics. Therefore, both methods have challenges for a rapid post-earthquake structural damage assessment in real time (Loos et al., 2020; Mangalathu and Jeon, 2020) as they are computationally expensive and time-consuming.

Tamhidi et al. (2021, 2022b) recently developed a method to simulate ground motion time series using a trained Gaussian process regression (GPR) model. This GPR model interpolates the discrete Fourier transform (DFT) coefficients of the observed nearby recorded ground motions to construct the time series at the target uninstrumented sites. The GPR model performs conditioned simulation of ground motions at an ensemble of target locations, imposing a comparatively low computational cost (Rasmussen and Williams, 2006).

The intrinsic uncertainty of ground motions affects earthquake engineering disciplines, such as the performance-based post-earthquake assessment and decision-making (Aghababaei et al., 2021; Roohi and Hernandez, 2020; Weatherill et al., 2015). Several studies attempted to quantify and model this randomness in various seismic problems (Alamilla et al., 2001; Wen et al., 2003; Yazdi et al., 2022). In this study, we focus on quantifying the uncertainty and validity of the generated motions using the GPR model introduced by Tamhidi et al. (2021). The model’s hyperparameter (the regularization factor) is fine-tuned based on observation densities, enabling the users to choose the optimum hyperparameter corresponding to the existing observed data set. Then, a methodology to

generate random realizations of ground motions using the trained GPR and an inter-frequency correlation model (Bayless and Abrahamson, 2019) is reviewed. This random realization methodology provides a means to quantify the uncertainty of the generated motions at the target sites. We implemented this methodology to investigate the simulated motions' accuracy and uncertainty using the 2019 **M**7.1 Ridgecrest earthquake data set recorded by the Community Seismic Network (CSN). In addition, the performance of the model as related to the spatial density of recording instruments is investigated for the 2019 Ridgecrest and 2020 **M**4.5 South El Monte earthquakes in Southern California.

## Theoretical background

Suppose the ground motion acceleration time series at site  $s$ ,  $a_s(t)$ , is constructed with  $N$  discrete data points,  $a_s(t_i)$ ,  $i = 1, \dots, N$ . The time series can be decomposed into its DFT coefficients  $A_k$  (e.g. Oppenheim et al., 1997) as follows:

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

where

$$A_k = \frac{1}{N} \sum_{i=0}^{N-1} a_s(t_i) [\cos(\omega_k t_i) - j \cdot \sin(\omega_k t_i)] = \mathcal{R}e_k + j \cdot \mathcal{I}m_k. \quad (2)$$

In Equations 1 and 2,  $\omega_k$  denotes the  $k^{\text{th}}$  DFT's frequency and  $j = \sqrt{-1}$ .  $\mathcal{R}e_k$  and  $\mathcal{I}m_k$  are the DFT coefficient's real and imaginary components at the  $k^{\text{th}}$  frequency, respectively. The DFT coefficients are assumed as random Gaussian variables, as demonstrated by Kameda and Morikawa (1992) and implemented by Konakli and Der Kiureghian (2012) to simulate ground motion fields. The  $\mathcal{R}e_k$  (or  $\mathcal{I}m_k$ ) at site  $s$  is correlated to  $\mathcal{R}e_k'$  (or  $\mathcal{I}m_k'$ ) at nearby sites  $s'$ . We implemented a GPR model to estimate  $\mathcal{R}e_k$  (and  $\mathcal{I}m_k$ ) given the observed  $\mathcal{R}e_k'$  (and  $\mathcal{I}m_k'$ ) at the neighboring sites. 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 site for different frequencies  $k$  and  $j$ , when we construct the mean estimated ground motions. Tamhidi et al. (2021) showed that mean estimated values for a multivariate Gaussian variable (here  $\mathcal{R}e_k$  or  $\mathcal{I}m_k$ ) are independent of the inter-frequency correlation between amplitudes at different frequencies. However, this inter-frequency correlation must be taken into account if we want to generate random realizations of time series at a site.

## Gaussian Process Regression

A Gaussian process (GP) is a set of indexed random variables, with every finite subset following multivariate Gaussian distribution (Rasmussen and Williams, 2006). The standard form of GP is shown in Equation 3:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (3)$$

In Equation 3,  $m(\mathbf{x})$  represents the mean function value at the input vector location  $\mathbf{x}$ , and  $k(\mathbf{x}, \mathbf{x}')$  is the covariance between vector locations  $\mathbf{x}$  and  $\mathbf{x}'$ . This article indicates vectors

and matrices with lower-case and upper-case boldface symbols, respectively. Suppose there are  $N_o$  and  $N_t$  number of observed and target locations, respectively. We denote  $\mathbf{f}$  as a  $N_o \times 1$  vector of observed values from a GP and  $\mathbf{f}_*$  as a  $N_t \times 1$  vector of unknown GP values at the target locations. Also, let us symbolize the observed locations'  $N_o \times d$  input matrix as  $\mathbf{X}$  ( $d$  is the number of features for each location), whose rows contain observed locations' input feature vectors,  $\mathbf{x}$ . Similarly, we call  $\mathbf{X}_*$  as the  $N_t \times d$  matrix of target locations. The predictive distribution of  $\mathbf{f}_*$  is given by (Rasmussen and Williams, 2006),

$$\mathbf{f}_* | \mathbf{X}_*, \mathbf{X}, \mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_{**}) \quad (4)$$

where

$$\boldsymbol{\mu}_* (N_t \times 1) = \boldsymbol{\mu} (N_o \times 1) + \mathbf{K}_{x_*x} (N_t \times N_o) \mathbf{K}_{xx}^{-1} (N_o \times N_o) (\mathbf{f} (N_o \times 1) - \boldsymbol{\mu} (N_o \times 1)) \quad (5)$$

$$\boldsymbol{\Sigma}_{**} (N_t \times N_t) = \mathbf{K}_{x_*x_*} (N_t \times N_t) - \mathbf{K}_{x_*x} (N_t \times N_o) \mathbf{K}_{xx}^{-1} (N_o \times N_o) \mathbf{K}_{xx_*} (N_o \times N_t) \quad (6)$$

In Equation 4,  $\boldsymbol{\mu}_*$  and  $\boldsymbol{\Sigma}_{**}$  stands for the posterior mean vector and covariance matrix of the GP values at target locations, respectively. In Equations 5 and 6,  $\boldsymbol{\mu}$  denotes a vector of GP prior mean;  $\mathbf{K}_{xx}$  and  $\mathbf{K}_{x_*x_*}$  are the covariance matrix of the DFT coefficients at the observed and target locations, respectively. Correspondingly,  $\mathbf{K}_{xx_*}$  (transpose of  $\mathbf{K}_{x_*x}$ ) represents the covariance between the observed and predicted DFT coefficients. The covariance matrices' elements are constructed with covariance kernel,  $k(r)$ , where  $r$  denotes the distance between input vectors. Tamhidi et al. (2021) indicated that the Matérn kernel with  $\nu = 1.5$  is the optimum covariance function for the GPR model to simulate the ground motion time series. Equations 7 and 8 illustrate Matérn ( $\nu = 1.5$ ) kernel function and the “distance” between two input vectors,  $\mathbf{x}$  and  $\mathbf{x}'$ , respectively. The “distance” can be the geographical distance and the difference between local soil conditions ( $V_{s30}$ ) at two sites as elaborated below.

$$k_{\nu=1.5}(r) = \sigma_f^2 (1 + \sqrt{3}r) \exp(-\sqrt{3}r) \quad (7)$$

$$r = \theta \sqrt{\sum_{i=1}^d (x_i - x'_i)^2} \quad (8)$$

In Equation 7,  $\sigma_f$  is the variance that governs how uncertain the GPR's estimate is. In Equation 8,  $x_i$  is the  $i^{\text{th}}$  component of the input vector at location  $\mathbf{x}$  and  $\theta$  is a positive scaling factor, as the inverse of length-scale,  $l$ , where  $\theta = \frac{1}{l}$ . The length-scale,  $l$ , is a parameter of the covariance function, which scales the “distance” between two locations. In other words, a larger length-scale results in a higher correlation by reducing the “distance.” In this study, all input vector's elements are scaled with the same  $\theta$ . Such a covariance function is called an isotropic covariance function.

Tamhidi et al. (2020, 2021) demonstrated that a four-dimension input vector,  $\mathbf{x} = \{x_1, x_2, x_3, \log(V_{s30})\}$ , is appropriate to represent sites.  $V_{s30}$  is the time-average shear wave velocity in the uppermost 30 m of the soil, and  $x_1$  through  $x_3$  are the Cartesian coordinates of the site on the 3D surface of the Earth. All four input attributes are normalized, so that each feature's mean and standard deviation are zero and one, respectively. The GPR model's parameters are the distance scaling factor,  $\theta$ , the GP prior mean,  $\boldsymbol{\mu}$ , and the



variance,  $\sigma_f$ , which need to be optimized for each  $\mathcal{R}e_k$  (and  $\mathcal{I}m_k$ ) at the target site using the observations which are all known ground motions  $\mathcal{R}e'_k$  (and  $\mathcal{I}m'_k$ ) within the corresponding event's data set. The maximum a posteriori estimates are chosen as the optimum model's parameters by maximizing the penalized log-likelihood of the observations. Denoting the parameters as  $\boldsymbol{\gamma} = (\theta, \boldsymbol{\mu}, \sigma_f)$ , Equation 9 displays the penalized log-likelihood of the observations:

$$\mathcal{Q}(\boldsymbol{\gamma}) = -\frac{1}{2}(\mathbf{f} - \boldsymbol{\mu})^T \mathbf{K}_{xx}^{-1}(\mathbf{f} - \boldsymbol{\mu}) - \frac{1}{2} \log |\mathbf{K}_{xx}| - \frac{N_o}{2} \log 2\pi - N_o d p_\lambda(\theta) \quad (9)$$

In Equation 9, T stands for the transpose operator and  $p_\lambda(\theta)$  is a non-negative penalty function for the scaling factor  $\theta$ . This study uses the L2 penalty function shown in Equation 10.

$$p_\lambda(\theta) = \lambda \theta^2 \quad (10)$$

The regularization factor,  $\lambda$ , is the hyperparameter of the model that determines how observations contribute the optimum parameters,  $\hat{\boldsymbol{\gamma}}$ .

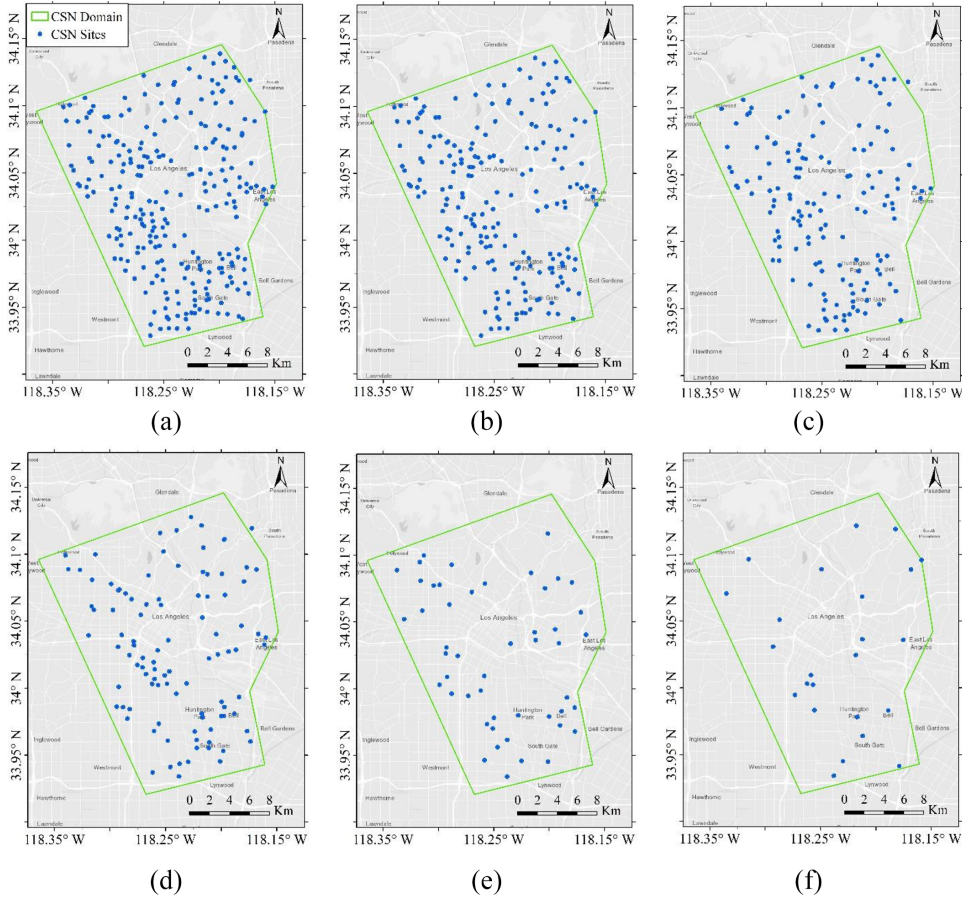
## Hyperparameter optimization

A higher penalty is needed when there are sparse observations (Li and Sudjianto, 2005). In other words, the optimum hyperparameter,  $\hat{\lambda}$ , depends on the observation density. Thus, it is required to tune the  $\hat{\lambda}$  for various observation densities. We used the recorded ground motions of the 2019 M7.1 Ridgecrest earthquake by the CSN within Los Angeles (Clayton et al., 2020).

The observation density for the 252 CSN recording sites distributed over a 464 km<sup>2</sup> region (CSN domain in Figure 1) is 0.54 sites/km<sup>2</sup>. To tune the optimum  $\hat{\lambda}$  for various observation densities, we make different data sets with the various number of observed sites by randomly selecting out of the 252 CSN sites. Six different data sets with 252, 201, 151, 100, 50, and 25 sites are chosen. The distribution of the randomly chosen subsets is shown in Figure 1. The  $V_{s30}$  values of the CSN sites are estimated using a proxy-based model described in Ahdi et al. (2020). The selection criterion for finding  $\hat{\lambda}$  is to minimize the average normalized root mean square error (NRMSE) between the recorded and generated ground motions' 5%-damped RotD50 response spectra (Boore, 2010). The NRMSE between the recorded and estimated motions' response spectra is calculated by Equation 11.

$$NRMSE = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} \frac{(PSA_i - \widehat{PSA}_i)^2}{\widehat{PSA}_i^2}} \quad (11)$$

In Equation 11,  $PSA_i$  and  $\widehat{PSA}_i$  are the response spectral ordinates of the estimated and recorded ground motions at the  $i^{\text{th}}$  period, respectively, and  $N_p$  equals to the number of periods included within the usable bandwidth. The usable bandwidth is considered as the mutual usable bandwidth among all observed motions, which is the reliable period range after the noise removal of the motions (Ancheta et al., 2014). We implemented the Leave-One-Out (LOO) cross-validation methodology (Vehtari et al., 2017) to find  $\hat{\lambda}$  for each subset shown in Figure 1. Nine different regularization factors, 0.01, 0.02, 0.05, 0.1, 0.2, 0.4,



**Figure 1.** Distribution of the randomly chosen subsets from CSN's recorded 2019 M7.1 Ridgecrest earthquake motions with (a) 252, (b) 201, (c) 151, (d) 100, (e) 50, and (f) 25 number of sites.

0.6, 0.8, and 1.0, called as  $\lambda_{test}$ , are examined for each subset through the following steps to find the  $\hat{\lambda}$ . These  $\lambda_{test}$  values are chosen, so that the  $\hat{\lambda}$  for each subset falls inside the 0.01–1.0 domain (cf. Table 1).

1. For each site,  $s$ , within the data set,  $s = 1, \dots, N_{sites}$  ( $N_{sites}$  = number of sites in the data set)
  - 1.1. Obtain the optimum parameters  $\hat{\gamma}_k$  at the  $k^{th}$  frequency using the observed ground motions, which are all the recorded ground motions (excluding at site  $s$ ) by maximizing  $Q(\gamma_k)$  using  $\lambda_{test}$ .
  - 1.2. Generate the ground motion time series at site  $s$  using posterior mean (Equation 5) of  $Re_k$  and  $Im_k$ ,  $k=0, \dots, N-1$ , given  $\hat{\gamma}_k$  in Step 1.1. We call this generated time series the mean estimated ground motion.
  - 1.3. Obtain the RotD50 spectrum of the mean estimated and recorded ground motions at site  $s$  and calculate NRMSE between them, called as  $Error_s$ .

**Table 1.** The optimum regularization factor,  $\hat{\lambda}$ , and corresponding  $Error_{avg}$  for various observation densities

| No. of observation | Density (station/km <sup>2</sup> ) | $\hat{\lambda}$ | 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                                   |

NRMSE: normalized root mean square error.

2. Compute the average of  $Error_s$  among all sites,  $s = 1, \dots, N_{sites}$ , and store it as  $Error_{avg}$  corresponding to  $\lambda_{test}$ .

Eventually, the  $\lambda_{test}$  corresponding lowest  $Error_{avg}$  is chosen as  $\hat{\lambda}$  associated with the target data set's observation density. Table 1 illustrates the obtained  $\hat{\lambda}$  for each data set and the corresponding  $Error_{avg}$  for the mean estimated ground motions.

We have employed the  $\hat{\lambda}$  given in Table 1 to generate the mean estimated ground motion time series using LOO analysis at each site within the corresponding data set. Figure 2 depicts the distribution of the NRMSE between the recorded and mean estimated ground motions' RotD50 spectrum. It is demonstrated that the predicted ground motions are reliably accurate for most of the target sites for the subsets with higher observation density (cf. Figures 2a, through c). However, the  $Error_{avg}$  increases as observation density decreases (cf. Figures 2d through f). The effect of observation density on the prediction accuracy and uncertainty is examined in further depth in section "Uncertainty quantification and sensitivity analysis."

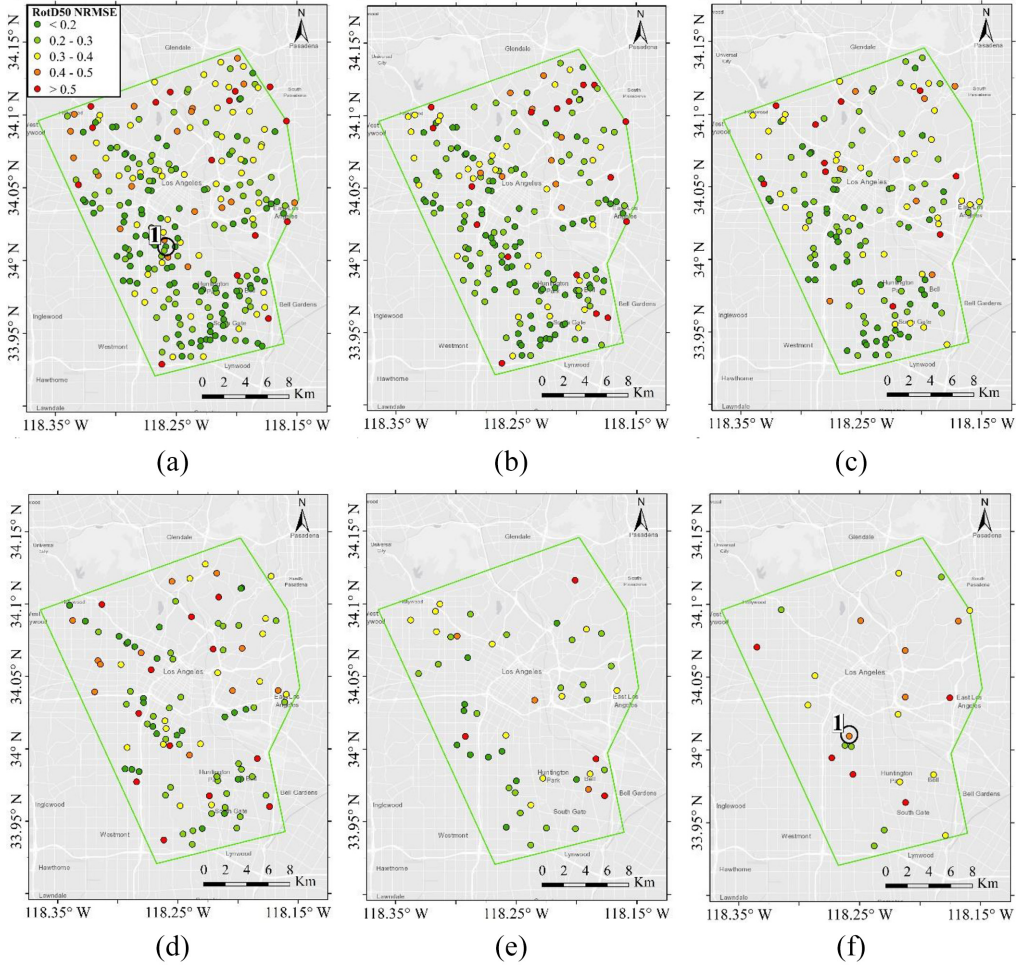
## Ground motion random realizations

It is desirable to quantify the uncertainty of the mean estimated ground motions at any target site. The posterior mean vector and covariance matrix for all target sites'  $k^{\text{th}}$  frequency DFT coefficients,  $k = 0, \dots, N-1$ , are given by Equations 5 and 6. In this study, we generate ground motion time series at one target site at each process. Therefore, Equations 5 and 6 can be converted to Equations 12 and 13, providing the scalar posterior mean,  $\mu_*$ , and standard deviation,  $\sigma_*$ , for each  $\mathcal{Re}_k$  (and  $\mathcal{Im}_k$ ):

$$\mu_* = \hat{\mu} + \mathbf{k}_{x,x(1 \times N_o)} \mathbf{K}_{xx}^{-1} (N_o \times N_o) \left( \mathbf{f}_{(N_o \times 1)} - \boldsymbol{\mu}_{(N_o \times 1)} \right) \quad (12)$$

$$\sigma_* = \widehat{\sigma_f} - \mathbf{K}_{x,x(1 \times N_o)} \mathbf{K}_{xx}^{-1} (N_o \times N_o) \mathbf{K}_{xx} (N_o \times 1) \quad (13)$$

In Equations 12 and 13,  $\mathbf{k}_{x,x}$  is the vector of covariance among the target site and observed sites  $\mathcal{Re}_k$  (or  $\mathcal{Im}_k$ ).  $\mathbf{k}_{x,x}$  and  $\mathbf{K}_{xx}$  are established using  $\hat{\gamma}$  at the corresponding frequency. The ground motion realizations at each target site are generated following below steps:



**Figure 2.** 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 the assessment of the introduced methodology for random generation of ground motions (see section “Ground motion random realizations”).

1. At each  $k^{\text{th}}$  frequency,  $k = 0, \dots, N-1$ :
  - 1.1. The posterior mean and standard deviation of  $\mathcal{R}e_k$  and  $\mathcal{I}m_k$  are calculated by Equations 12 and 13.
  - 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 data set except target site) ground motions. Consequently, a  $2 \times 2$  covariance matrix for the  $(\mathcal{R}e_k, \mathcal{I}m_k)$  is established using the estimated correlation and standard deviations resulted in Step 1.1.

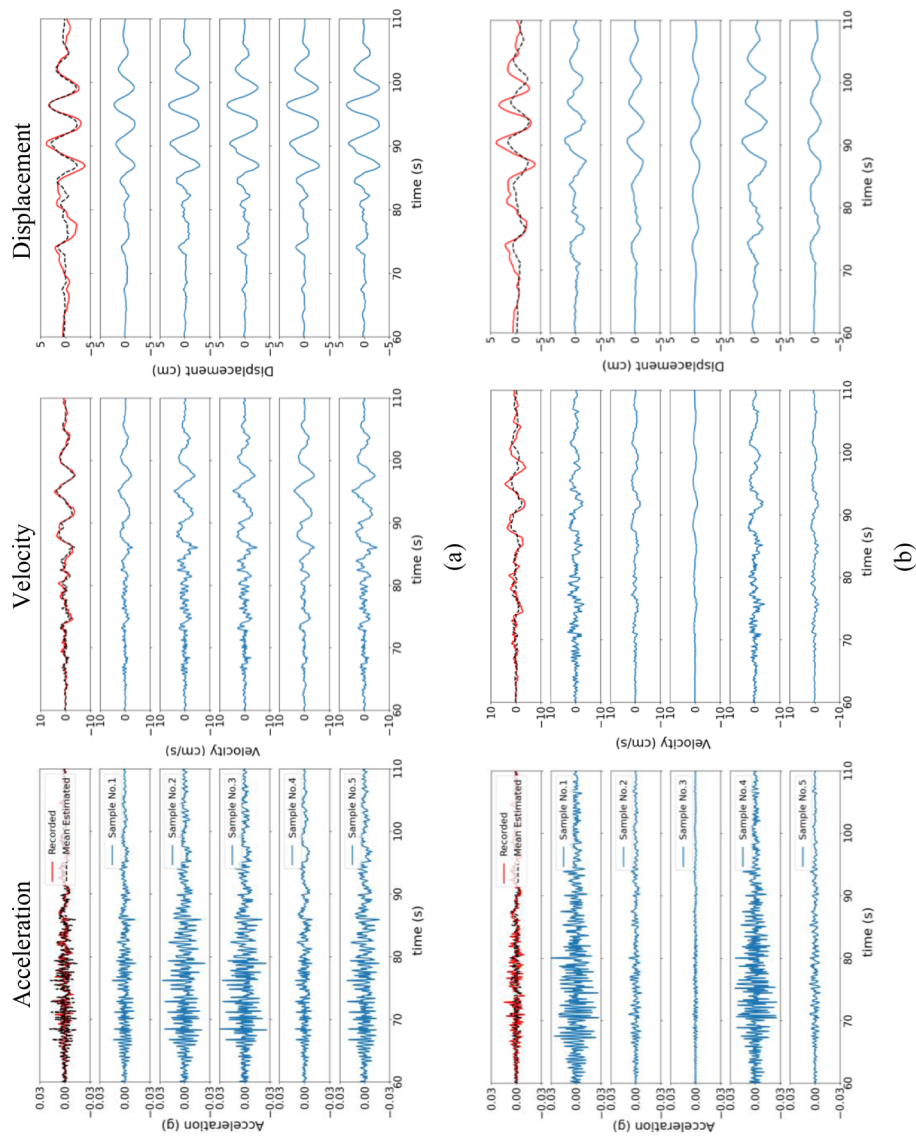
**Table 2.** Site properties of the Test Site 1 shown in Figure 2

| Coordinates (longitude, latitude) | $V_{s30}$ (m/s) | $Z_{1.0}$ (km) | $Z_{2.5}$ (km) | $R_{rup}$ (km) | Hypocentral distance (km) |
|-----------------------------------|-----------------|----------------|----------------|----------------|---------------------------|
| (118.258°W, 34.009°N)             | 290             | 0.62           | 4.48           | 191.7          | 204.1                     |

- 1.3. A set of random samples of  $2 \times 1$  vectors of  $(\mathcal{R}e_k, \mathcal{I}m_k)$  are produced using the estimated  $2 \times 1$  mean vector (Step 1.1) and  $2 \times 2$  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  $N \times N$  covariance matrix of  $\log(|A_k|)$ ,  $k = 0, \dots, N-1$ , are constructed using the inter-frequency correlation values given by Bayless and Abrahamson (2019) model and calculated standard deviations in Step 1.4.
3. Random Gaussian  $N \times 1$  vector samples of  $\log(|A_k|)$  are produced using the  $N \times 1$  mean vector (Step 1.4) and  $N \times N$  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 Fourier amplitude spectrum (FAS) in Step 3 are combined with the Phase spectrum constructed with the posterior mean DFT coefficients to generate ground motion time series realizations.

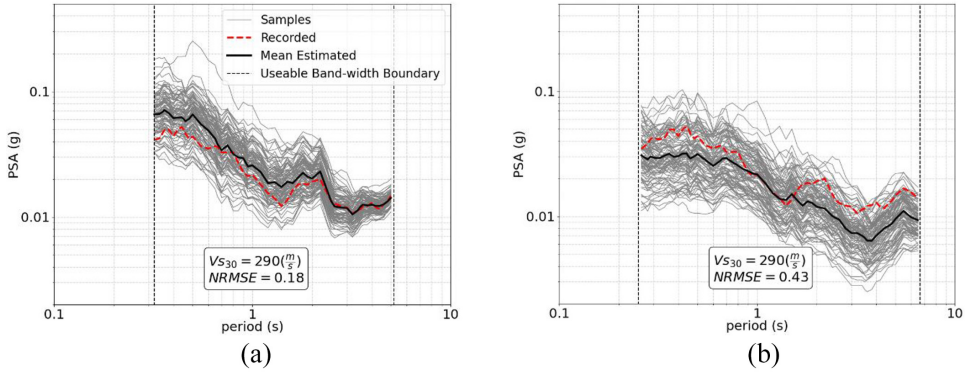
We also examined the randomization of Fourier phase spectra; yet, the results were not as promising as the outcomes stated in Step 4 above. We employ the 2019 M7.1 Ridgecrest earthquake data set recorded over the CSN sites to evaluate the proposed methodology. Test Site 1 shown in Figures 2a and f is the target site. The geotechnical properties of Test Site 1 are summarized in Table 2. In Table 2,  $Z_{1.0}$  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 CSM-S4 model (Nweke et al., 2018). In Table 2,  $R_{rup}$  is the closest distance to the coseismic rupture.

Two different observed sets are considered to generate ground motion realizations at Test Site 1; first, all 251 CSN sites in Figure 2a, and second, all 24 CSN sites in Figure 2f. The  $\hat{\lambda}$  for each case is chosen from Table 1. We generated 100 ground motion realizations at Test Site 1. Figure 3 indicates the mean estimated and five ground motion time series realizations along the East–West (EW) direction at the Target Site 1. Figure 3a 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 3b. In addition, it is observed from Figure 3a 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 3a). However, the higher frequency content of the generated motions shows a greater degree of uncertainty even using 251 observations (cf. accelerations in Figure 3a). However, Figure 3b 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 3b). The reason for this is because the average distance between the 24 observed sites is not close enough to predict long waves of the motion.

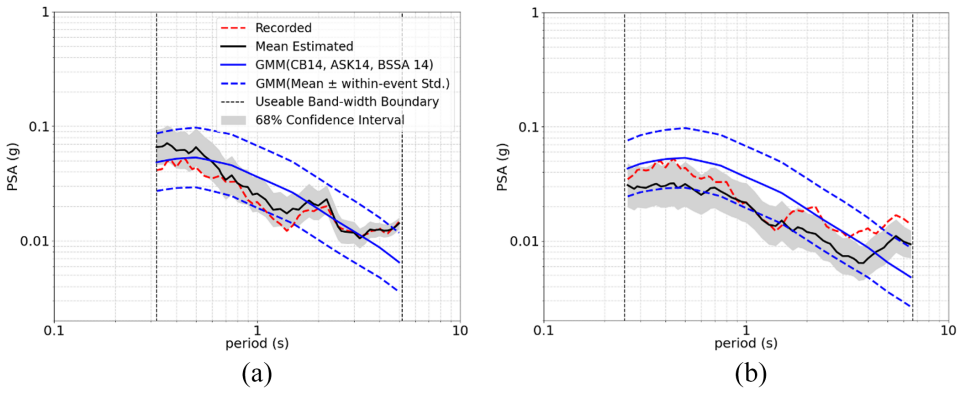


**Figure 3.** Estimated mean and five generated ground motion realizations time series along EW direction at the Test Site 1 within the **M7.1** Ridgecrest earthquake CSN data set shown in Figures 2a and f using (a) 251 and (b) 24 observed sites.





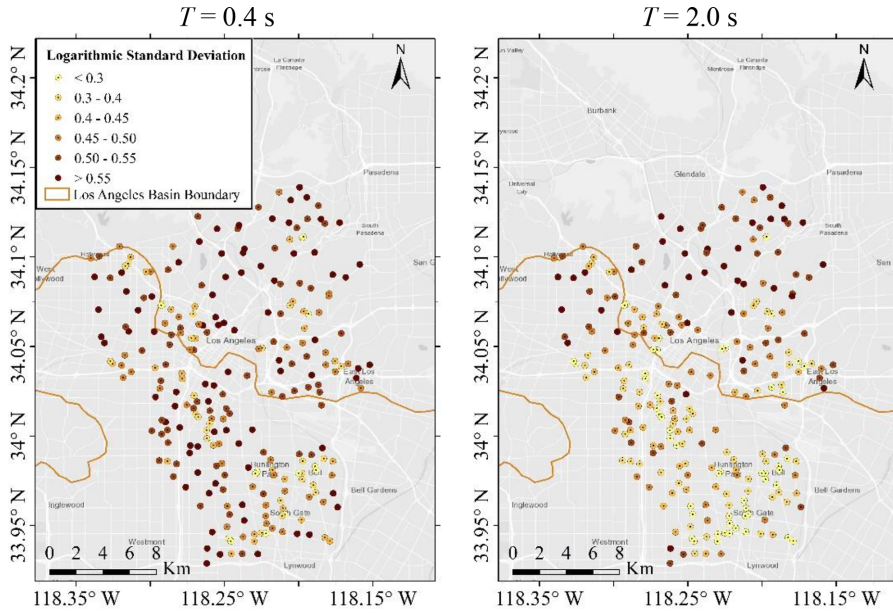
**Figure 4.** The 5%-damped RotD50 spectrum of generated ground motion realizations at the Test Site I using the (a) 251 (Figure 2a) and (b) 24 (Figure 2f) observed sites.



**Figure 5.** The 68% CI of RotD50 spectrum of generated ground motion realizations at the Test Site I using the (a) 251 (Figure 2a) and (b) 24 (Figure 2f) observed sites.

Figure 4 depicts the mean estimated and 100 ground motion realizations' 5%-damped RotD50 spectra at Test Site 1 using 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 4 indicates that the long-period prediction (longer than 1 s) has minor variation and error for having 251 observed sites than those estimated with 24 observations. However, 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 4a and b that the short periods' variation is high and does not differ significantly from having 251 observations to 24 ones.

Figure 5 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 5 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 6.** The PSA logarithmic standard deviation of the estimated ground motions along EW direction at two periods  $T = 0.4$  and  $2.0$  s for the **M7.1** Ridgecrest earthquake CSN data set.

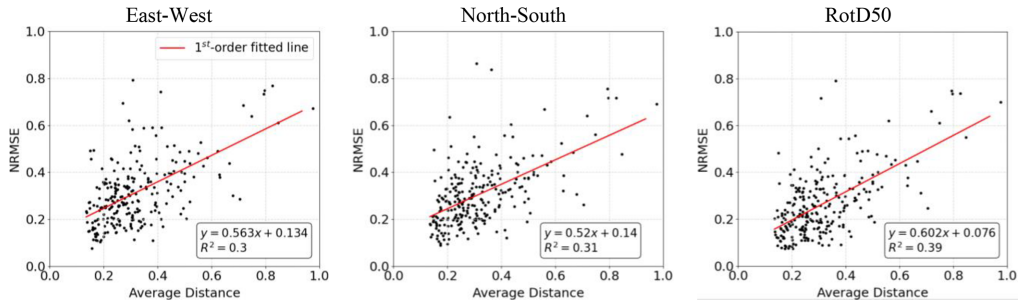
Figure 5a indicates that the recorded ground motion response spectrum falls inside 68% CI of the generated motions using 251 observations for the majority of periods. Furthermore, Figure 5a shows that within-event uncertainty for the average GMMs is greater than that of generated motions using 251 observations. Figure 5a 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. However, Figure 5b 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 5 indicates that the standard deviation from short to long periods does not alter considerably for having fewer observations. Interested readers are referred to Tamhidi et al. (2022a) to know more about the observation density's effect on 68% CI of estimated motions at other target sites of CSN.

## Uncertainty quantification and sensitivity analysis

Accuracy and uncertainty of the generated time series are quantified in this section. We employed 252 CSN sites' LOO analysis results for the 2019 **M7.1** Ridgecrest earthquake. The logarithmic standard deviation of 100 generated pseudo-spectral accelerations (PSAs) at two periods,  $T = 0.4$  and  $2.0$  s, is obtained as a measure of generated motions' uncertainty at short and long periods, respectively. Figure 6 depicts the distribution of the EW PSAs' logarithmic standard deviation at  $T = 0.4$  and  $2.0$  s.

Figure 6 illustrates that estimated ground motion realizations at CSN sites on the Los Angeles basin show minor variations at long periods ( $T = 2.0$  s) compared to those located outside the basin. However, the generated motions' uncertainty at short period,



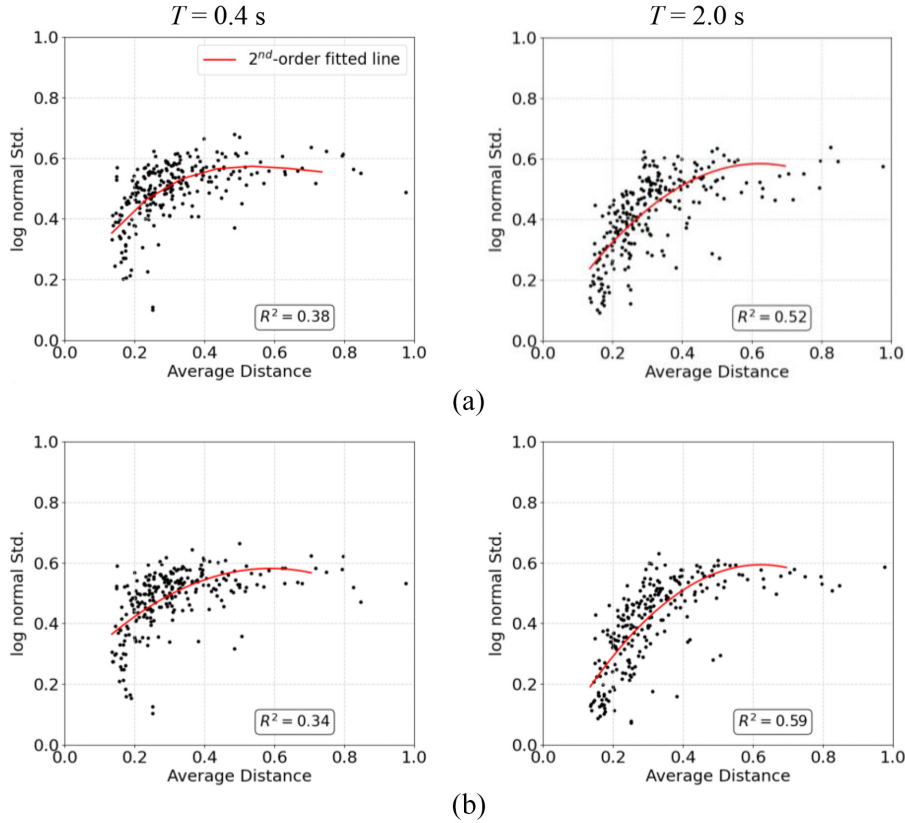


**Figure 7.** Scatter plot of the PSA NRMSE with respect to the average separation distance for the 2019 M7.1 Ridgecrest earthquake CSN data set.

$T = 0.4$  s, changes insignificantly between CSN sites inside and outside the Los Angeles basin. Comparing results at  $T = 0.4$  with 2.0 s in Figure 6 reveals that the PSA’s logarithmic standard deviation at long period is smaller than those of short period 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 southern part of the CSN is high enough to produce reliable long-period motions. Furthermore, the sites atop the Los Angeles basin receives 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.

As a metric of observation density surrounding each target site, we determined the average distance (inside the 4D space established in previous sections) between each target site and its four nearest observed neighbors. In this article, we refer to this distance as “average separation distance.” The shorter average separation distance indicates a higher observation density surrounding the target site. Figure 7 depicts the scatter plot of the mean estimated motions’ PSA NRMSE within usable bandwidth along EW, North–South (NS), and RotD50 concerning the average separation distance. Figure 7 also depicts the fitted lines to the scatter plots and their R-squared,  $R^2$ . The separation distance in Figure 7 is unitless as the feature vectors are all normalized, as elaborated in the “Theoretical background” section.

Figure 7 shows that estimation error and average separation distance have a general direct correlation. One can recognize that having more observations closer to the target site results in more accurate ground motion prediction, as expected and now quantified in Figure 7. Figure 8 indicates the scatter plot of the PSA’s logarithmic standard deviation along EW and NS at  $T = 0.4$  and 2.0 s relative to the average separation distance. Figure 8 displays that the estimation uncertainty at  $T = 2.0$  s grows as the average separation distance increases. In addition, it is noticeable that in a general trend, the uncertainty of the long-period estimation is sensitive to the observation density; but, the short-period estimation ( $T = 0.4$  s) is not significantly correlated with the number of observations surrounding the target site. 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  $T = 0.4$  and 2.0 s in Figure 8 reveals that the long-period motions have less variability than short-period ones at shorter average separation distances. Figure 8 demonstrates that for a target site with average

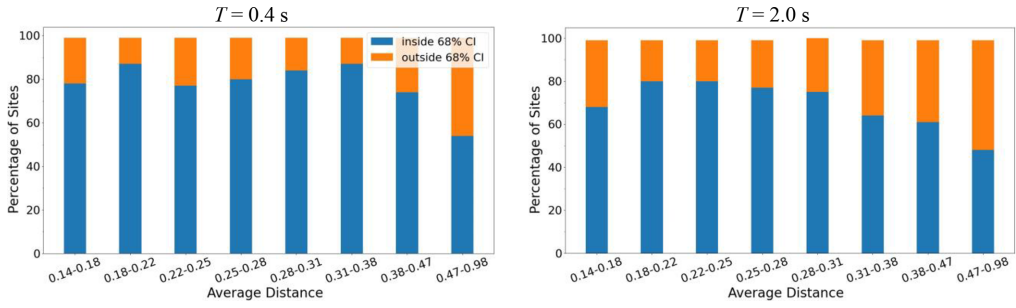


**Figure 8.** Scatter plot of the logarithmic standard deviation of PSA along (a) EW and (b) NS for the M7.1 Ridgecrest earthquake CSN data set.

distance of 0.2 from its four nearest neighbors, the estimated logarithmic standard deviation for PSA is around 0.45 and 0.30 at  $T = 0.4$  and  $2.0$  s, respectively. Furthermore, Figure 8 depicts that both short and long periods' uncertainty saturates for very long 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 9 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 regard to the average separation distance. The eight spans of average separation distance shown in Figure 9 are selected so that each span includes an approximately same number of target sites. Figure 9 indicates that the percentage of sites where the recorded PSA locates outside of the 68% CI rises as 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  $T = 0.4$  and  $2.0$  s, respectively.

Consequently, it may be concluded that the target sites with higher observation densities close to them are more likely to have the recorded PSA within their 68% CI estimation. About 76% and 74% of the 252 target sites' estimated PSA at  $T = 0.4$  s captures the recorded one within 68% CI along EW and NS, respectively. Similarly, 70% and 76% of



**Figure 9.** 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 data set.

the target sites' generated PSA at  $T = 2.0$  s includes the recorded spectral ordinate within 68% CI along EW and NS, respectively. The effect of other governing parameters, such as uncertainty of the predicted site conditions ( $V_{s30}$ ) and the surface slope of the nearby instrumented sites on estimations, is studied by Tamhidi et al. (2022a).

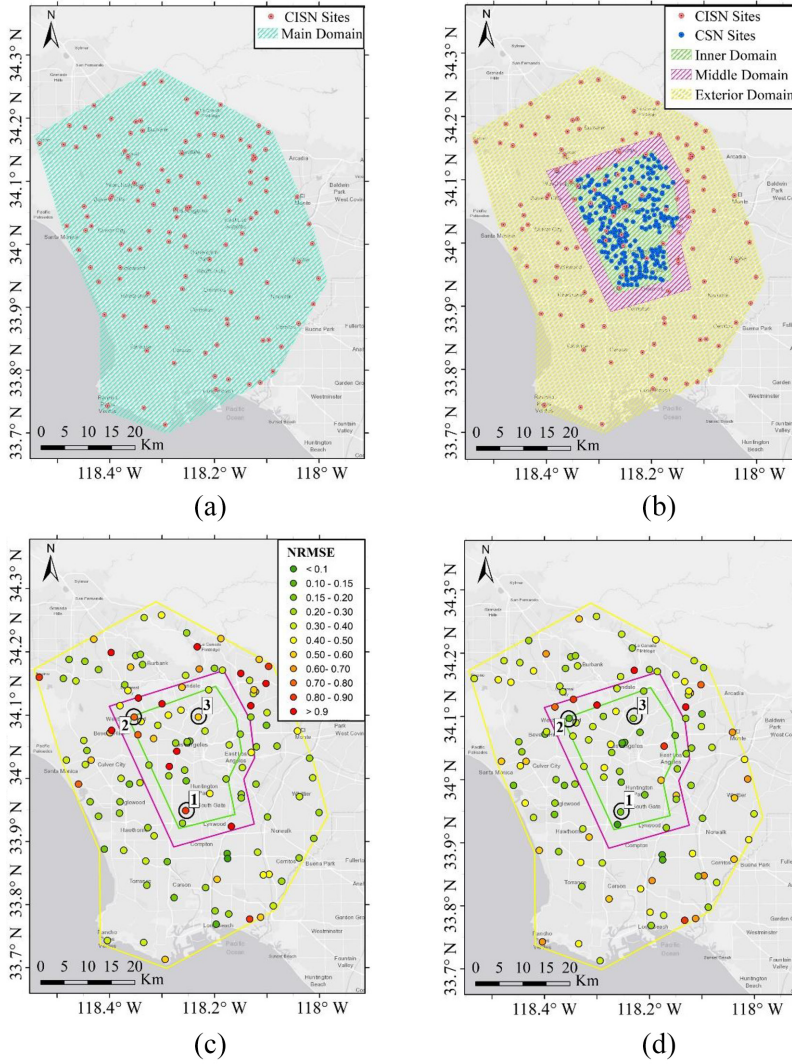
In summary, Figure 8 shows that increasing the density of instrumentation closer to the target site reduces the variability of the generated ground motions. As a result, the higher observation density is expected to decrease uncertainty of structural engineering demand parameters derived from nonlinear response history analyses. Figure 9 indicates that for the target sites with a greater observation density, the uncertainty of long-period 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 shorter than 0.3, the PSA realizations are about 80% likely to capture the recorded spectra within their mean  $\pm$  standard deviation bandwidth.

## Performance evaluation on combined network data sets

Herein, we 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, we execute LOO ground motion prediction at each CISN station as a target site using all other CISN sites (except the target site) as observation. Second, we 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 M7.1 Ridgecrest and (2) 2020 M4.5 South El Monte earthquakes, as elaborated below.

### 2019 M7.1 Ridgecrest earthquake

We selected 121 ground-level sites from CISN that recorded the 2019 M7.1 Ridgecrest earthquake in Los Angeles. These 121 recording sites are widely dispersed throughout a 3100 km<sup>2</sup> region the so-called Main domain (see Figure 10a), whereas the 252 CSN's sites are placed over a smaller 460 km<sup>2</sup> region. The distribution of the CSN and CISN



**Figure 10.** 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 2019 M7.1 Ridgecrest earthquake.

installations over Los Angeles is shown in Figure 10b. The  $\hat{\lambda}$  depends on the observation density for each target site, as illustrated in Table 1. Thus, we separated the Main domain into three subdomains: (1) Inner, (2) Middle, and (3) Exterior Domains (Figure 10b). There is one observation density when we use just CISN sites as observation. In contrast, the observation density and the required  $\hat{\lambda}$  are different when we integrate observations from CISN and CSN. Table 3 depicts the observation density and corresponding  $\hat{\lambda}$  for each domain. In Table 3, “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 Inner, Middle, and Exterior target domains is derived by dividing the number of sites available inside Inner, Inner plus Middle, and Main domains by their respective areas.

**Table 3.** The implemented  $\hat{\lambda}$  for 2019 M7.1 Ridgecrest earthquake

| Observations | Target domain | Area (km <sup>2</sup> ) | Observation density (site/km <sup>2</sup> ) | $\hat{\lambda}$ |
|--------------|---------------|-------------------------|---------------------------------------------|-----------------|
| 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            |

CISN: California Integrated Seismic Network; CSN: Community Seismic Network.

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 the most existing seismic networks. The aforementioned estimate of the overall observation density for the Exterior target region is a suggested approximation by authors to use corresponding  $\hat{\lambda}$  given in Table 1.

Table 3 demonstrates how the sites within the Exterior domain (Figure 10b) require  $\hat{\lambda} = 0.2$  for having more observations (CISN and CSN). In contrast, the same sites with fewer observations (Figure 10a) 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  $\hat{\lambda} = 0.05$ .

We 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 end is implemented to ensure that all motions start and finish at the same Universal Time Coordinated (UTC). Finally, the lowest sampling rate among all recorded motions is chosen as the target site's generated motion's sampling rate. Figures 10c and d show the distribution of the mean estimated motions' RotD50 NRMSE. The average RotD50 NRMSE for all target sites using 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 0.3 indicates a reasonably precise estimation in terms of both time series and response spectrum. However, an NRMSE larger than 0.4 demonstrates a poor estimation. Interested readers are referred to the Appendix section of Tamhidi et al. (2022a), which contains a variety of instances for predictions' NRMSE.

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 10d). Figure 10b 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 an inaccurate ground motion estimations as evidenced by Tamhidi et al. (2021). Table 4 compares the average of the mean estimated ground motions' NRMSE along each horizontal component and RotD50 spectra for various domains.

Table 4 demonstrates that additional CSN sites generally improve the generated motions' accuracy along both horizontal components for Inner domain target sites. Furthermore, Table 4 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

**Table 4.** The prediction error along EW, NS, and RotD50 response spectra in different domains for the 2019 **M7.1** Ridgecrest earthquake data set

| Domain   | Observations | EW            |                                  | NS            |                     | RotD50        |                     |
|----------|--------------|---------------|----------------------------------|---------------|---------------------|---------------|---------------------|
|          |              | Average NRMSE | Error reduction <sup>a</sup> (%) | 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          |                     |

NRMSE: normalized root mean square error; CISN: California Integrated Seismic Network; CSN: Community Seismic Network.

<sup>a</sup>Error reduction shows the reduction in the average NRMSE among all CISN target sites due to the added CSN sites.

domain. The effect of having more observations on the predictions’ error and uncertainty at different periods is investigated by Tamhidi et al. (2022a).

Three CISN target sites are chosen (Figures 10c and d) to indicate the improvement of the generated motions after adding CSN sites. Figure 11 displays the predicted motions’ RotD50 spectra and velocity time series along EW. Figures 11a and b illustrate how the amplitude of the velocity time series fits closer to the recorded one after observing additional sites from CSN. Similarly, the response spectrum of the prediction matches more precisely to the recorded one, having more observations.

2020 **M4.5** South El Monte earthquake

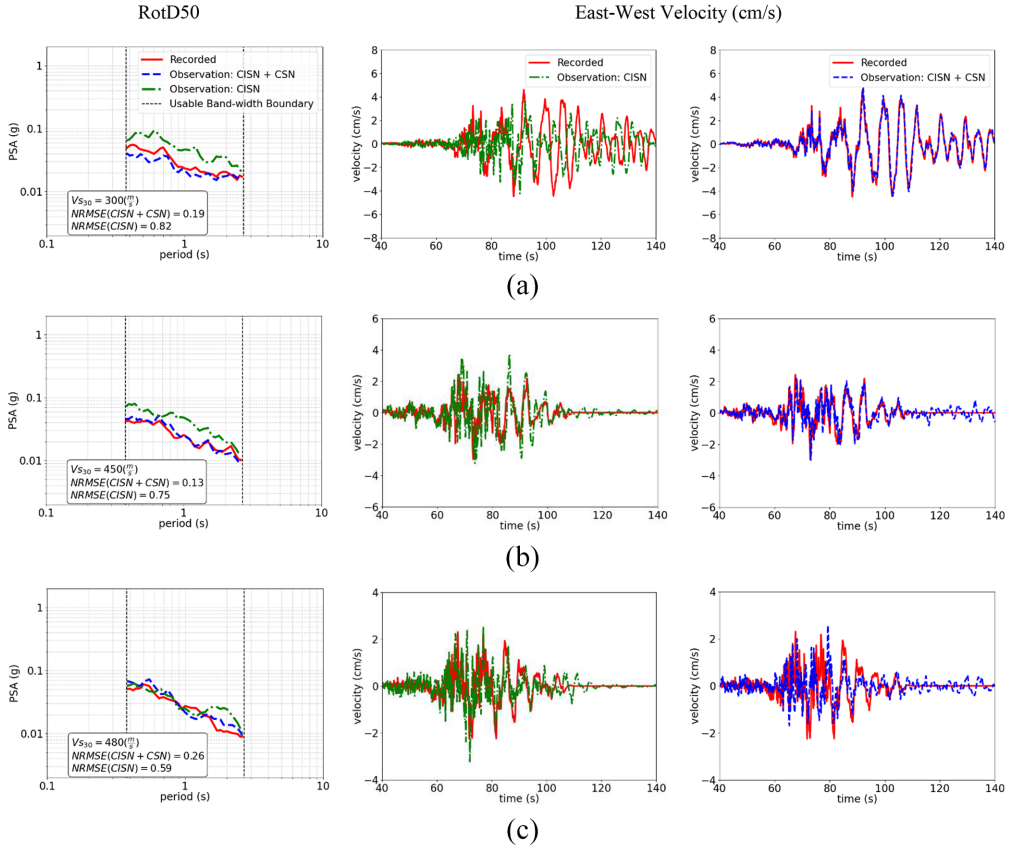
In addition, we evaluate the influence of added observations for the recently recorded ground motions of the 2020 **M4.5** South El Monte earthquake. Table 5 outlines the **M4.5** South El Monte earthquake characteristics (U.S. Geological Survey (USGS), 2020). We used 95 and 215 ground-level recording sites for CISN and CSN in Los Angeles, respectively (see Figures 12a and b). The number of sites is obtained by eliminating those with a too narrow usable bandwidth.

Table 6 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 1, respectively.

Figures 12c and d 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% (12 sites) of the target sites inside the Inner domain had an NRMSE smaller than 0.32 (Figure 12d). 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.

Table 7 compares the NRMSE of the mean estimated ground motions for each target domain. Table 7 and Figure 12 indicate that the addition of observed sites from CSN,





**Figure 11.** The RotD50 and velocity time series of the prediction using CISN and CISN plus CSN observation along EW direction for the test sites (a) No. 1, (b) No. 2, and (c) No. 3 for 2019 M7.1 Ridgecrest earthquake.

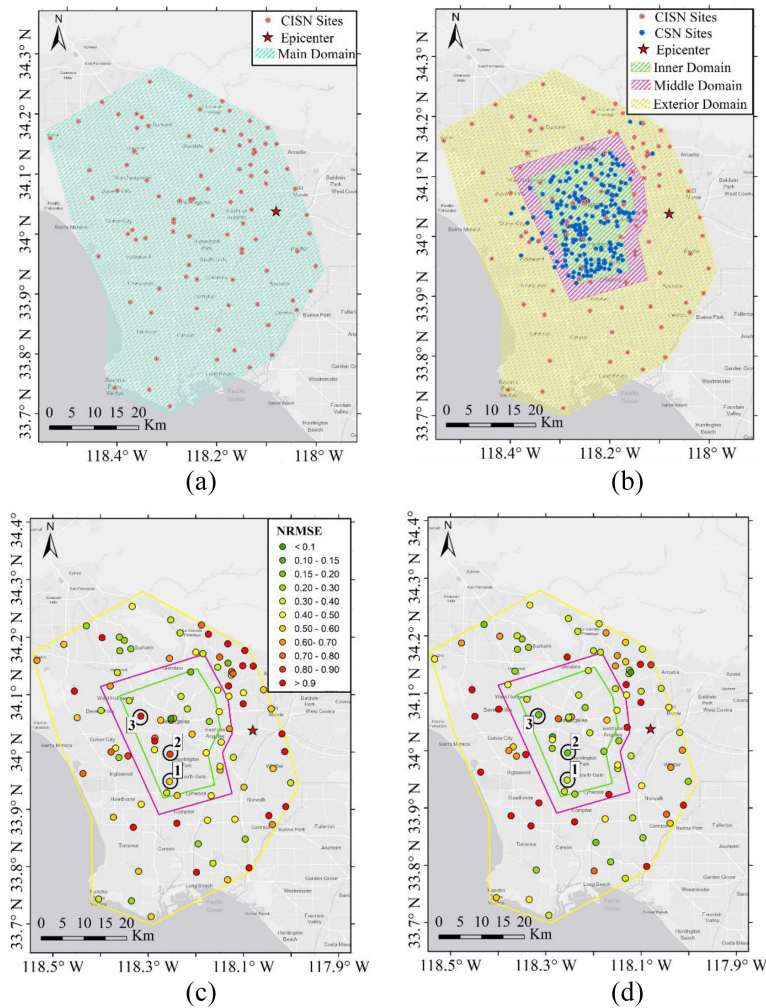
**Table 5.** The 2020 M4.5 South El Monte earthquake features (USGS, 2020)

| Date              | UTC time | $M_w$ | Epicenter      | Depth   |
|-------------------|----------|-------|----------------|---------|
| 19 September 2020 | 06:38:46 | 4.5   | South El Monte | 16.9 km |

UTC: Universal Time Coordinated.

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 12d). Comparing Table 7 and Table 4 reveals that the influence of added CSN sites for the M4.5 South El Monte earthquake is less than that for the M7.1 Ridgecrest earthquake. There are two reasons for the latter. First, the mutual usable bandwidth of the estimated motions for the South El Monte earthquake (0.11–0.55 s) is narrower and shorter than that for the Ridgecrest earthquake (0.38–2.8 s) and it is discussed that the effect of the





**Figure 12.** 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 6.** The implemented  $\hat{\lambda}$  for 2020 **M**4.5 South El Monte earthquake

| Observations | Target domain | Area (km <sup>2</sup> ) | Observation density (site/km <sup>2</sup> ) | $\hat{\lambda}$ |
|--------------|---------------|-------------------------|---------------------------------------------|-----------------|
| 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            |

CISN: California Integrated Seismic Network; CSN: Community Seismic Network.

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

**Table 7.** The prediction error along EW, NS, and RotD50 response spectra in different domains for the 2020 M4.5 South El Monte earthquake data set

| Domain   | Observations | EW            |                                  | NS            |                     | RotD50        |                     |
|----------|--------------|---------------|----------------------------------|---------------|---------------------|---------------|---------------------|
|          |              | Average NRMSE | Error reduction <sup>a</sup> (%) | 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          |                     |

NRMSE: normalized root mean square error; CISN: California Integrated Seismic Network; CSN: Community Seismic Network.

<sup>a</sup>Error reduction shows the reduction in the average NRMSE among all target sites due to the added CSN sites.

CSN observations might have a negligible effect on improving the estimations for the 2020 M4.5 South El Monte earthquake data set.

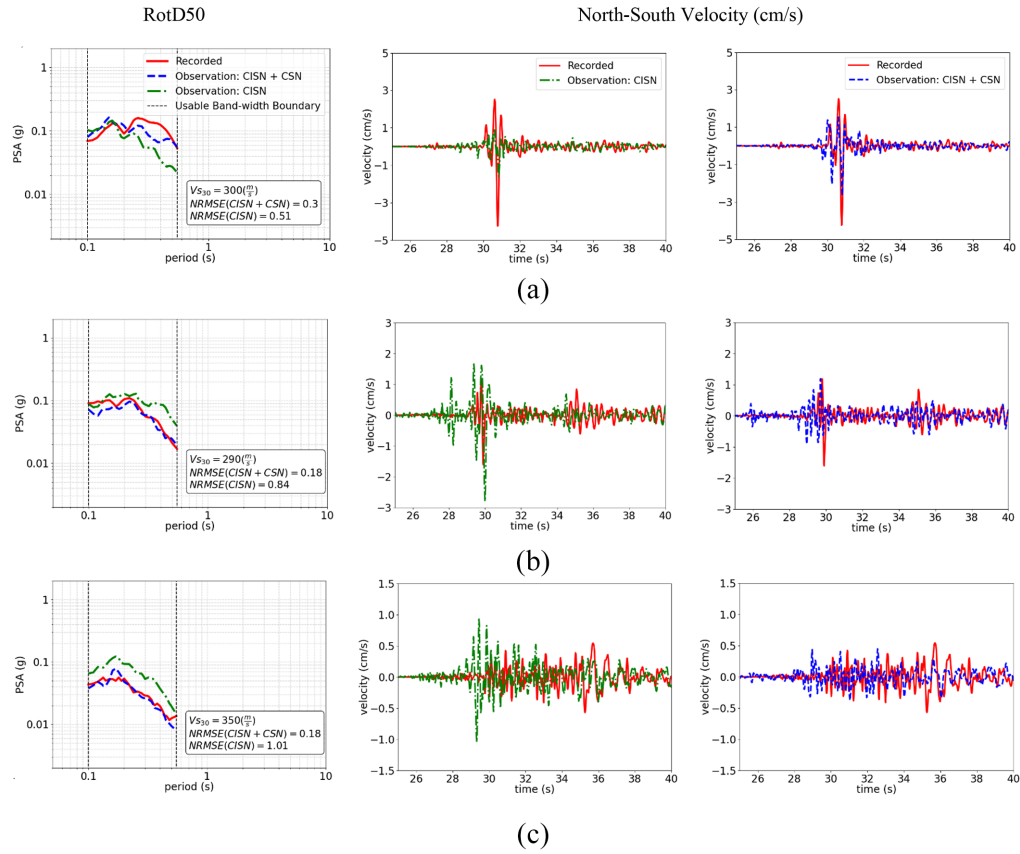
The influence of added observations is negligible for the sites within the Middle or Exterior domains and, in some cases, can worsen the estimations. It should be noted that the number of available CISN sites within the Middle domain is sparse (9 sites), which can affect the statistical inference of the added CSN observations' effect in that region.

Three CISN target sites are selected to demonstrate the estimated ground motion velocity time series using CISN-only, and CISN-plus-CSN observed sites (Figures 12c and d). Figure 13 illustrates the estimated velocity time series along the NS direction using two sets of observations. Figure 13 shows how the velocity time series for the CISN target sites inside the Inner domain fits closer to the recorded one's amplitude after the GPR model observed more CSN sites.

## Concluding remarks

This research aimed to generate post-earthquake ground motion time series at uninstrumented ("target") sites. We developed a GPR model to generate such ground motions. We explored the influence of observation spatial density (instrumented sites) on the model's optimal hyperparameter. The optimized hyperparameter allows users to implement the GPR model for various observation densities. It was demonstrated that the required regularization factor is smaller for the region with a higher observation density. In contrast, greater regularization is needed where there are fewer observations.

We also produced ground motion realizations at the target points. This approach offers an ensemble of estimated ground motion time series for uninstrumented sites to conduct the site-specific nonlinear response history analysis, which reveals the uncertainty of the predicted structural damage resulting from record-to-record variation. Quantification of uncertainty of structural responses is a future study which is under development by the authors. The uncertainty of the estimated ground motions is assessed using the generated ground motion realizations at different CSN sites using the 2019 M7.1 Ridgecrest earthquake recorded data set. It is concluded that the number of observed sites closer to the



**Figure 13.** The RotD50 and velocity time series of the prediction using C1SN and C1SN plus CSN observation along NS direction for the test sites (a) No. 1, (b) No. 2, and (c) No. 3 for 2020 **M4.5** South El Monte earthquake.

target site plays a vital role in the accuracy and uncertainty of the predictions, particularly at longer periods.

The effect of having additional observations from various recording networks on prediction accuracy was also investigated using the motions recorded during the 2019 **M7.1** Ridgecrest and 2020 **M4.5** South El Monte earthquakes. As expected, the results demonstrated that the prediction for the target sites located inside the added observed sites' borders generally improved; yet, there were a few sites with adjacent non-uniform observations that their estimations worsened after observing more ground motions. The influence of additional observations on estimations for the target sites outside the added network's boundaries was insignificant.

## Acknowledgments

The authors thank Prof. Tadahiro Kishida for his efforts in signal processing of the recorded ground motions. Prof. Chukwuebuka Nweke and Prof. Pengfei Wang have kindly provided the estimation of the  $V_{s30}$ ,  $Z_{1.0}$ , and  $Z_{2.5}$  values at the recording stations. Cooperation of Prof. Monica Kohler and Dr

Richard Guy on Community Seismic Network (CSN) data is greatly acknowledged. The comments from two Earthquake Spectra anonymous reviewers are appreciated.

## Data and Resources

The raw M7.1 2019 Ridgecrest and M4.5 2020 South El Monte earthquakes data recorded by the Community Seismic Network were obtained from <https://csn.caltech.edu/data/> (last accessed May 2022). The processed ground motions for the M7.1 2019 Ridgecrest earthquake data set recorded by CSN and CISN can be retrieved from <https://www.risksciences.ucla.edu/nhr3/gmdata> (last accessed May 2022). The ground motions for the M4.5 2020 South El Monte earthquake recorded by CISN are obtained from the Center for Engineering Strong Motion Data (CESMD) at <https://www.strong-motioncenter.org/> (last accessed May 2022). The RotD50 and orthogonal directions linear response spectra of the ground motions were constructed using the R package for computation of earthquake ground motion response spectra (Wang et al., 2017) which is accessible through <https://peer.berkeley.edu/peer-reports> (last accessed May 2022).

## Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the University of California, Los Angeles (UCLA) Graduate Fellowship to the first author, which is highly appreciated. Partial support from the National Science Foundation (award no. 2025310), California Department of Transportation, and Pacific Gas and Electric Company is also gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed are those of the authors and do not necessarily reflect those of the supporting agencies.

## ORCID iDs

Aidin Tamhidi  <https://orcid.org/0000-0002-3254-1720>

Nicolas M Kuehn  <https://orcid.org/0000-0002-3512-5300>

## References

- Aagaard BT, Brocher TM, Dolenc D, Dreger D, Graves RW, Harmsen S, Hartzell S, Larsen S, McCandless K, Nilsson S and Petersson NA (2008) Ground-motion modeling of the 1906 San Francisco earthquake, part II: Ground-motion estimates for the 1906 earthquake and scenario events. *Bulletin of the Seismological Society of America* 98(2): 1012–1046.
- Abrahamson NA, Schneider JF and Stepp JC (1991) Empirical spatial coherency functions for application to soil-structure interaction analyses. *Earthquake Spectra* 7(1): 1–27.
- Abrahamson NA, Silva WJ and Kamai R (2014) Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra* 30(3): 1025–1055.
- Aghababaei M, Koliou M, Watson M and Xiao Y (2021) Quantifying post-disaster business recovery through Bayesian methods. *Structure and Infrastructure Engineering* 17(6): 838–856.
- Ahdi SK, Mazzoni S, Kishida T, Wang P, Nweke CC, Kuehn NM, Contreras V, Rowshandel B, Stewart JP 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): 1474–1494.

- Alamilla J, Esteva L, García-Pérez J and Díaz-López O (2001) Simulating earthquake ground motion at a site, for given intensity and uncertain source location. *Journal of Seismology* 5(4): 475–485.
- Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, Wooddell KE, Graves RW, Kottke AR, Boore DM and Kishida T (2014) NGA-West2 database. *Earthquake Spectra* 30(3): 989–1005.
- Atkinson GM 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): 48–60.
- Bayless J and Abrahamson NA (2019) An empirical model for the interfrequency correlation of epsilon for Fourier amplitude spectra. *Bulletin of the Seismological Society of America* 109(3): 1058–1070.
- Boore DM (2010) Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bulletin of the Seismological Society of America* 100(4): 1830–1835.
- Boore DM, Stewart JP, Seyhan E and Atkinson GM (2014) NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra* 30(3): 1057–1085.
- Campbell KW 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): 1087–1115.
- Clayton RW, Kohler M, Guy R, Bunn J, Heaton T and Chandy M (2020) CSN-LAUDS network: A dense accelerometer network in Los Angeles Schools. *Seismological Research Letters* 91(2A): 622–630.
- Cui XZ and Hong HP (2020) Conditional simulation of spatially varying multicomponent nonstationary ground motions: Bias and ill condition. *Journal of Engineering Mechanics* 146(2): 04019129.
- Deodatis G (1996) Non-stationary stochastic vector processes: Seismic ground motion applications. *Probabilistic Engineering Mechanics* 11(3): 149–167.
- Fraser WA, Wald DJ and Lin KW (2008) Using shakemap and shakecast to prioritize post-earthquake dam inspections. In Manzari MT and Hiltunen DR (eds) *Geotechnical Earthquake Engineering and Soil Dynamics IV*. Reston, VA: American Society of Civil Engineers, pp. 1–10.
- Heredia-Zavoni E and Santa-Cruz S (2000) Conditional simulation of a class of nonstationary space-time random fields. *Journal of Engineering Mechanics* 126(4): 398–404.
- Hu L, Xu YL and Zheng Y (2012) Conditional simulation of spatially variable seismic ground motions based on evolutionary spectra. *Earthquake Engineering & Structural Dynamics* 41(15): 2125–2139.
- Kameda H and Morikawa H (1992) An interpolating stochastic process for simulation of conditional random fields. *Probabilistic Engineering Mechanics* 7(4): 243–254.
- Kohler MD, Filippitzi F, Heaton T, Clayton RW, Guy R, Bunn J and Chandy KM (2020) 2019 Ridgecrest earthquake reveals areas of Los Angeles that amplify shaking of high-rises. *Seismological Society of America* 91(6): 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): 495–513.
- Li R and Sudjianto A (2005) Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. *Technometrics* 47(2): 111–120.
- Lin K, Wald DJ, Kircher CA, Slosky D, Jaiswal K and Luco N (2018) USGS shakecast system advancements. In *11th National Conference on Earthquake Engineering*, 2018 June, EERI pp. 3458–3468.
- Loos S, Lallemand D, Baker J, McCaughey J, Yun SH, 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): 1695–1718.

- Mangalathu S and Jeon JS (2020) Ground motion-dependent rapid damage assessment of structures based on wavelet transform and image analysis techniques. *Journal of Structural Engineering* 146(11): 04020230.
- Nweke CC, Wang P, Brandenburg SJ and Stewart JP (2018) Reconsidering basin effects in ergodic site response models. UCLA. Retrieved from <https://escholarship.org/uc/item/6048v74k>
- Oppenheim AV, Willsky AS and Nawab SH (1997) *Signals and Systems*. Prentice hall Inc., Upper Saddle River, NJ.
- Rasmussen CE and Williams CKI (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.
- Rodda GK and Basu D (2018) Coherency model for translational and rotational ground motions. *Bulletin of Earthquake Engineering* 16(7): 2687–2710.
- Rodgers AJ, Anders Petersson N, Pitarka A, McCallen DB, 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): 1268–1284.
- Roohi M and Hernandez EM (2020) Performance-based post-earthquake decision making for instrumented buildings. *Journal of Civil Structural Health Monitoring* 10(5): 775–792.
- Shinozuka M, Deodatis G (1996) Simulation of multi-dimensional Gaussian stochastic fields by spectral representation. *Appl Mech Rev* 49 (1): 29–53. <https://doi.org/10.1115/1.3101883>
- Southern California Earthquake Data Center (2022) Available at: <https://service.scedc.caltech.edu/SCSNStationMap/station.html> (accessed May 2022).
- Tamhidi A, Kuehn NM and Bozorgnia Y (2022a) *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, Ghahari SF, Rodgers AJ, Kohler MD, 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): 331–347. DOI: 10.1785/0120210054.
- Tamhidi A, Kuehn N, Ghahari SF, Rodgers AJ, Taciroglu E and Bozorgnia Y, (2022b) Earthquake ground motion conditioned simulation using sparsely distributed observed motions for analysis and design of lifeline structures. *Proceedings of Lifelines 2022 Conference, Los Angeles, CA, February, Reston, VA: American Society of Civil Engineers.*
- Tamhidi A, Kuehn NM, Kohler MD, Ghahari F, Taciroglu E and Bozorgnia Y (2020) 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, September 2020.*
- Thomson EM, Bradley BA and Lee RL (2020) Methodology and computational implementation of a New Zealand Velocity Model (NZVM2. 0) for broadband ground motion simulation. *New Zealand Journal of Geology and Geophysics* 63(1): 110–127.
- U.S. Geological Survey (2020). M4.5-3km WSW of South El Monte, CA. Available at: <https://earthquake.usgs.gov/earthquakes/eventpage/ci38695658/executive> (accessed May 2022.).
- Vanmarcke EH and Fenton GA (1991) Conditioned simulation of local fields of earthquake ground motion. *Structural Safety* 10(1–3): 247–264.
- 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): 1413–1432.
- Wald D, Lin KW, 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): 533–553.
- Wang P, Stewart JP, Bozorgnia Y, Boore DM and Kishida T (2017) *R package for computation of earthquake ground motion response spectra*. Pacific Earthquake Engineering Center, Report 2017/09.
- Weatherill GA, Silva V, Crowley H and Bazzurro P (2015) Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bulletin of Earthquake Engineering* 13(4): 957–981.

- Wen YK, Ellingwood BR, Veneziano D, Bracci J. (January, 2003) *Uncertainty modeling in earthquake engineering*. Mid-America Earthquake Center Project FD-2 report
- Worden CB, Thompson EM, Baker JW, Bradley BA, Luco N and Wald DJ (2018) Spatial and spectral interpolation of ground-motion intensity measure observations. *Bulletin of the Seismological Society of America* 108(2): 866–875.
- Wu Y, Gao Y and Li D (2015) Error assessment of multivariate random processes simulated by a conditional-simulation method. *Journal of Engineering Mechanics* 141(5): 04014155.
- Yazdi M, Motamed R and Anderson JG (2022) A new set of automated methodologies for estimating site fundamental frequency and its uncertainty using horizontal-to-vertical spectral ratio curves. *Seismological Society of America* 93(3): 1721–1736.
- 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): 37–48.