Title
Geo-spatial Learning and Modeling for Seismic Site Responses in Los Angeles County

Author
Wang, Pengfei

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Geo-spatial Learning and Modeling for Seismic Site Responses in Los Angeles County

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Pengfei Wang

2020

ABSTRACT OF THE THESIS

Geo-spatial Learning and Modeling for Seismic Site Responses in Los Angeles County

by

Pengfei Wang

Master of Science in Statistics

University of California, Los Angeles, 2020

Professor Frederic R Paik Schoenberg, Chair

Earthquakes in seismically-active regions, such as in California, present a significant human and financial risk to communities. Ground motion models (GMMs) have been developed to account for earthquake impacts when infrastructures are designed. However, GMMs assume ergodic hypothesis that is ground motions behave the same globally over time. The assumption was made as the global data had to be combined for modeling due to very limited available data. As more and more data collected, it was realized that spatial variations of ground motions are too large to be neglected. This study proposes a Bayesian hierarchical model to extract seismic site terms, the spatial site response bias from ergodic GMMs, to develop non-ergodic seismic site responses. The model was then implemented on the data from earthquake stations in Los Angeles County. The Kriging prediction is also conducted to generate heat map for seismic site responses visualization.

The thesis of Pengfei Wang is approved.

Mark Stephen Handcook

Nicolas Christou

Frederic R Paik Schoenberg, Committee Chair

University of California, Los Angeles

2020

To my family.

For their endless love, support and encouragement

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

Funding for this study was provided by the California Strong Motion Instrumentation Program, California Geological Survey, under Agreement Number 1016-985. Partial support was also provided by the US Geological Survey under contract number G17AP00018. This support is gratefully acknowledged. The work presented here represents the views and opinions of the authors and does not reflect the policy, expressed or implied, of the State of California or the US Government. The support of lecture notes from Dr. Nicolas Christou is also greatly acknowledged.

VITA

2011–2015 B.S. (Traffic Engineering), Tongji University, Shanghai, China.

2015–2016 M.S. (Civil Engineering), UCLA, Los Angeles, California.

2016– Ph.D. Candidate (Civil Engineering), UCLA, Los Angeles, California.

PUBLICATIONS

Wang, P., Zimmaro, P., Ahdi, S.K., Kwak, D.Y., Stewart, J.P.. (2020). Shear Wave Velocity Database and Its Application for Analysis of Non-Ergodic Site Amplification Effects. *USGS*. Report No. G17AP00018.

Brandenberg, S.J., Stewart, J.P., Wang, P., Nweke, C.C., Hudson, K., Goulet, C.A., Meng, X., Davis, C.A., Ahdi, S.K., Hudson, M.B., Donnellan, A., Lyzenga, G., Pierce, M., Wang, J., Winters, M.A., Delisle, M.-P., Lucey, J., Kim, Y., Gallien, T.W., Lyda, A., Yeung, J.S., Issa, O., Buckreis, T., Yi, Z.. (2020). Ground Deformation Data from GEER Investigations of Ridgecrest Earthquake Sequence. *Seismological Research Letters*. https://doi.org/10.1785/0220190291.

Wang, P., Stewart, J.P.. (2019). Data-Derived Site Response and its Predictability Using Ergodic and Site-Specific Methods. *Proceeding of SMIP2019 Seminar on Utilization of Strong Motion Data*, California Strong Motion Instrumentation Program, University of California Los Angeles, California. October 18, 2019. https://escholarship.org/uc/item/3wb9h9fq.

Stewart, J.P., Wang, P., Teague, D.P., Vecchietti, A.. (2019). Applications of non-ergodic

site response in ground motion modeling. Proceeding of 7th International Conference on Earthquake Geotechnical Engineering (Invited Keynote), Rome, Italy. 07/17-20/2019. https://escholarship.org/uc/item/5427j7f3.

Wang, P., Stewart, J. P., Bozorgina, Y., Boore, D. M., Kishida, T. (2017). R Package for Computation of Earthquake Ground Motion Response Spectra. Report No. 2017/09. *PEER*, UC Berkeley.

Zheng, N., Dantsuji, T., Wang, P., Geroliminis, N., (2017). Macroscopic Approach for Optimizing Road Space Allocation of Bus Lanes in Multimodal Urban Networks Through Simulation Analysis. *Journal of the Transportation Research Board*, No. 2651, DOI:10.3141/2651-05.

CHAPTER 1

Introduction

1.1 Motivation

Earthquakes as one of the most ruinous hazards have caused tremendous damage, especially in the areas where there are many active faults, like California. However, there is still not an effective way to forecast when and where a big earthquake could happen. To mitigate the damage of earthquakes on infrastructures, structural engineers account for earthquake effects when designing structures by using ground motion models (GMMs). GMMs were developed to predict ground motion intensity measures given parameters descriptive of source, path, and site conditions. These GMMs incorporate source, path, and site response models that represent approximately the average conditions in the database from which the GMMs were derived. In the most commonly used NGA-type GMMs [1, 3, 4, 5], global earthquake data is used so the predictions represent global averages. In contrast, when GMMs are applied for a specific engineering project, the source, path, and site response attributes of interest are those local to the site, which may depart from the global averages represented by the GMM. In this context, I refer to the source, path, and site models in the GMM as spatially ergodic. Alternative models that consider local, or site-specific features, are considered spatially *non-ergodic*, and have the potential to significantly reduce the ground motion bias and variability.

The most accepted method to correct ergodic GMM to be spatially non-ergodic model is using the actual earthquake observation data from the sites to extract the site-specific bias, denoted as site term (η_S). Thus, the site-specific GMM will be the sum of ergodic GMM and site terms and then the site-specific site response will be the sum of site response from

ergodic GMM and site terms. The traditional approach to extract the site terms is utilizing fixed effect model on total residuals. This approach is widely used as it is unbiased estimator of site terms and is easy to implement and interpret. However, the approach is not robust to outliers and it could overfit for some sites where the amount of available earthquake data is limited. Therefore, an alternative model is needed.

Furthermore, there is a last problem needs to be resolved before implementing site-specific site response into the design. The earthquake observation ground motions data are recorded at strong motion stations distantly located. For example, in California where the density of stations is the highest in the U.S., there are around 2,000 stations mainly locating along the coast and clustering around San Francisco and Los Angeles. With the highest density of stations, in Los Angeles County, however, we are still unable to get a reasonably high resolution of site-specific site response. This is because the distance between two stations is generally a few kilometers away. To generate a higher resolution map of site-specific site response, a prediction model (or interpolation) is imperative.

1.2 Organization

This study seeks to propose a more robust method to calculate seismic site terms and site-specific site responses, and then generate a higher resolution map of site terms for Los Angeles County. The main body of the thesis consists of 3 chapters.

The Chapter 2 first introduces the commonly used earthquake intensity measure (the ground motion metric), pseudo spectra acceleration, as it is the prediction result of GMM. And then it describes the ergodic GMM and its prediction of site response. Additionally, it also presents the earthquake ground motion data set for Los Angeles County and will be used in this study.

The Chapter 3 describes the traditional approach (Frequentist) and robust approach (Bayesian) to extract site terms from total residuals of GMM. And then both methods are implemented on the data set in Los Angeles County to calculate site terms. The difference between the two methods will then be illustrated and discussed.

The Chapter 4 first describes the basic theory of Kriging. The author then describes how to implement Kriging to construct site-specific site terms and site responses for Los Angeles County. The heat map of site terms is also generated.

Finally, in the last Chapter, it concludes findings and limitations.

CHAPTER 2

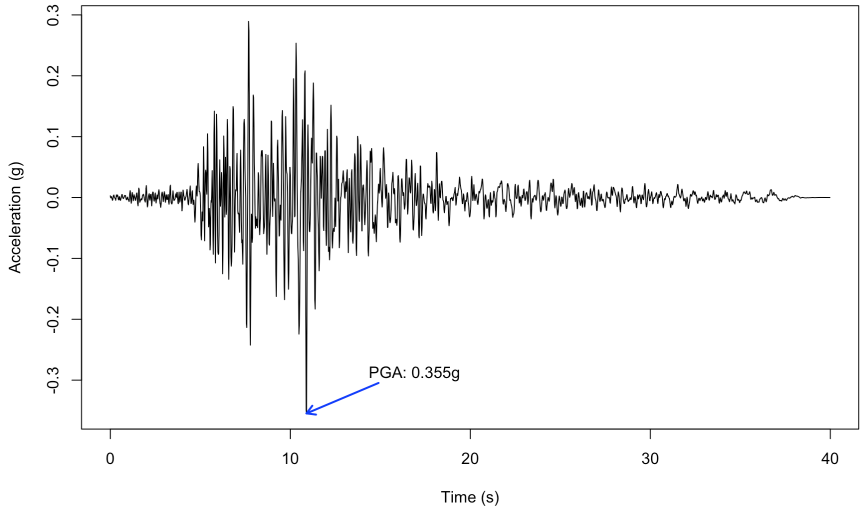
Ergodic GMM and Data Source

In this chapter, I first introduce a most commonly used earthquake ground motion intensity measure, pseudo spectra acceleration, as it reflects more real infrastructures response and also it is the prediction result of GMM. And then I will present one ergodic GMM and its prediction for site response. The data set of ground motion used in this study is also presented.

2.1 Pseudo Spectra Acceleration

The actual observed ground motions are in time series format, as one example shown in Fig. 2.1. This time series is from 1994 Northridge earthquake (Magnitude is 6.69) at Los Angeles Obregon Park strong motion station. For ease of description and modeling use, many intensity measures were proposed to numerically express how strong the shaking is. Peak Ground Acceleration (PGA) is one of the most common measures. PGA is defined as the maximum absolute ground motion acceleration. As it is shown in Fig. 2.1, the PGA of the time series is 0.355 g. g is the unit of gravity ($\sim 9.8\ m/s^2$), so the PGA is about $0.355 \times 9.8 \approx 3.5 m/s^2$ which is a very strong shaking. This is because the Obregon Park station is only 40 kilometers from the epicenter.

Figure 2.1: One Horizontal Time Series Recorded at Los Angeles Obregon Park Station from 1994 Northridge Earthquake



It is no doubt that PGA is widely used intensity measure, however, PGA cannot tell distinctions of different buildings responses under the same ground motions. The reason is that PGA is the metric objectively describes how strong the ground motion is, it does not consider any effect from the buildings themselves. This is why another intensity measure was came up that is pseudo spectra acceleration (PSA). It was proposed from the assumption that the responses of buildings are governed by Single-degree-of-freedom (SDF) System.

Figure 2.2: The schematic of SDF system [6]

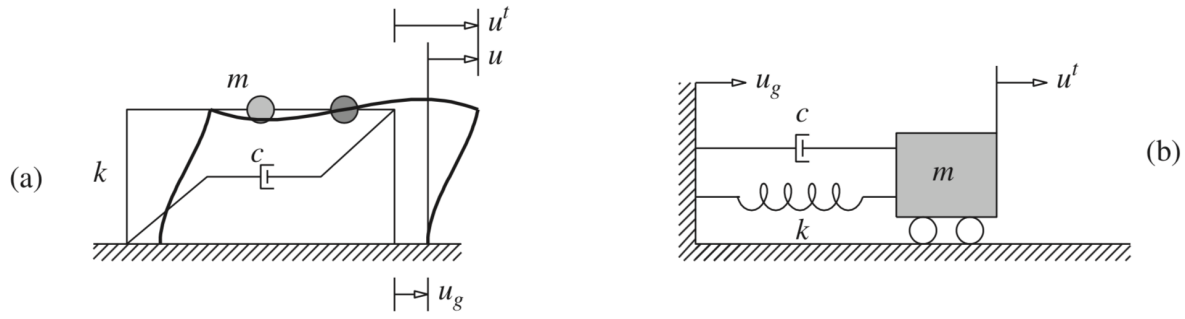


Fig. 2.2 (a) shows the response of a simplified building to ground motion u_g. The de-

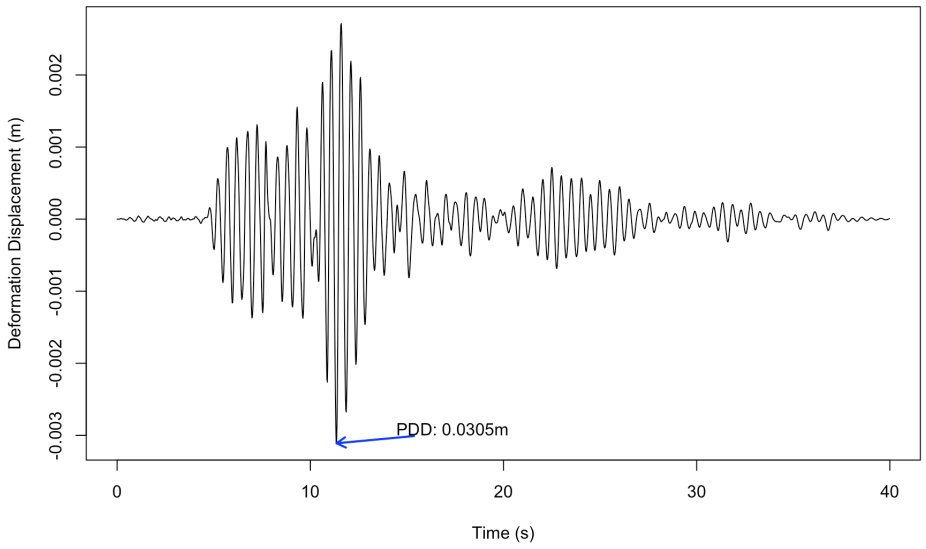
formation displacement (response) of top respect to bottom of the building is u while the absolute displacement of top respect to previous position before shaking is u^t. The mass, stiffness, and damping coefficient are denoted by m, k, and c. This response can be expressed by Fig. 2.2 (b), the schematic of SDF system. By Newton law, the motion equation is,

$$\ddot{u} + 2\,\xi\,\omega_n\,\dot{u} + \omega_n^2\,u = -\ddot{u}_g(t) \tag{2.1}$$

where damping ratio $\xi = \frac{c}{2\sqrt{km}}$ and natural frequency $\omega_n = \sqrt{k/m} = \frac{2\pi}{T_n}$. Then the deformation displacement can be written as, $u \equiv u(t, T_n, \xi)$. Usually, we set $\xi = 0.05$ as default for typical structure design. Therefore, the deformation displacements of different buildings will be fully determined by their natural periods T_n. Generally speaking, shorter buildings are stiffer and the natural periods are smaller. On the other hand, the taller buildings would have longer natural periods (up to a few periods).

Given a natural period of interest, by using Eq. 2.1, the deformation displacement can be calculated (using the R package [14]). For example, suppose $T_n = 0.5$ sec, given the same ground motion as shown in 2.1, the deformation displacement (along time) can be calculated and plotted in Fig. 2.3.

Figure 2.3: Deformation Response with Natural Period $T_n = 0.5$ sec to Ground Motion from Fig. 2.1

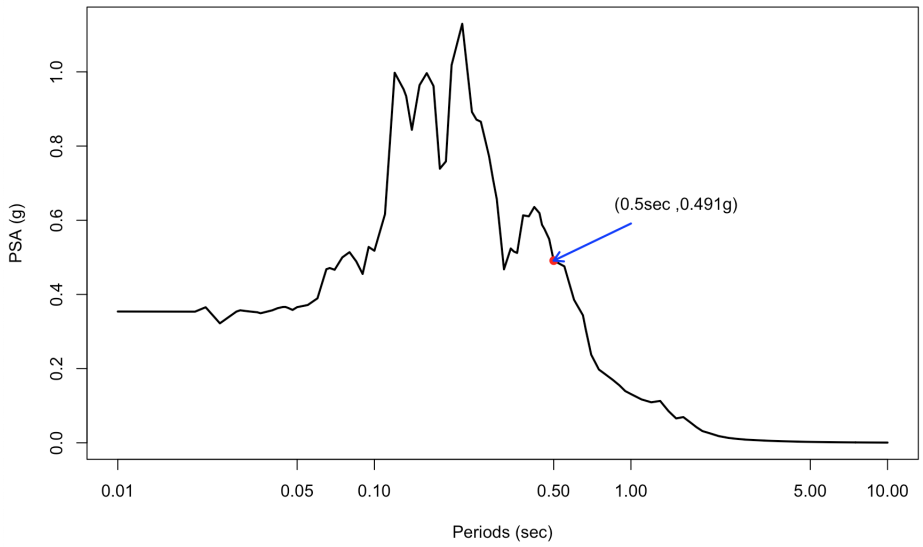


Note PDD stands for Peak Deformation Displacement in the same manner as PGA, the maximum absolute of the values. PDD can also be transformed into acceleration by the following equation,

$$S_a(T_n) = \omega_n^2 \, \text{PDD} \tag{2.2}$$

where $S_a(T_n)$ is named as pseudo spectra acceleration (PSA). Spectra is referring to periods and pseudo comes from the conversion from PDD (as the real spectra acceleration should be derived directly from Eq. 2.1). If a sequence of natural periods are inputted, then a sequence of S_a will be calculated shown in Fig. 2.4. The red point is corresponding to natural period $T_n = 0.5$ sec, the example in Fig. 2.3.

Figure 2.4: PSA of Ground Motion from Fig. 2.1

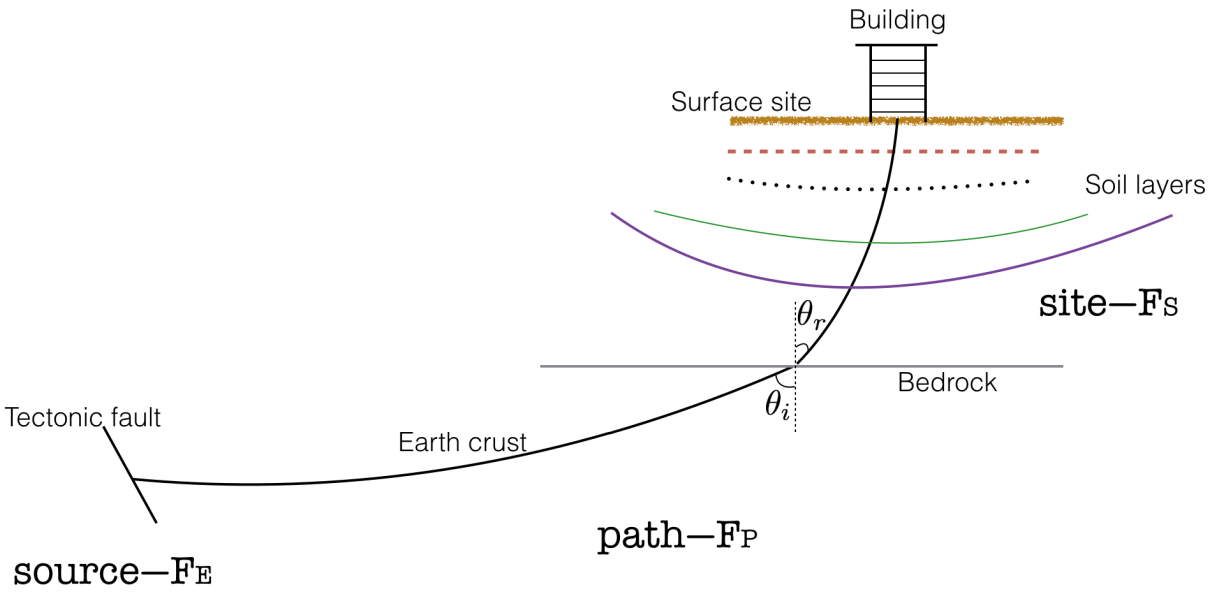


Then given the PSA curve of any ground motion and natural period of interest, structural engineers will be able to estimate the building response and potential damage very easily. Due to the valuable advantage, PSA is the most important intensity measure used in earthquake engineering and selected as the intensity measure in GMMs.

2.2 Ergodic GMM

GMM is an empirical model to numerically describe the source, path, and site response effects of ground motion. Fig. 2.5 is a schematic to show the whole process that the ground shaking has been transmitted from earthquake epicenter through earth crust and relatively soft soil layers to the surface site of our building. Then in GMM, it describes each of these three effects.

Figure 2.5: The schematic of ground motion generation and propagation



You may notice that the earthquake is generated along the tectonic fault and then transmitted within deep earth crust not shallow soils. Actually the wave is transmitted everywhere, however, the stronger motions to our site are the ones with smaller energy loss which comes through relatively harder and less damping deep earth crust. This is the reason why only this path is showed in Fig. 2.5. Another note is that the wave gradually becomes vertical as the wave propagates up from bedrock. This behavior can be explained by Snell's Law that the angle of refraction, θ_r, becomes smaller than the angle of incidence, θ_i, if the velocity in the second medium is smaller. The wave velocity in the upper softer soil is indeed

slower than the deeper harder soil, then the wave propagation becomes gradually vertical. Therefore, the source effect mainly focuses on the tectonic fault, path effect describes the transmission within deep earth crust, and the site response accounts for the wave propagation from bedrock to the surface. You can imagine the site response should not depend too much on where and how earthquake happens as it is the local site behavior and could be well-predictable under well studies. This is hypothetical reason of this study.

2.2.1 Structure of Ergodic GMM

There are five commonly used GMMs [1, 3, 4, 5, 8] for crustal earthquakes which were developed based on NGA-West2 project database [2]. Their model structures are the same, while some detailed terms differ. BSSA14 model [3] is selected to illustrate the ergodic GMM and its prediction for site response, and will be used in this study. The full BSSA14 model can be expressed as

$$\ln Y = F_E(\mathbf{M}, mech) + F_P(R_{JB}, \mathbf{M}, region) + F_S(V_{S30}, R_{JB}, \mathbf{M}, region, z_1) + \varepsilon_n \sigma(\mathbf{M}, R_{JB}, V_{S30}) \quad (2.3)$$

where $\ln Y$ is the natural log of PSA at a period of interest; F_E is the source function dependent on earthquake magnitude $\mathbf{M}$ and earthquake focal mechanism $mech$; F_P is the path function determined by the distance R_{JB}, $region$ (as the earth curst hardness varies in the world), and earthquake magnitude $\mathbf{M}$ (as the frequency contents of different magnitude earthquakes differ and then wave transmissions vary); F_S is the site response function (explained in detail below), σ is the total standard deviation, and ε_n is the number of standard deviations of a predicted value away from the mean.

2.2.2 Ergodic Site Response

In this ergodic model, the site response F_S is a function of time-averaged shear wave velocity at the top 30 meters V_{S30}, distance R_{JB}, earthquake magnitude $\mathbf{M}$, site region $region$, and depth that is used to represent basin effect z_1. V_{S30} and z_1 are the variables that roughly describes the stiffness and thickness of soil layer. The combination of R_{JB}, $\mathbf{M}$, and $region$ will account for the effects of site response from ground shaking. This is because soil could

be linear or nonlinear behaviors under weak or strong shaking.

Then we can partition the whole ergodic site response F_S into three parts, linear component of site amplification $\ln(F_{lin})$, nonlinear component of site amplification $\ln(F_{nl})$, and the effect of basin depth $F_{\delta z_1}(\delta z_1)$,

$$F_S(V_{S30}, R_{JB}, \mathbf{M}, region, z_1) = \ln(F_{lin}) + \ln(F_{nl}) + F_{\delta z_1}(\delta z_1) \tag{2.4}$$

the coefficients of this function parameters are period-dependent (for PSA) and are regressed based on global database, the detailed approach is discussed in [3].

The linear site amplification and nonlinear site amplification were originally proposed by [12]. The linear site amplification F_{lin} describes the scaling of ground motion with V_{S30} for linear soil response conditions, typically with small strains under weak ground motions. The natural logarithm of F_{lin} is expressed as

$$\ln(F_{lin}) = \begin{cases} c\ln(\frac{V_{S30}}{V_{ref}}) & V_{S30} \leq V_c \\ c\ln(\frac{V_c}{V_{ref}}) & V_{S30} > V_c \end{cases} \tag{2.5}$$

where c is the V_{S30} scaling, V_c is the maximum velocity beyond which ground motions no longer scale with V_{S30}, and V_{ref} is a reference velocity (760 m/s). Parameter c and V_c are period-dependent and given by BSSA14.

The nonlinear site amplification F_{nl} in natural log is

$$\ln(F_{nl}) = f_1 + f_2\ln(\frac{PGA_r + f_3}{f_3}) \tag{2.6}$$

where f_1, f_2, and f_3 are model coefficients, and PGA_r is the median peak horizontal acceleration for reference rock where $V_{S30} = 760m/s$. The model takes $f_1 = 0$ to force $\ln(F_{nl})$ to 0 for $PGA_r \ll f_3$. f_3 is a transition intensity measure (IM) between linear behavior (lower than f_3) and linear decrease at rate of $f_2\ln(IM)$ (higher than f_3), it is taken as $f_3 = 0.1g$. f_2 is the degree of nonlinearity as a function of V_{S30},

$$f_2 = f_4[\exp\{f_5(\min(V_{S30}, 760) - 360)\} - \exp\{f_5(760 - 360)\} \tag{2.7}$$

where f_4 and f_5 are period-dependent coefficients, also given by BSSA14.

And the last part of site amplification is basin term $F_{\delta z_1}$, it is formulated as

$$F_{\delta z_1}(\delta z_1) = \begin{cases} 0 & T < 0.65 \\ f_6 \delta z_1 & T \geq 0.65 \& \delta z_1 \leq f_7/f_6 \\ f_7 & T \geq 0.65 \& \delta z_1 > f_7/f_6 \end{cases} \tag{2.8}$$

where f_6 and f_7 are model coefficients, provided by BSSA14. δz_1 (in km) is computed as

$$\delta z_1 = z_1 - \mu_{z_1}(V_{S30}) \tag{2.9}$$

where $\mu_{z_1}(V_{S30})$ is a function of V_{S30} relating to z_1. For California, it can be computed as

$$\ln(\mu_{z_1}) = \frac{-7.15}{4} \ln(\frac{V_{S30}^4 + 570.94^4}{1360^4 + 570.94^4}) - \ln(1000) \tag{2.10}$$

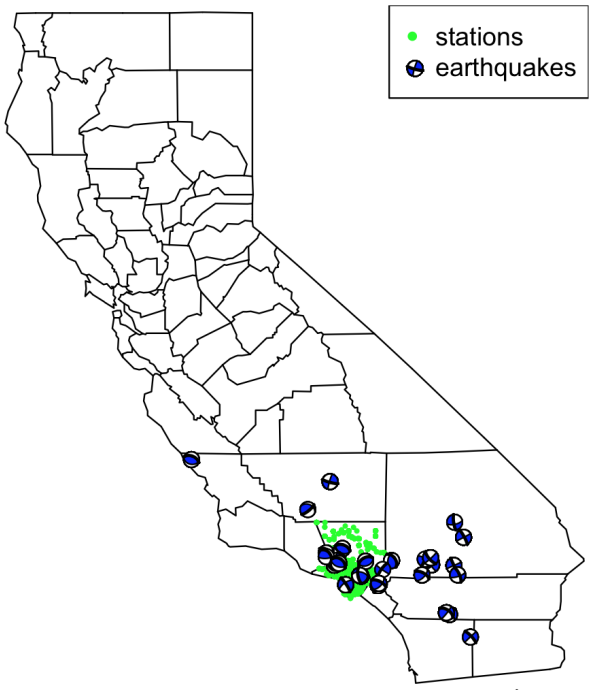
where μ_{z_1} is also in km. Therefore, given V_{S30}, z_1, and PGA_r, the ergodic site response F_S can be estimated.

2.3 Data Source

The dataset used in this study is from Pacific Earthquake Engineering Research (PEER) center [1]. It contains the strong ground motion recordings since 1930s until 2011 from all over the world. There are 21540 processed and calculated PSAs in total. Since this study focuses on Los Angeles County, the data outside of Los Angeles County are filtered out. The filter results in 2608 records in Los Angeles County which are collected by 344 unique stations and from 134 earthquake events. The earthquakes (blue-white beachballs) and stations (green dots) of the data are plotted below. The beachball is a convenient way to show focal mechanisms (fault strike direction).

[1] http://peer.berkeley.edu/ngawest2/databases/

Figure 2.6: Spatial distribution of earthquakes and stations



CHAPTER 3

Frequentist and Bayesian Models for Site Terms

When GMMs are applied for a specific engineering project, the source, path, and site response effects of interest are those local to the site, which will depart from the global averages represented by the ergodic GMM. The systematic bias of site response effect between ergodic site response from GMM at the specific site and the actual site response will be called site term (denoted by η_S). The same to event term, the systematic bias of source effect between ergodic source from GMM in a specific event and the actual source effect (denoted by η_E). Therefore, the site-specific site response will be the sum of ergodic site response from GMM and the site term (the site systematic bias), $\tilde{F_S} = F_S + \eta_S$. Then the goal in this chapter is to extract site term from residuals.

From residual analysis, we have the following equation,

$$\ln Y_{ij} - \ln \hat{Y}_{ij} = R_{ij} = \eta_{E_i} + \delta W_{ij} = \eta_{E_i} + \eta_{S_j} + e_{ij} \tag{3.1}$$

where i represents event i, j stands for station or site j; $\ln Y_{ij}$ is the natural log of actual observed ground motion and $\ln \hat{Y}_{ij}$ is the natural log of predicted ground motion from GMM; R_{ij} is total residual from deviation of observation and GMM, δW_{ij} is within-event residuals, and e_{ij} is the random noise. To extract the site terms η_{S_j}, event terms η_{E_i} should be removed from total residuals. How to estimate event terms and site terms? There are two methods: Frequentist versus Bayesian.

3.1 Frequentist Approach

From Frequentist statistical point of view, Equation 3.1 is an additive model. Event terms η_{E_i} and site terms η_{S_j} are two predictors. Based on two assumptions,

- within-event residuals δW_{ij} independently follows multivariate normal distribution $\mathcal{N}(0, \phi^2)$;
- random noise e_{ij} also independently follows multivariate normal distribution $\mathcal{N}(0, \sigma_e^2)$

Then, either applying least square or maximum likelihood estimation, we can have the following results showing in the Figure 3.1.

Figure 3.1: Frequentist additive model

Event terms

$$R_{ij} = \eta_{E_i} + \delta W_{ij}$$

Site terms

$$R_{ij} - \eta_{E_i} = \delta W_{ij} = \eta_{S_j} + e_{ij}$$

$$\mathbf{R} = \begin{bmatrix} r_{1,1} \\ r_{1,2} \\ \vdots \\ r_{1,n_1} \\ r_{2,1} \\ \vdots \\ r_{i,n_i} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \eta_{E_1} \\ \eta_{E_2} \\ \vdots \\ \eta_{E_i} \end{bmatrix} + \begin{bmatrix} \delta w_{1,1} \\ \delta w_{1,2} \\ \vdots \\ \delta w_{1,n_1} \\ \delta w_{2,1} \\ \vdots \\ \delta w_{i,n_i} \end{bmatrix} \quad \delta \mathbf{W} = \begin{bmatrix} \delta w_{1,1} \\ \delta w_{1,2} \\ \vdots \\ \delta w_{1,k_1} \\ \delta w_{2,1} \\ \vdots \\ \delta w_{j,k_j} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \eta_{S_1} \\ \eta_{S_2} \\ \vdots \\ \eta_{S_j} \end{bmatrix} + \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ \vdots \\ e_{1,k_1} \\ e_{2,1} \\ \vdots \\ e_{j,k_j} \end{bmatrix}$$

$$\eta_{E_i} = \frac{\sum_{j=1}^{j=n_i} r_{i,j}}{n_i} \qquad \eta_{S_j} = \frac{\sum_{i=1}^{i=k_j} \delta w_{i,j}}{k_j}$$

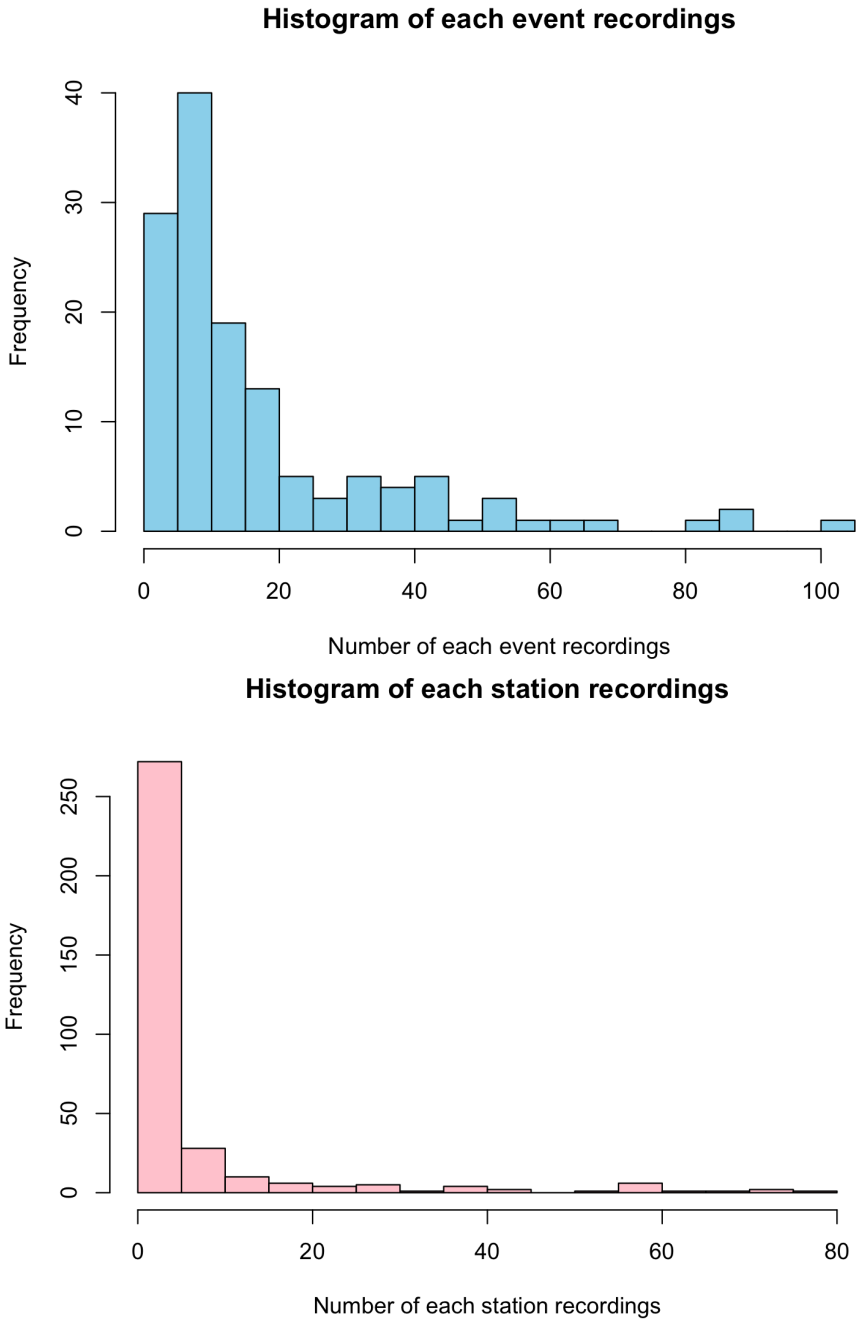
From Figure 3.1, the estimators based on Frequentist additive model are the averages. For the ith event term η_{E_i}, it is the average of total residuals for the recordings from event i; and the jth site term η_{S_j}, it is the average of within-event residuals (total residuals removed event terms) for the recordings from station j.

3.2 Bayesian Hierarchical Approach

The estimators of Frequentist approach are not robust but very sensitive to outliers when the dataset is relatively small. The histograms of number of recordings within each event

and each station are showed in Figure 3.2.

Figure 3.2: Histograms of number of recordings



The majority of events and stations have less than 10 or even less than 5 recordings so that using means is not very appropriate. In this situation, Bayesian hierarchical approach is superior. Stewart et al [13] gave the analytic solutions to event terms and site terms,

however, they did not show mathematical reasoning and explain in detail. Here, according to Gelman et al [7], I will give more detailed explanations.

Bayesian approach is more preferable here is because there is a prior distribution which can be trained by using the entire dataset. The total dataset is large enough to have a relative stable solutions although each individual event and station does not. Then event terms and site terms will be estimated by both prior distribution and their own individual datasets. In another word, although each individual event terms and site terms are independent of each other, they help each other to improve estimations from connection of prior distribution.

First, there are two prior distributions for event terms and site terms, respectively,

$$\eta_{E_i} \sim \mathcal{N}(0, \tau^2)$$

$$\eta_{S_j} \sim \mathcal{N}(0, \phi_S^2)$$

and two conditional distributions,

$$R_{ij}|\eta_{E_i} \sim \mathcal{N}(\eta_{E_i}, \phi^2)$$

$$\delta W_{ij} = R_{ij} - \eta_{E_i}|\eta_{S_j} \sim \mathcal{N}(\eta_{S_j}, \sigma_e^2)$$

To estimate event terms, we can have,

$$\begin{aligned}\mathbb{P}(\eta_{E_i}|R_{ij}) &\propto \pi(\eta_{E_i})\mathbb{P}(R_{ij}|\eta_{E_i}) \\ &= \varphi(\frac{\eta_{E_i} - 0}{\tau})\varphi(\frac{R_{ij} - \eta_{E_i}}{\phi}) \\ &= \varphi(\frac{\eta_{E_i} - \mu_i}{V_i})\end{aligned}$$

where $\varphi(\cdot)$ is the standard normal probability density function, μ_i and V_i are the mean and varaince of posterior distribution $\eta_{E_i}|R_{ij}$, N_i is the number of recordings in event i. They are,

$$\mu_i = \frac{\frac{N_i}{\phi^2}\overline{R_i}}{\frac{1}{\tau^2} + \frac{N_i}{\phi^2}}$$

$$V_i^2 = \frac{1}{\frac{1}{\tau^2} + \frac{N_i}{\phi^2}}$$

then, the Bayesian estimator of event term is

$$\hat{\eta}_{E_i} = \mathbb{E}(\eta_{E_i}|R_{ij}) = \mu_i = \frac{\frac{N_i}{\phi^2}\overline{R_i}}{\frac{1}{\tau^2} + \frac{N_i}{\phi^2}} \tag{3.2}$$

But the variance of event terms τ^2 and variance of within event residuals ϕ^2 are unknown. So iteration method is applied, the procedure is,

- initialize parameters τ^2 and ϕ^2;
- plug them into Equation 3.2 to get event terms $\hat{\eta}_{E_i}$;
- compute the sample variance of event terms $\hat{\tau}^2$ and within event residuals $\hat{\phi}^2$;
- repeat step 2 and 3 until they converge.

After obtaining event terms, we can compute within event residuals $\delta W_{ij} = R_{ij} - \eta_{E_i}$, and then apply the same approach to calculate site terms,

$$\begin{aligned}
\mathbb{P}(\eta_{S_j}|\delta W_{ij}) &\propto \pi(\eta_{S_j})\mathbb{P}(\delta W_{ij}|\eta_{S_j}) \\
&= \varphi(\frac{\eta_{S_i} - 0}{\phi_S})\varphi(\frac{\delta W_{ij} - \eta_{S_j}}{\sigma_e}) \\
&= \varphi(\frac{\eta_{S_j} - \mu_j}{V_j})
\end{aligned}$$

then, the Bayesian estimator of site term is

$$\hat{\eta}_{S_j} = \mathbb{E}(\eta_{S_j}|\delta) = \mu_j = \frac{\frac{n_j}{\sigma_e^2}\overline{\delta W_{ij}}}{\frac{1}{\phi_S^2} + \frac{1}{\sigma_e^2}}$$

where n_j is the number of recordings at station j. The iteration is also applied to compute $\hat{\phi}_S^2$, $\hat{\sigma}_e^2$, and $\hat{\eta}_{S_j}$.

In R, I used package *nlme* [11] (for mixed effects model) to estimate all these parameters.

3.3 Comparison of Two Approaches

The residuals described above are in terms of pseudo spectral acceleration (PSA) which has been described in Chapter 2. It is period dependent variable. For buildings, the typical

fundamental periods of interest are, 0.010, 0.050, 0.500, and 2.500 sec. They represent the relatively stiff short buildings to relatively soft tall buildings. Then the PSA at these periods are the most important. Therefore, event terms and site terms at these periods are imperative to calculate and will be studied in this thesis. The calculations of event terms and site terms from both Frequentist approach and Bayesian model are shown in Fig. 3.3 and Fig. 3.4 below.

From statistical view of point, Frequentist additive model gives unbiased estimators, while estimators from Bayesian hierarchical model are biased. However, Bayesian hierarchical model shrinks event terms and site terms to avoid overfitting, and reduces variance of event terms and site terms. This is reflected in Fig. 3.3 and Fig. 3.4 that the points of Bayesian are consistently absolutely smaller and closer to zero. It is also clear to see this trend in Table 3.1. It summarizes the standard deviations of event terms and site terms at the periods of 0.010, 0.050, 0.500, and 2.500 sec for both Bayesian and Frequentist methods. The Bayesian approach gives consistently smaller values than Frequentist.

Moreover, Bayesian hierarchical model is more flexible and robust to outliers. The event or station which has only a few recordings can obtain additional information from other events and/or stations through prior distribution. Due to these features, I will use Bayesian hierarchical model to compute event terms and site terms for spatial visualization in next chapter.

Table 3.1: Standard deviations of event terms ($\hat{\tau}$) and site terms ($\hat{\phi}_S$) at the periods T of 0.010, 0.050, 0.500, and 2.500 sec

Period	**Frequentist** $\hat{\tau}$	**Bayesian** $\hat{\tau}$	**Frequentist** $\hat{\phi}_S$	**Bayesian** $\hat{\phi}_S$
0.010 sec	0.4875	0.3887	0.5675	0.2680
0.050 sec	0.5145	0.4102	0.5549	0.2645
0.500 sec	0.5071	0.3836	0.6106	0.2765
2.500 sec	0.6441	0.5624	0.5565	0.2839

Figure 3.3: Event terms of 134 earthquakes by Frequentist and Bayesian approach

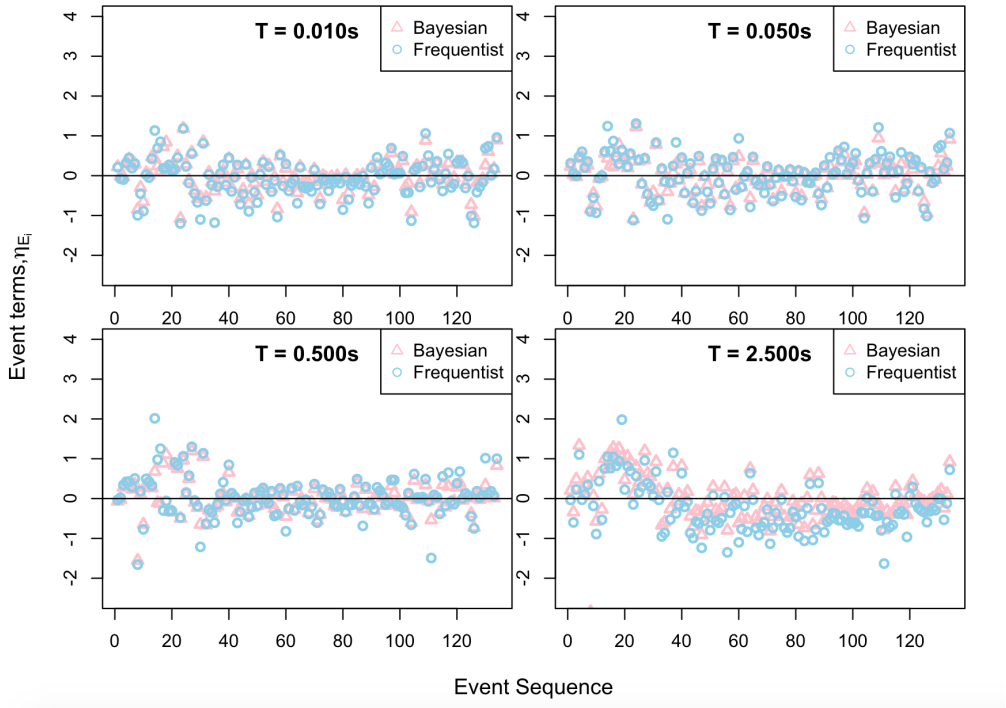
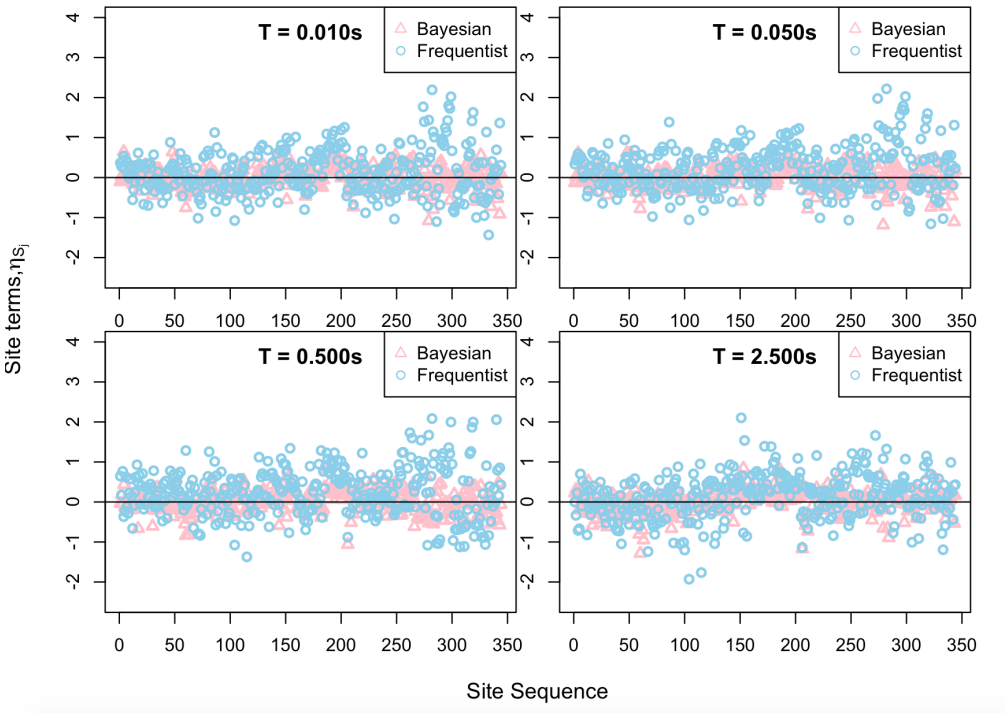


Figure 3.4: Site terms of 344 sites by Frequentist and Bayesian approach



CHAPTER 4

Kriging

Site terms and site responses could be well-predictable so that we can use them for future structural design. However, even in Los Angeles County where the density of ground motion stations is almost the highest in the U.S., it is very likely the there is no ground motion station nearby (within 2 km) to derive site-specific site response. So it is imperative to develop a model to predict site terms where there is no nearby station. In this study, I implement the most popular geostatistical method–Kriging to conduct prediction (or also called interpolation) on site terms. Then this chapter will first introduce the basic theory of Kriging; and then implement Kriging to generate the heat map of site terms of Los Angeles County.

4.1 Basic Theory of Kriging

In this section, I will introduce the basic theory of Kriging in seismic site term context. The basic form of Kriging is quite simple, which is the sum of weighted measurements,

$$\hat{\eta}_{S_0} = w_1\eta_{S_1} + w_2\eta_{S_2} + \cdots + w_n\eta_{S_n} = \sum_{j=1}^{n} w_j\eta_{S_j} \tag{4.1}$$

where $\{\eta_{S_j}, j = 1, 2, \cdots, n\}$ is the site term of the j-th station; η_{S_0} is the site term at the site of our interest without any ground motion records; and w_j is the weight of η_{S_j} used to combine n observed site terms to estimate unknown η_{S_0}.

To preform the prediction, the key is to obtain the weight of each η_{S_j}. Obviously, the weight of η_{S_j} should be higher if the station S_j is closer to our site S_0. Inverse distance weighting or other similar method could be good options, however, they do not consider

spatial variations. In contrast, Kriging perfectly combines spatial correlation and distance to estimate the weights.

To quantify the relation between spatial correlation and distance, variogram (denoted as $2\gamma()$ or semivariogram $\gamma()$) were inverted by [10]. The variogram $2\gamma()$ function describes variability based on the separation distance. The function can be written as,

$$2\gamma(h) = \mathbb{V}\left[\eta_S(h) - \eta_S(0)\right] \tag{4.2}$$

where $\mathbb{V}$ represents variance, h is the separation distance between any two data points which are separated by h. If we bin h and calculate the corresponding $2\hat{\gamma}(h)$, we would be able to get many experimental estimates of $2\hat{\gamma}(h)$ versus h. By fitting smooth continuous functions, we then can easily make forward application based on the fitting function. Fig. 4.1 shows an example of experimental semivariogram of site terms and the spherical fitting function (one of the most commonly used functions). The formula is,

$$\gamma(h) = \begin{cases} c_0, & h = 0 \\ c_0 + c_1(\frac{3}{2}(\frac{h}{\alpha}) - \frac{1}{2}(\frac{h}{\alpha})^3), & 0 < h \leq \alpha \\ c_0 + c_1, & h > \alpha \end{cases}$$

where c_0 is nugget (≈ 0.05 in Fig. 4.1), the semivariogram value at distance $h = 0$, was first suggested by [9]. It believed that microscale variation can cause a discontinuity at the origin. $c_0 + c_1$ is sill (≈ 0.64 in Fig. 4.1), the quantity of plateau shown in Fig. 4.1, and α ($\approx 900km$ in Fig. 4.1) is range that is corresponding to the starting point of plateau. We can see as separation distance increases, the semivariogram is increasing. When the distance reaches range, semivariogram becomes plateau, then the correlation inversely becomes very small. We will say there is almost no correlation or they are independent.

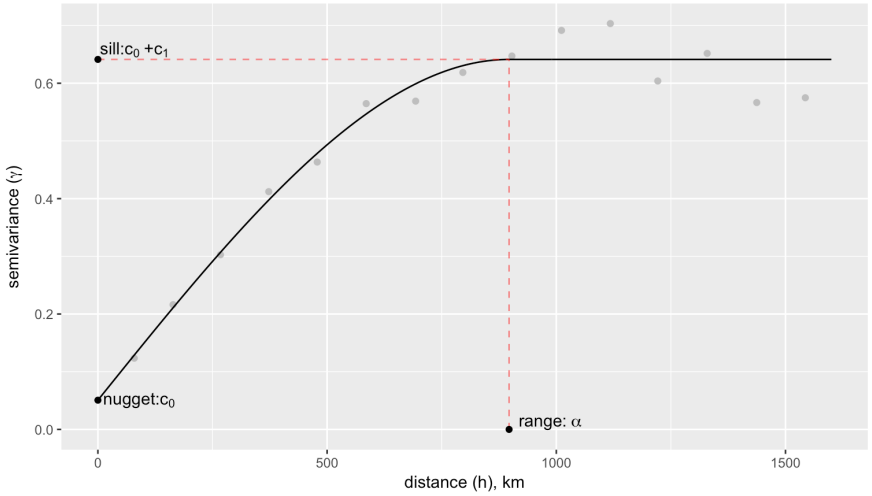


Figure 4.1: Spherical fitting curve and its nugget, sill, and range

By having the tool of semivariogram, then we can solve for the weights. From Eq. 4.1, we have

$$\hat{\eta}_{S_0} = \sum_{j=1}^{n} w_j \eta_{S_j} \tag{4.3}$$

where we also have constraint $\sum_{j=1}^{n} w_j = 1$ to ensure unbiased estimation. Our objective is to minimize to the mean squared error of prediction,

$$\min \quad \sigma_e^2 = \mathbb{E}\left[\eta_{S_0} - \hat{\eta}_{S_0}\right]^2 = \mathbb{E}\left[\eta_{S_0} - \sum_{j=1}^{n} w_j \eta_{S_j}\right]^2 \tag{4.4}$$

where $\mathbb{E}$ is the expectation. By assuming the expectation of any site $\mathbb{E}(\eta_S) = \mu$ is a constant (ordinary Kriging assumption), we can plug Eq. 4.3 into Eq. 4.4 and conduct some algebra to get,

$$\sigma_e^2 = \mathbb{E}\left[\eta_{S_0} - \sum_{j=1}^{n} w_j \eta_{S_j}\right]^2 = 2\sum_{j=1}^{n} w_j \gamma(S_0 - S_j) - \sum_{j=1}^{n}\sum_{k=1}^{n} w_j w_k \gamma(S_j - S_k) \tag{4.5}$$

Considering the constraint $\sum_{j=1}^{n} w_j = 1$, we will have

$$\min \quad 2\sum_{j=1}^{n} w_j \gamma(S_0 - S_j) - \sum_{j=1}^{n}\sum_{k=1}^{n} w_j w_k \gamma(S_j - S_k) - 2\lambda\Big(\sum_{j=1}^{n} w_j - 1\Big) \tag{4.6}$$

where λ is the Lagrange multiplier. By differentiating Eq. 4.6 with respect to each $w_1, w_2, \ldots, w_n$ and λ and set the derivatives equal to zero, we eventually will get the best weight. The

weights are (expressed in matrix form),

$$\boldsymbol{W} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} \tag{4.7}$$

where

$$\boldsymbol{W} = (w_1, w_2, \ldots, w_n, \lambda)$$

which includes Lagrange multiplier.

$$\boldsymbol{\gamma} = (\gamma(S_0 - S_1), \gamma(S_0 - S_2), \ldots, \gamma(S_0 - S_n), 1)^\top$$

where $\top$ means transpose. It is the vector of semivariogram between all observations S_j and our interested point S_0 and 1. 1 is associated to Lagrange multiplier λ.

$$\boldsymbol{\Gamma} = \begin{cases} \gamma(S_j - S_k), & j = 1, 2, \ldots, n, \;\; k = 1, 2, \ldots, n \\ 1, & j = n+1, \;\; k = 1, 2, \ldots, n \\ 1, & k = n+1, \;\; j = 1, 2, \ldots, n \\ 0, & j = n+1, \;\; k = n+1 \end{cases}$$

which is a $(n+1) \times (n+1)$ matrix. The another way to represent Kriging system $\boldsymbol{\Gamma W} = \boldsymbol{\gamma}$ will be,

$$\begin{bmatrix} \gamma(S_1 - S_1) & \gamma(S_1 - S_2) & \cdots & \gamma(S_1 - S_n) & 1 \\ \gamma(S_2 - S_1) & \gamma(S_2 - S_2) & \cdots & \gamma(S_2 - S_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \gamma(S_n - S_1) & \gamma(S_n - S_2) & \cdots & \gamma(S_n - S_n) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(S_0 - S_1) \\ \gamma(S_0 - S_2) \\ \vdots \\ \gamma(S_0 - S_n) \\ 1 \end{bmatrix}$$

Therefore, if the semivariogram between any two of sites (including unknown site S_0) are given, we can calculate weights and then estimate η_{S_0} by Eq. 4.3. And the semivariogram fitting function is just the tool to provide semivariogram between any two points given their separate distance.

4.2 Implementation of Kriging on Site Terms

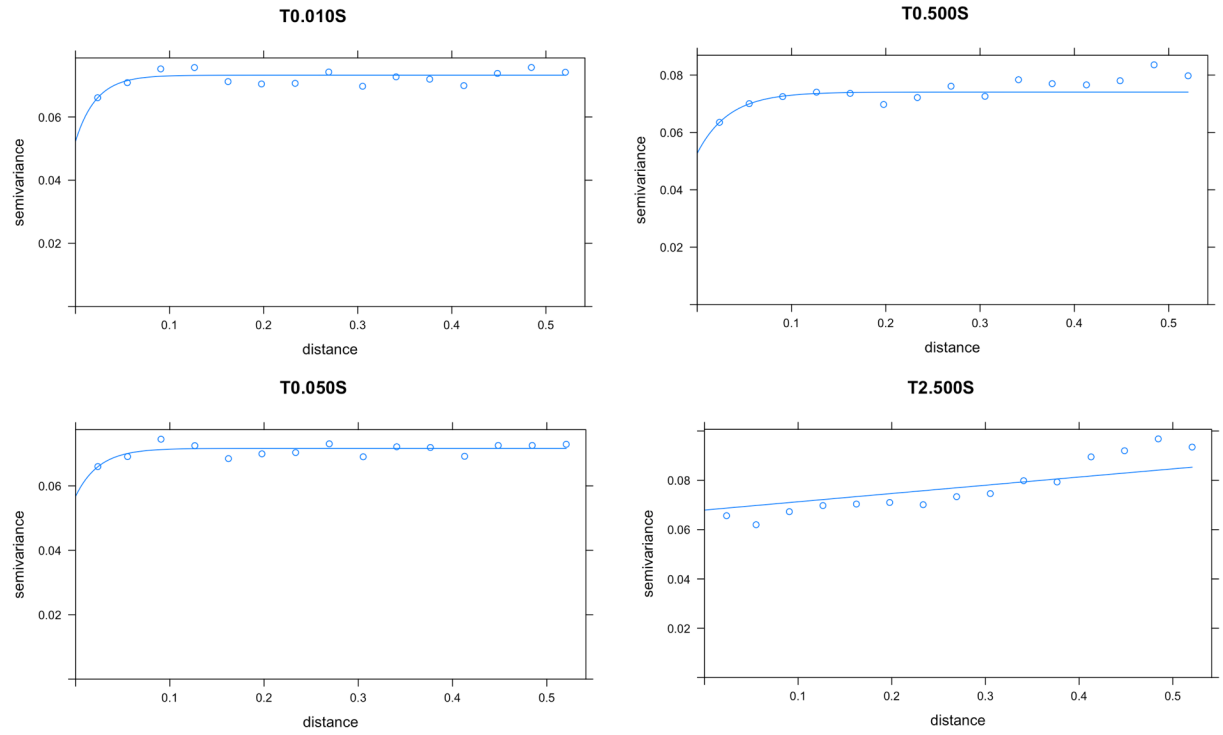
Based on the derivations of semivariogram and Kriging from last section, the site terms at any given location can be estimated from the observed site terms. If a dense grid in Los

Angeles County is generated, then we can implement Kriging to calculate site terms for each dot in the grid and then procedure a high resolution map of site terms.

The reason to predict for site terms rather than site responses is because site terms are relatively clearer and simpler (without linear, nonlinear, and other effects involved) and also site terms meet the assumption for ordinary Kriging. For ordinary Kriging, the expectation of the variable is constant. Apparently, site responses are not constant, however, site terms could be constant if the ergodic site response from GMM is unbiased. The unbiasedness of ergodic GMM can be validated by Fig. 3.3 and Fig. 3.4 where the overall average of site terms is around zero.

To perform Kriging prediction, first semivariogram is needed. In Fig. 4.2, I present the experimental semivariograms at the four representative periods, 0.010, 0.050, 0.500, and 2.500 sec, of PSA and the best fitting functions.

Figure 4.2: Semivariogram functions

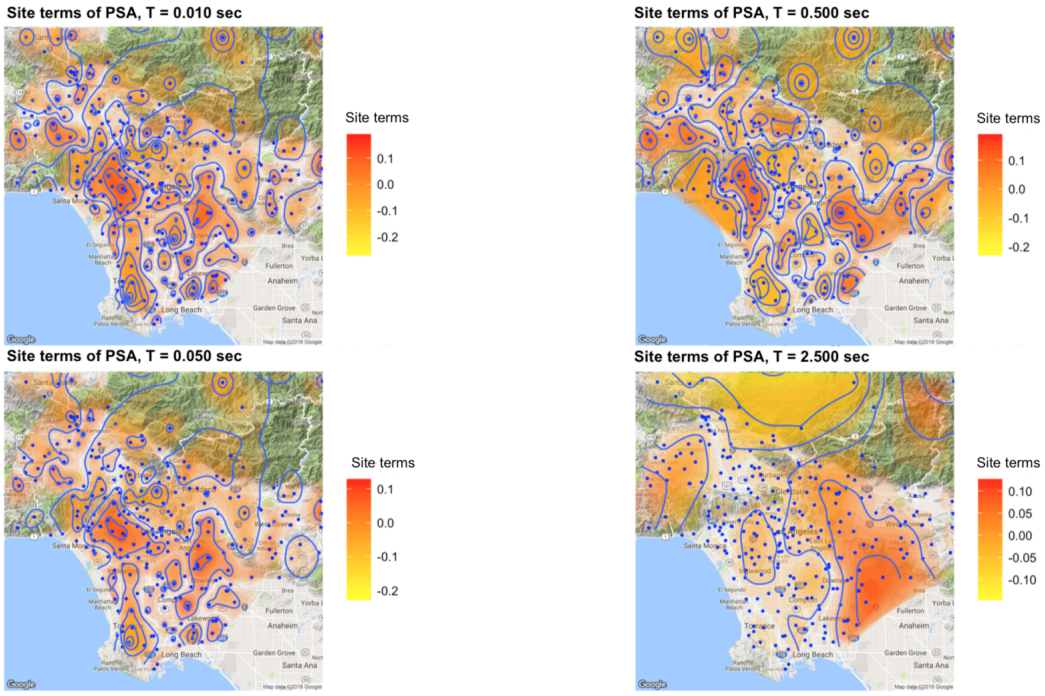


Note, the unit of distance in Fig. 4.2 is degree (as the location of station is described by longitude and latitude). The range α starts from around 0.05 at period of 0.01 sec, to 0.07

at period of 0.05 sec, to 0.1 at period of 0.5 sec, to very far at period of 2.5 sec. It reflects the spatial correlation range is further at longer period. This is because PSA at longer periods reflect the behavior of deeper soil structure and the deeper soil structure could be more continuous than shallower soil (due to less disturbance and geotectonic development) so that the spatial correlation of deeper soil structure and PSA at longer periods can go further.

Based on fitting functions at four periods and the derivations in last section, the site terms can be predicted at any location. Then I generate a dense grid in Los Angeles County and then calculate the site terms for each dot. The following figure shows heat map of site terms in Los Angeles County

Figure 4.3: Kriging prediction for site terms



In the above heat map, darker red and yellow represent larger positive site terms and larger negative site terms. The overall average of site terms is about zero (validating the constant expectation assumption of ordinary Kirging). The larger positive site terms mainly locate around center of Los Angeles county where it is in Los Angeles basin. It means that ergodic site response from GMM under-estimates ground motion the most in Los Angeles

basin. The reason to this result is due to focusing effect. The basin substructure can locally focus the seismic wave energy to amplify ground motion higher than the sites without basin. Although the basin effect has been accounted in ergodic GMM, the basin effect is still under-estimated for Los Angeles basin. Thus, if ergodic GMM is utilized in structural design, the infrastructures are very likely damaged due to the model under-estimation. Therefore, it is crucial to extract site terms and calculate site-specific site response to estimate ground motion for structural design.

Then given the site term at site j, site-specific site response can be estimated by

$$\tilde{F_S} = \eta_{S_j} + F_S(V_{S30}, R_{JB}, \mathbf{M}, region, z_1) \tag{4.8}$$

where $F_S(V_{S30}, R_{JB}, \mathbf{M}, region, z_1)$ is ergodic site response from GMM given by Eq. 2.4.

CHAPTER 5

Conclusion

GMM has been the crucial and mandatory calculation for earthquake risk estimation and management. However, as this study and many previous studies state that GMM is ergodic and could give very large biased prediction. If the bias is negative, the ergodic GMM under-estimates ground motion, the structural designs are safe but could be conservative to waste resources. In contrast, if the bias is positive, then the structural designs are not stable or resistant to any expected potential earthquakes. Therefore, ergodic GMM is not suggested to be used as the final seismic risk estimation, non-ergodic GMM should be highly recommended.

This study presents a method to correct the site response effect in ergodic GMM to be spatially non-ergodic or site-specific site responses. It corrects the site responses by adding the systematic site response bias, site term (η_S). Two approaches to extract site term from residuals analysis are investigated and compared. Based on the actual analysis on ground motion data in Los Angeles County, Bayesian approach is recommended especially for limited data. This is because Bayesian approach is robust to outliers, is flexible to capture complex hierarchical data structure, and avoids overfitting.

This study also proposes to use Kriging to predict site terms for sites where there are no ground motion data available for direct site term calculation and then generate high resolution heat map of site terms in Los Angeles County. Based on the heat map, I found GMM could not capture Los Angeles basin effects, particularly for short periods. The larger positive site terms indicate ergodic site response in GMM under-estimates ground motions. This finding strengthens the statement that non-ergodic GMM or at least site-specific site response GMM should replace ergodic GMM for earthquake risk assessment.

The limitation of this study is that the data set is still small so the site terms derived may not be the truth. Additionally, the predicted site terms from Kriging are not verified with any data and could not be used in actual structural design. However, as the number of ground motion stations increases, data will also increase, we will be able to get more accurate site terms so is site-specific site response to mitigate earthquake damage.

References

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

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

[3] Boore, D. M., Stewart, J. P., Seyhan, E., 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.

[4] Campbell, K. W., 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.

[5] Chiou, B. S.-J., Youngs, R. R. (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*. **30(3)**, pp 1117-1153.

[6] Chopra, A. K. (2006) Dynamics of Structures. *Pearson*. pp 168-169.

[7] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). Bayesian Data Analysis, 3rd edition. *CRC Press.*

[8] Idriss, I. M. (2014) An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthquake Spectra*. **30(3)**, pp 1155-1177.

[9] Matheron, G. (1962). 2, Traité de géostatistique appliquée, Tome I: Mémoires du Bureau de Recherches Géologiques et Minières. *Editions Technip*. **14**, pp 333.

[10] Matheron, G. (1963). Principles of geostatistics. *Economic Geology*. **58**, pp 1246-1266.

[11] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team. (2016). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-128. *http://CRAN.R-project.org/package=nlme.*

[12] Seyhan, E., Stewart, J. P. (2014). Semi-empirical nonlinear site amplification from NGA-West 2 data and simulations. *Earthquake Spectra.* **30(3)**, pp 1241-1256.

[13] Stewart, J. P., Afshari, K., Goulet, C. A. (2017) Non-Ergodic site response in seismic hazard analysis. *Earthquake Spectra*. **30(3)**, pp 1241-1256.

[14] Wang, P., Stewart, J. P., Bozorgnia, Y., Boore, D. M. (2017) R Package for Computation of Earthquake Ground Motion Response Spectra. *Report No. 2017/09* Pacific Earthquake Engineering Research Center, UC Berkeley, CA..