UNIVERSITY OF CALIFORNIA
Los Angeles

Prediction Model Development of Seismic Building Responses

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Han Sun

2018

ABSTRACT OF THE THESIS

Prediction Model Development
of Seismic Building Responses

by

Han Sun
Master of Science in Statistics
University of California, Los Angeles, 2018
Professor Arash Ali Amini, Chair

The ability to predict building responses subjected to an earthquake could be used to identify building damage which would largely reduce human inspection effort and operation down-time. This thesis explores various of machine learning methods to formulate prediction model for seismic building responses over the great Los Angeles region using three actual earthquake scenario data (1994 Northridge, USA, 1999 Chi-Chi, Taiwan and 2000 Tottori, Japan). The result shows that the geospatial interpolation method kriging outperforms other candidates among all earthquakes in both accuracy and model stability using criteria such as cross-validation and median absolute residual difference. Some inconsistency in accuracy levels between different earthquakes are caused by 1)earthquake characteristics and 2)representativeness of data samples of each event.

The thesis of Han Sun is approved.

Qing Zhou

Yingnian Wu

Arash Ali Amini, Committee Chair

University of California, Los Angeles

2018

To my mother ...

who encouraged me discover science

since my childhood

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to acknowledge and thank all the professors I have been taking classes and discussed with at Statistics department of UCLA who not only delivers splendid lectures but also introduces deep insights, Prof.Chad Hazlett, Prof.Songchun, Zhu, Prof.Yingnian, Wu, Prof.Qing Zhou, Dr. Nicolas Christou and my thesis committee chair Prof. Arash Amini for his patience and guidance.

CHAPTER 1

Introduction

1.1 Background and Motivation

Seismic hazard is one of the most severe natural threats to modern urbanized areas where buildings are often condensed and tall. Due to the inherent dynamic characteristics of both buildings and frequency contents of earthquake, tall and slender buildings are often more vulnerable to earthquakes. In seismic active regions such as US west coast, it is crucial for the urban centers to be resilient when subject to frequent and large earthquakes. Bruneau and Reinhorn [2] defined building resilience as an ability of the structure to possibly avoid the shock, or to absorb such a shock and to restore to its normal performance after subjecting to the shock. The first phase of the ability can be referred as a robust design of structures which is mandatory required by the building design code; however, given the uncertainty of earthquake characteristics, there is always a possibility that the building goes into second and third phase.

In order to timely recover from a damage state to the normal state of the building, rapid damage detection is essential. There are considerably large amount of research being done for damage detection in both theoretical simulation and experiment testing perspective. For example, Pandey and Biswas [3] developed a damage detection algorithm based on changes in structural stiffness demonstrated by numerical analysis of a truss structural system; Kosmatka and Ricles [4] proposed an experimental procedure that relies on changes of system modal vibration characteristics which determines structural damage state. All these techniques requires accurate knowledge of dynamic responses of structures such as Peak Story Drift Ratio (PSDR) and Peak Floor Acceleration (PFA).

It is through structural health monitoring and sensing techniques to retrieve actual structural response under earthquake in field. However, it is practically not realistic to install sensors on every building at all interested measurement locations within each floor and over all floors. Therefore, it is important to develop an interpolation scheme which utilizes known structural responses to predict unknowns. This process contains two major components, intra-building (interpolate along building height) and inter-building (interpolate between different buildings). There are numerical research in former aspects. For example, Naeim [5] developed a cubic interpolation method for measured structural responses along building height; Limongelli [6] developed a spline function which interpolates deformed structural shape based on measurement and determines structural damage accordingly. On the contract, there is very few research on inter-building interpolation. Several research has been done in quantifying correlation patterns between different buildings. Loth and Baker [7] studied spatial correlation between different buildings by simplifying building responses as their Spectral Acceleration at first mode period (SA_{T1}). They developed semi-variogram models between different SA_{T1} which represents different buildings. Another study by DeBock *et. al.* [8] extended the interested variable to be actual building responses using simulated strucutral responses result from Nonlinear Response History Analysis (NRHA). They evaluated correlation patterns of PSDR and PFA of actual building responses with respect to SA_{T1}. They further proposed an empirical model to predict building responses using SA_{T1}. However, SA_{T1} is usually not available at most locations of majority buildings. To practically interpolate building responses, more advanced statistical techniques are necessary.

1.2 Objective

This study intends to explore several statistical and machine learning based interpolation methods for two major measure of building seismic responses, PSDR and PFA respectively, that are directly related to building damage. Considering that building damage are primarily related to maximum PSDR and PFA over building height, only the maximum building responses are used to formulate statistical models.

1.3 Source of Data

Three major earthquake scenarios are considered in this study. They are 1994, Northridge in United States; 1999, Chi-Chi in Taiwan and 2000 Tottori in Japan. They all occurred in high density urban areas and caused tremendous damage to buildings and were recorded over a large geo-area. This provides sufficient amount and magnitude ground motion acceleration time history data available at Peer NGA2 West Database [9] to conduct the analysis. The three earthquakes all have large magnitude from 6.61 to 7.62. Often, Peak Ground Acceleration (PGA) is a general measure of earthquake intensity at a particular site. However, the maximum geometric mean PGA differ a lot from each other as summarized in Table 1.1. This is mainly due to various site properties such as soil and earthquake characteristics. Another notable differences between the three earthquakes is the geo-spatial coverage. As demonstrated in Figure 1.1, black dot represents epicenter and red crosses are where stations locate. Northridge earthquake covers an area of approximate 18500 km^2. In contrast, 400 stations from Chi-Chi earthquake covers 36200 km^2 and Tottori has covered 112000 km^2. As a result, the recorded stations has uneven distributions over the area which affects the prediction models and their result. This will be discussed in later chapter as well.

Table 1.1: Summary of Earthquake Statistics

Earthquake	*Mw.*[1]	*No. of Records*	*Year*	*Fault Type*	*Vector Sum PGA*[2]
Chi-Chi, Taiwan	7.62	400	1999	Strike Slip	0.047 - 2.01
Northridge, USA	6.7	152	1994	Blind Thrust	0.008 - 1.15
Tottori, Japan	6.61	414	2000	Strike Slip	0.001 - 1.13

[1] Magnitude.

[2] Peak Ground Accerlation (PGA) in g.

A vector sum PGA histogram of Northridge earthquake is shown in Figure 1.2. It is observed that the distribution of PGA concentrates at 0.2 g and decreases in an exponential manner while the peak PGAs are as high as 1.5 and 2 g. This indicates that building responses distribute in a highly nonlinear pattern due to the irregular distribution of earth-

quake intensity measured as PGA. It further suggests that linear prediction models are not suitable for the purpose of this study.

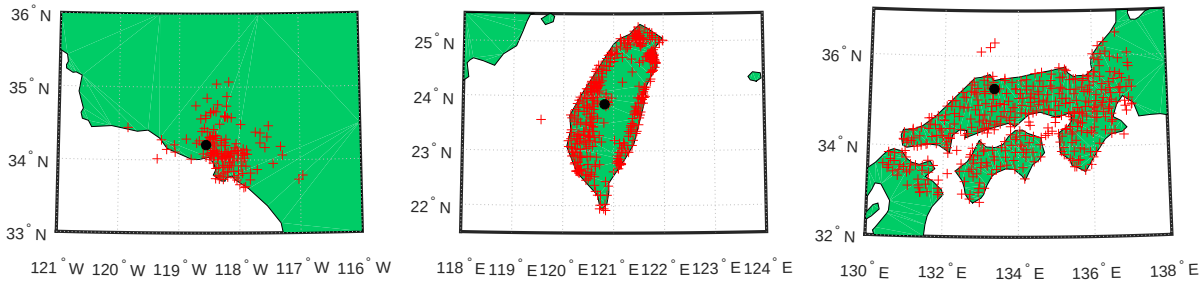


Figure 1.1: Locations of Recorded Stations for Each Earthquake Scenario

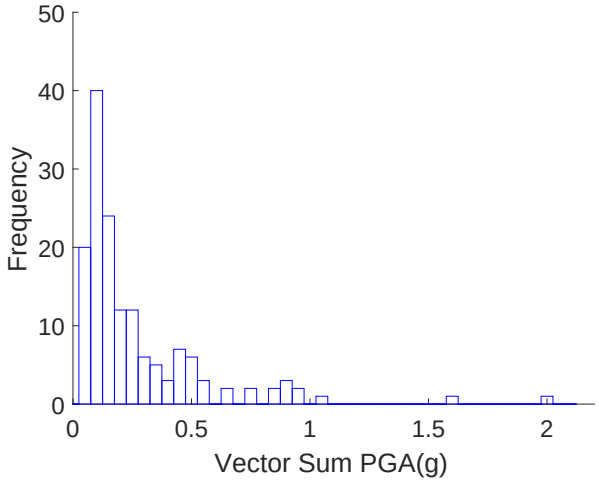


Figure 1.2: PGA Histogram of Northridge Earthquake

NRHA was conducted to retrieve building's seismic responses under the pre-described ground motions from the three earthquakes using OpenSees [10]. A series of reinforced concrete moment frame building models developed by Haselton [11] are selected to formulate a representative pool. The major structural dissimilarity between buildings is building height and corresponding design criteria. From NRHA, interested EDPs (PFA and PSDR) were extracted and used as data in this study. An example PSDR histogram is shown in Figure 1.3.

Table 1.2: Building Models for Nonlinear Response History Analysis

Building Index	*1st Mode Period (sec)*	*Number of Stories*
CMF-2	0.66	2
CMF-4	1.12	4
CMF-8	1.71	8
CMF-12	2.01	12
CMF-20	2.63	20

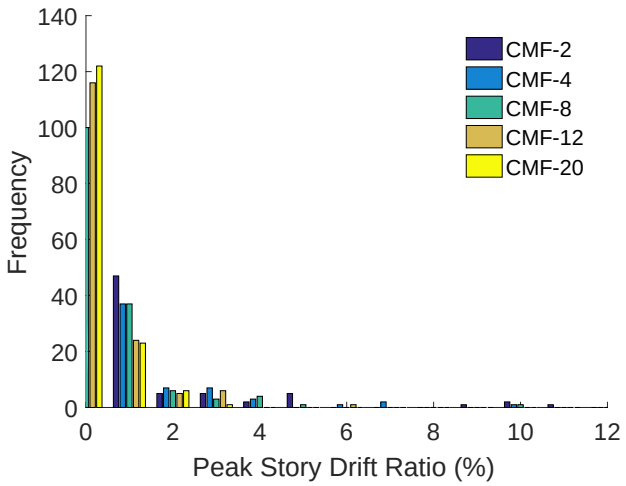


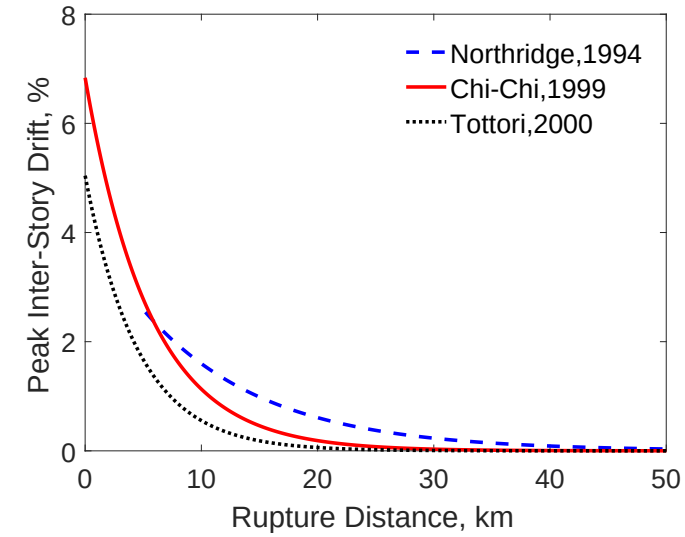
Figure 1.3: PSDR Histogram under Northridge Earthquake

Comparing between the three earthquakes, an exponential fitted trend of PSDR and PFA over epic distance of 2-story building is demonstrated in Figure 1.4. It shows that PSDR is significantly dominated by epic distance and drops faster than that of PFA. In addition, the trend of PSDR between Tottori and Chi-Chi is consistent while that of PFA crosses each other at 30km epic distance. These observations, also observed in other buildings, indicate that PSDR is more consistent with geo-spatial features while PFA might contain contents that is not able to be explained by geo-spatial features which is discussed in later chapter too.

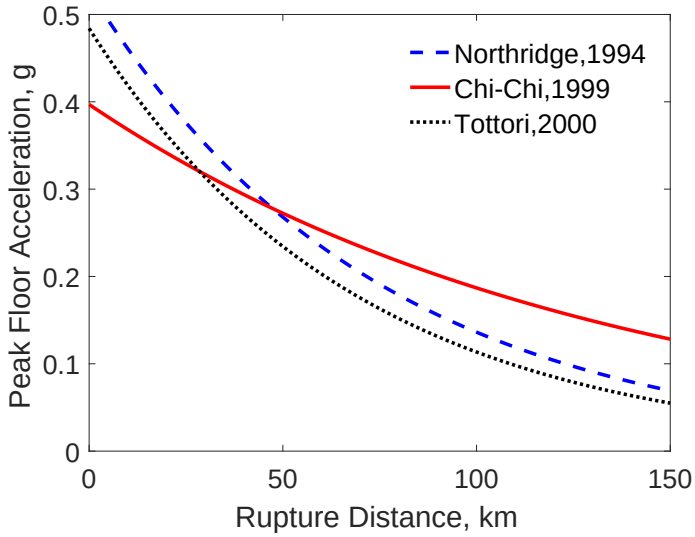
To summarize, there are two types of EDPs available at each site from each earthquake. For each EDP, it can come from five different buildings. All EDPs are treated as response variable Y and the predictors are latitude and longitude of a site.

1.4 Proposed Methodology

There are three categories of methodology being used in this study. The first one is a traditional geo-statistical tool, Kriging or Gaussian processing regression. The other two are recent popularized Kernel based machine learning techniques, Kernel ridge regression and Kernel support vector machine respectively. These methods are applied and compared to classical regression models such as ordinary least squares and naive model. The naive model is selected to be a baseline comparison and ordinary least squares is used as a linear model benchmark. Their performance is evaluated based on a non-replacement Bootstrap which gives an error distribution and will be discussed in detail in later chapter.



(a) PSDR



(b) PFA

Figure 1.4: EDP Exponential Fitted Trend over Rupture Distance of 2-story Building

CHAPTER 2

Methodology

All methods being used in this study are derived in this chapter using frequentist's approach. The uncertainity caused by underlying distribution of model parameters is considered through a non-replacement Bootstrap procedure. The geo-data in this study can be naturally fitted into the geo-spatial model Kriging while the geo-coordinate latitude and longitude can be considered as features for other general models.

2.1 Naive Model

The naive model is used as a baseline which makes predictions by averaging training response variables as shown in 2.1.

$$\hat{Y} = \bar{Y} \tag{2.1}$$

2.2 Ordinary Least Squares

Ordinary Least Squares (OLS) is commonly used in a wide range of engineering, biology and social research field. It is a linear regression method to predict dependent variables using a series of predictors. The model is expressed in matrix form in Equation 2.2 with the predictor being a matrix of dimension $n \times p$. Y is a $n \times 1$ vector (response variable) and ϵ is the residual and assumed to follow Normal distribution with mean 0 and variance σ_n^2.

$$Y = X\beta + \epsilon \tag{2.2}$$

β, model parameter of dimension $p \times 1$, can be estimated by Equation (4) (OLS estimator) which is solved by minimizing Residual Sum of Squares (RSS): $(Y - X\beta)^T(Y - X\beta)$. RSS

for OLS is a matrix form of objective function. There is no penalty term in OLS model. It has a close-form solution, or unique global minimum as shown in Equation 2.3.

$$\hat{\beta}_{OLS} = (X^T X)^{-1} XY \tag{2.3}$$

The OLS estimator is given by Equation 2.4.

$$\hat{f}(X) = X\hat{\beta}_{OLS} \tag{2.4}$$

This derivation is the frequentist's view which assumes that model parameter has a single true value instead of a distribution. Therefore, it is equivalent to the Maximum Likelihood Estimator (MLE) of the residual ϵ which is showed in Equation 2.5.

$$\begin{aligned}
\hat{\beta}_{MLE} &= \max_{\beta} p(Y|X,\beta) \\
&= \max_{\beta} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(- \frac{(y_i - \beta x_i)^2}{2\sigma_n^2} \right) \\
&= \max_{\beta} \log\left(\frac{1}{\sqrt{2\pi}\sigma_n} \right) - \sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{2\sigma_n^2} \\
&= \min_{\beta} (Y - X\beta)^T (Y - X\beta)
\end{aligned} \tag{2.5}$$

2.3 Ordinary Kriging

Ordinary Kriging is a linear interpolation scheme which assumes a random field expressed through a variogram [12] over spatial locations $s_1, s_2, ...s_n$. It is suitable for geo-spatial data prediction and therefore selected for this study[13]. The model is given by Equation 2.6.

$$f(s) = \mu + \delta(s) \tag{2.6}$$

Where $\delta(s)$ is a location dependent stochastic component defined by semivariogram $\gamma(\cdot)$ and $f(s)$ is response variable Y at site s. The model estimator is given by Equation 2.7 subjected to $\sum_{i=1}^{n} w_i = 1$ to ensure unbiasness[12].

$$\hat{f}(s_0) = \sum_{i=1}^{n} w_i f(s_i) \tag{2.7}$$

Where w_i is an estimated weight. Same as OLS, the derivation starts with RSS in Equation 2.8 which can also be derived from MLE as well.

$$\min_{W} = E\big[\big(f(s) - \sum_{i=1}^{n} w_i f(s_i)\big)^2\big] \tag{2.8}$$

By assuming $f(s)$ is intrinsically stationary and incorporate the constraint using Lagrange multiplier λ, Equation 2.8 can be written to:

$$\min_{W} 2\sum_{i=1}^{n} w_i\gamma(s - s_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(s_i - s_j) - 2\lambda\big(\sum_{i=1}^{n} w_i - 1\big) \tag{2.9}$$

By taking derivative with respect to each w_i and λ, Equation 2.8 has a global minimum and is solved by Equation 2.10.

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{W}} \\ \lambda \end{bmatrix} &= \begin{bmatrix} \boldsymbol{Var_{f(s_i)}} & \mathbf{1} \\ \mathbf{1^T} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Cov_{f(s_i),f(s)}} \\ \mathbf{1} \end{bmatrix} \\
&= \begin{bmatrix} \gamma(s_1 - s_1) & \gamma(s_1 - s_2) & \dots & \gamma(s_1 - s_n) & 1 \\ \gamma(s_2 - s_1) & \gamma(s_2 - s_2) & \dots & \gamma(s_2 - s_n) & 1 \\ \dots & \dots & \ddots & \dots & 1 \\ \gamma(s_n - s_1) & \gamma(s_n - s_2) & \dots & \gamma(s_n - s_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(s - s_1) \\ \gamma(s - s_2) \\ \vdots \\ \gamma(s - s_n) \\ 1 \end{bmatrix}
\end{aligned} \tag{2.10}$$

Equation 2.10 is named ordinary Kriging based on its assumption in Equation 2.6 that there is a constant mean μ of this process. This assumption may not hold true for the given dataset. Another Kriging, universal Kriging is also used which assumes a trend existing in the data instead of constant. By doing so, the universal Kriging model is formulated as below and its solution can be refereed to [14].

$$f(s) = \mu(s) + \delta(s) \tag{2.11}$$

2.4 Kernel Ridge Regression

Kernel Ridge Regression (KRR) has flexibility in transferring X into high dimensional space using various of Kernels and is beneficial in dealing with complex data structure. There are

several ways to derive it. We start again with a model similar to Equation 2.2 and 2.6 as shown by Equation 2.12.

$$Y = \hat{f}(X) + \epsilon \tag{2.12}$$

Using a Kernel machine to represent the model, the KRR estimator of a single x is given by Equation 2.13.

$$\hat{f}(x) = \sum_{j=1}^{n} \alpha_j K(x, x_j) \tag{2.13}$$

The frequentist's approach is to add a l_2 regularization term, $\lambda|\alpha|^2$ to control model complexity such that the loss function which minimizes both RSS and regularizes λ is shown in Equation 2.14 and its matrix form in Equation 2.15 where K is the Kernel matrix:

$$\min_{\alpha} \sum_{i}^{n} (y_i - \alpha_j K(x_i, x_j))^2 + \sum_{i}^{n} \sum_{j}^{n} \alpha_i \alpha_j K(x_i, x_j) \tag{2.14}$$

$$\min \frac{1}{2} ||Y - K\alpha||^2 + \lambda \alpha^T K \alpha \tag{2.15}$$

Take derivative with respect to α and set it to zero,

$$K(K\alpha - Y) + \lambda K\alpha = 0 \tag{2.16}$$

$$(K + \lambda I)K\alpha = KY \tag{2.17}$$

In numerical cases, K may not be always invertible, use Single Value Decomposition (SVD) on K,

$$K = U\Sigma V^T \tag{2.18}$$

$$U(\Sigma + \lambda I)\Sigma V^T \alpha = U\Sigma V^T Y \tag{2.19}$$

Let $\tilde{\alpha} = V^T\alpha$ and $\tilde{Y} = V^T Y$

$$\left[\begin{pmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \lambda I & \mathbf{0} \\ \mathbf{0} & \lambda I \end{pmatrix} \right] \tilde{\alpha} = \begin{pmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tilde{Y} \tag{2.20}$$

Which can be solved as:

$$\tilde{\alpha} = \begin{pmatrix} (\Sigma_1 + \lambda I)^{-1} & \mathbf{0} \\ \mathbf{0} & \lambda^{-1} I \end{pmatrix} \begin{pmatrix} \Sigma_1 \tilde{Y} \\ \mathbf{0} \end{pmatrix} \tag{2.21}$$

This approximate α gives global minimum of the loss function 2.15 of the KRR model. The Kernel trick has been applied and is not discussed in detail here.

2.4.1 Kernel Construction

There are various of ways to construct a Kernel and the efficiency of learning is dependent on the data structure. For example, element in Gaussian Kernel matrix is defined as Equation 2.22.

$$k_g(x_i, x_j) = e^{-\frac{D(x_i,x_j)^2}{2\sigma^2}} \tag{2.22}$$

Where $D(x_i, x_j)$ is the Euclidean distance between x_i and x_j and σ is a control variable of the Kernel's function shape. A KRR with linear Kernel shown in 2.23 is often called ridge regression.

$$k_l(x_i, x_j) = x_i \cdot x_j \tag{2.23}$$

Several Kernels are visualized in Figure 2.1.

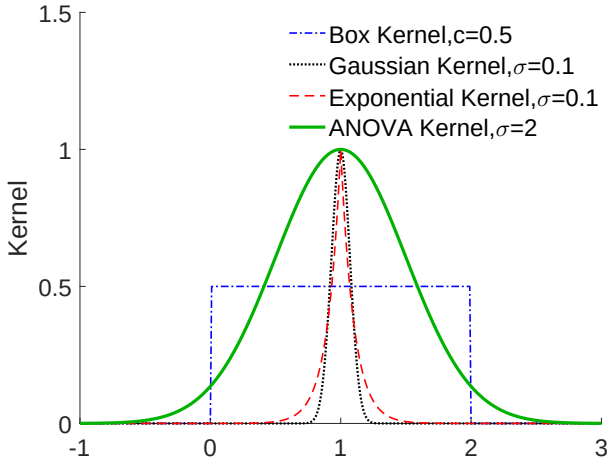


Figure 2.1: Demonstration of Kernels

2.5 Model Training

2.5.1 Non-Replacement Bootstrap

In order to evaluate model performance, A total of 100 non-replacement Bootstrap procedure is performed. Within each procedure, the data is randomly separated into training batch and testing batch based on a prefixed ratio 7 : 3. The Median Absolute Relative Devation (MARD)(Equation 2.24) is reported for each model where n_t is number of testing data such that a distribution of error measure MARD can be visualized.

$$MARD = median\{|\frac{y_i - \hat{f}(x_i)}{y_i}|, i = 1, 2, ..n_t\} \tag{2.24}$$

2.5.2 K-Fold Cross Validation

Within each procedure, model parameters are determined by k-fold cross validation using training data. For each $k = 1, 2, 3, ...K$, fit the model with parameter λ to the other $K - 1$ parts, the prediction error is computed as:

$$error_k(\lambda) = \sum_{i \in k^{th} part} (y_i - \hat{f}(x_i))^2 \tag{2.25}$$

Which gives cross validation error:

$$CV(\lambda) = \frac{1}{K} \sum_{k=1}^{K} error_k(\lambda) \tag{2.26}$$

The λ value which generates least CV is selected.

2.5.3 Response Variable Logarithm Transfer

An additional nature logarithm transfer is performed on all EDPs of the dataset. This is based on knowledge that EDPs and IMs often follows log-normal distribution over geo-spatial domain and such transformation makes normal assumption valid of the prediction residual ϵ.

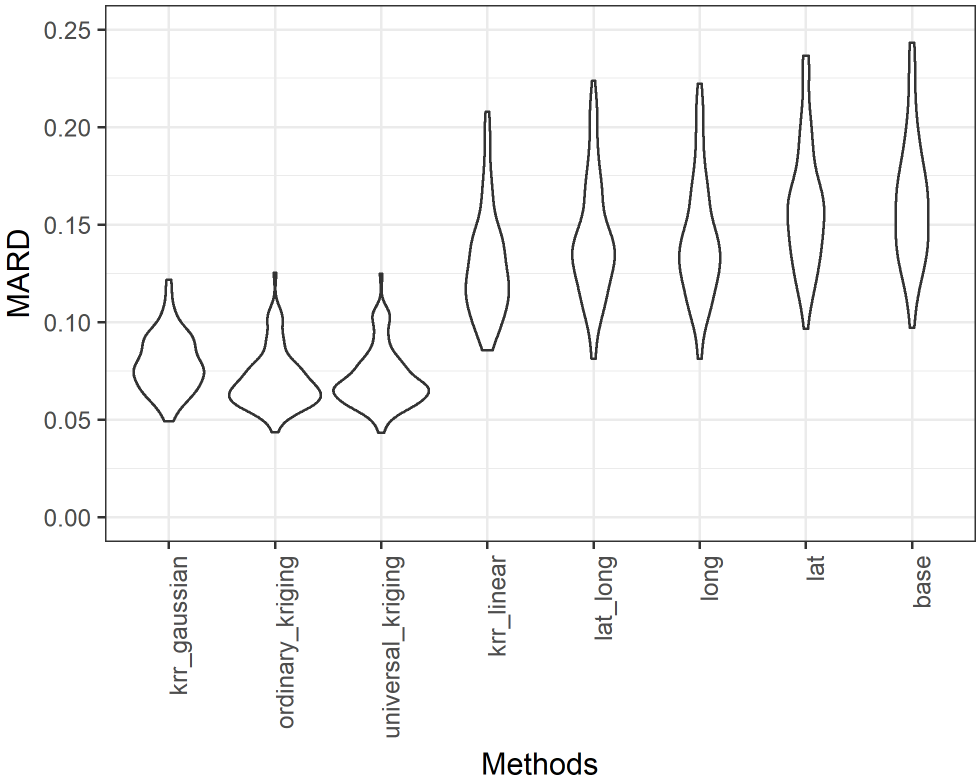
CHAPTER 3

Result Analysis and Discussion

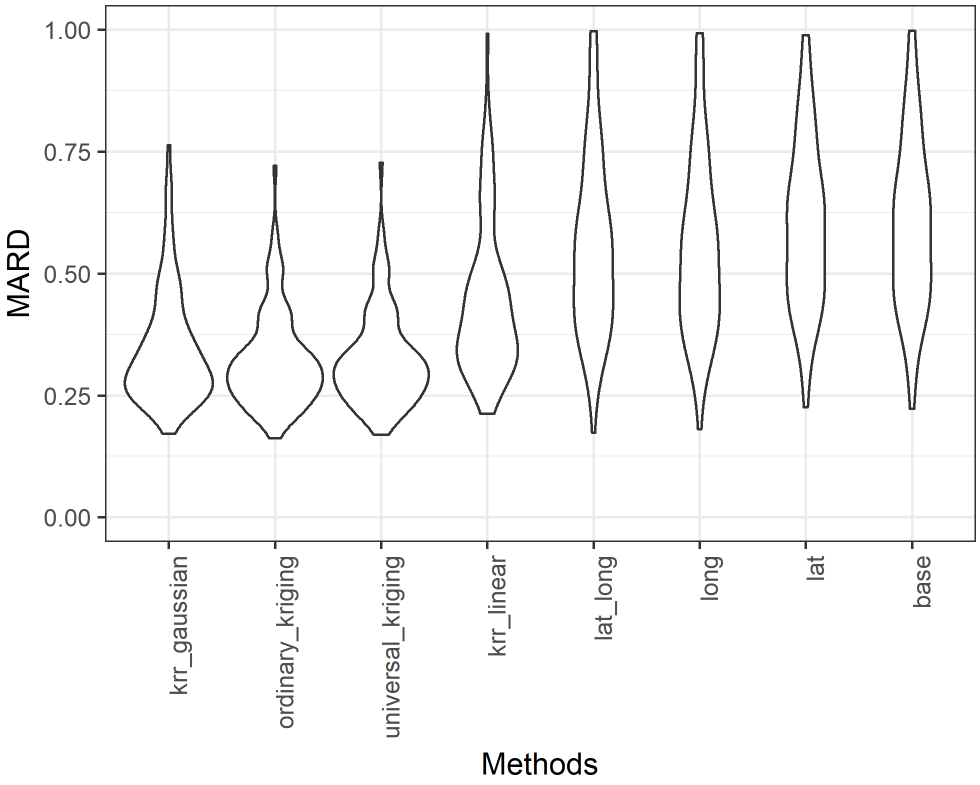
The prediction performance is evaluated by the MARD distribution form Non-replacement Bootstrap. Both EDPs, PFA and PSDR, are evaluated for each considered earthquake. *base* is the baseline model, *lat*, *long*, and *lat long* is three OLS model with corresponding predictors. *krr gaussian* and *krr linear* is two KRR model using Guassian and linear Kernel. *ordinary kriging* and *universal kriging* is two Kriging models with no trend and linear trend respectively.

3.1 Bias and Variance of Prediction

MARD distribution of PSDR and PFA of 4-story building is shown in Figure 3.1 which shows prediction performance for PSDR is more robust than that of PFA. In Figure 1.4, it shows consistent observation that PFA is probably less explained by geo-spatial features compared to PSDR. Regardless of which predictors to use, linear models have the highest error median at 18% for PSDR and 60% for PFA and the violin plot for linear models MARD is quite symmetric which indicates that they are unbiased with high variance. This result is consistent with nature sense that linear models without regularization turns to have high variance in testing error. This can be derived in the following.



(a) PSDR



(b) PFA

Figure 3.1: MARD Distribution of 4-Story Building Subjected to Northridge Earthquake

$$
\begin{aligned}
testing\,error &= E_{\hat{\beta},\tilde{\epsilon}}\big[\|Y+\tilde{\epsilon}-X\hat{\beta}\|^2|X\big] \\
&= E_{\hat{\beta},\tilde{\epsilon}}\big[\|Y-X\hat{\beta}\|^2+\tilde{\epsilon}^2|X\big] \\
&= E_{\hat{\beta},\tilde{\epsilon}}\big[\|Y-X\beta_{best}\|^2+\|X\beta_{best}-X\hat{\beta}\|^2|X\big] \\
&= E_{\hat{\beta},\tilde{\epsilon}}\big[\|Y-X\beta_{best}\|^2|X\big]+E_{\hat{\beta},\tilde{\epsilon}}\big[\|\hat{\beta}-\beta_{best}\|^2|X\big] \\
&= Bias^2_{model}+MSE
\end{aligned}
\tag{3.1}
$$

$$
\begin{aligned}
MSE &= E_{\hat{\beta},\tilde{\epsilon}}\big[\|\hat{\beta}-\beta_{best}\|^2|X\big] \\
&= E\big[\|\hat{\beta}-E(\hat{\beta})+E(\hat{\beta})-\beta_{best}\|^2\big] \\
&= E\big[\|\hat{\beta}-E(\hat{\beta})\|^2\big]+E\big[\|E(\hat{\beta})-\beta_{best}\|^2\big] \\
&= Var+Bias^2_{\beta}
\end{aligned}
\tag{3.2}
$$

In Equation 3.1, the test error is decomposed into a bias term from the model and a Mean Squared Error (MSE) term which is further decomposed into variance and estimator bias (Equation 3.2). By adding back a noise term, testing error of linear model is decomposed into:

1. Bias of model: can be improved by increasing model complexity;
2. Bias of estimator: can be improved by regularization of the model;
3. Variance of estimator: can be improved by regularization of the model;
4. Noise.

The above discussion provides some evidences for poor performance from linear model and supports the idea to add regularization for model complexity as well as use nonlinear models which can both be achieved by KRR. Figure 3.1 shows KRR with linear Kernels, or ridge regression, has slight better performance than that of all linear models indicating that bias and variance of estimator is reduced by regularization. KRR with Gaussian Kernel further reduces testing error significantly which suggests that bias of model is reduced. As a result, the prediction error is mainly caused by complex data structure and can be improved through

using complex models such as KRR with Gaussian Kernel or the Kriging models. There is an inherent relationship between them which is discussed in the following section.

3.2 Kriging and Kernel Ridge Regression

Kriging model is equivalent to Gaussian process regression while μ is assumed to be zero (often referred as simple Kriging) which is essentially a Bayesian framework of KRR with Gaussian Kernel. We first make the connection between Kriging and Gaussian Process (GP) regression. In Equation 2.6, Ordinary Kriging model is defined to be a constant mean μ plus a location dependent Gaussian term $\delta(s)$. Assuming the constant mean is zero, we have:

$$f(s) = \delta(s) = \hat{f}(X) \sim \mathcal{N}(0,\, \Sigma) \tag{3.3}$$

Now, let's make connection from Equation 3.3 to KRR by project X into a N-dimensional space which is much greater than p such that we have:

$$\hat{f}(X) = \phi(X)^T w, \quad Y = \hat{f}(X) + \epsilon, \quad \epsilon \sim \mathcal{N}(0,\, \sigma_n^2) \tag{3.4}$$

The likelihood of residual ϵ to follow normal distribution is:

$$p(Y|X,\, w) = \frac{1}{(2\pi\sigma_n^2)^2} exp\bigg(- \frac{1}{2\sigma_n^2}|Y - \phi(X)^T w|^2 \bigg) = \mathcal{N}(\phi(X)^T w,\, \sigma_n^2 I) \tag{3.5}$$

The Bayesian formalism needs a belief knowledge on the model parameters w so we assume that $w \sim \mathcal{N}(0,\, I_N)$. According to Bayes's rule,

$$p(w|X, Y) = \frac{p(Y|X,\, w)p(w)}{p(Y|X)} \tag{3.6}$$

The normalizing constant, also known as marginal likelihood can be ignored as it is independent to weights. The posterior is then:

$$p(Y|X,\, w) \propto exp\bigg[- \frac{1}{2}(w - \hat{w})^T (\frac{1}{\sigma_n^2}\phi(X)\phi(X)^T + \Sigma_N^{-1})(w - \hat{w}) \bigg] \tag{3.7}$$

Which is a Gaussian distribution with mean $\hat{w} = \sigma_n^{-2}(\sigma_n^{-2}\phi(X)\phi(X)^T + \Sigma_N^{-1})^{-1}\phi(X)Y$. Let the weight covariance matrix Σ_N^{-1} be $\tau^2 I_N$ and $K = \phi(X)\phi(X)^T$, use Equation 3.3 and posterior predictive, the prediction equation becomes:

$$\hat{f}(x) = (K + \frac{\tau^2}{\sigma_n^2} I_N)^{-1} Y k(x, X) \tag{3.8}$$

This is same form as KRR. Back to Figure 3.1, KRR with Gaussian distribution is same as Gaussian process regression, which is a simple Kriging model that is not very different from ordinary Kriging on the same figure. Universal Kriging with linear trend is very similar to the above two as well. It is also shown that error from the complex models, KRR with Gaussian and Kriging, distributes unsymmetrically with shorter confidence interval. This suggests that the error distribution has shorter tails which is preferred compared to longer tail linear models.

3.3 Comparison between Buildings

Besides the result presented in Figure 3.1 from 4-story building subjected to Northridge earthquake, 2-story building result is shown in Figure 3.2. It is observed that the 2-story building has higher error distributed over a larger confidence interval for both PFA and PSDR for both linear models and complex models. A summary figure which plots number of stories vs. mean of MARD is shown in Figure 3.3. The first observation is that baseline model is almost identically poorly performed as OLS using latitude suggesting that the most of the responses content is not able to be explained linearly by latitude along. It is also shown that OLS with both geographical coordinates has the same level of performance as OLS with longitude only which is consistent with previous observation for both PFA and PSDR. The ridge regression, or KRR with linear Kernel, is 60% better overall compared to base model which indicates that model regularization contributes a lot to testing accuracy for this dataset. The best models are KRR with Gaussian Kernel and Ordinary Kriging where both models outperform all linear models.

Comparing between Figure 3.3a and Figure 3.3b, as number of stories increases, predictions of PSDR become better, i.e., mean of MARD from KRR with Gaussian Kernel drops from 8.3% to 6% as number of stories increases from 2 to 20. However, the trend is opposite in PFA. The best performing model, Ordinary Kriging and KRR with Gaussian, has mean of MARD at 75% for 20-story building. Therefore, the proposed prediction scheme is not working well with PFA. This might be caused by high variance in PFA data when it comes

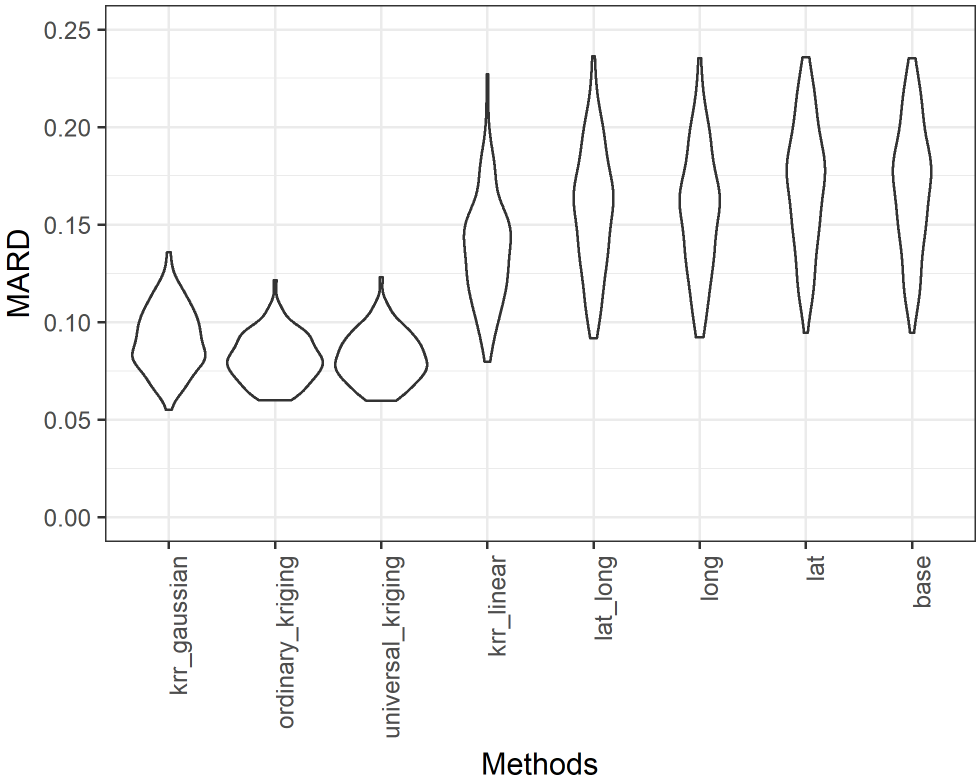
to taller buildings.

3.4 Comparison between Earthquakes

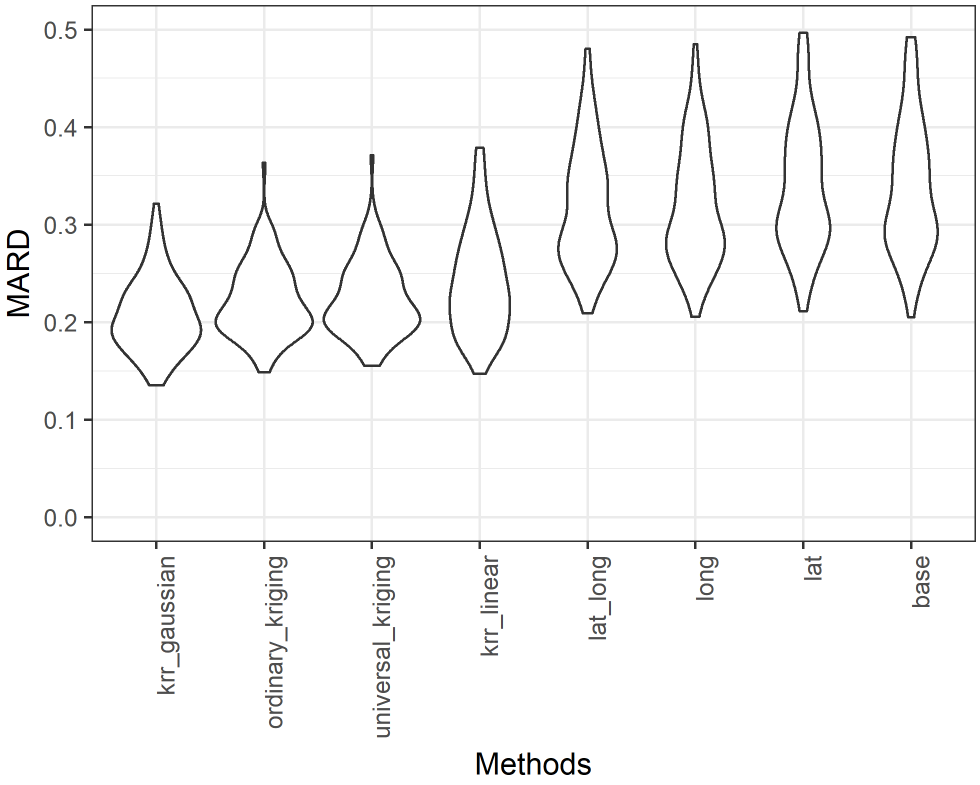
Besides building responses from Northridge, another two earthquakes, 1999 Chi-Chi and 2000 Tottori, are used to evaluate models' prediction performance for different seismic events. However, it should be noted that each seismic event has its unique characteristics and complexity given location and depth of epicenter, energy released, transferring path and soil conditions. These spatial dissimilarities caused ground motion on-site to be different. The proposed methodology relies on that these spatial dissimilarities are implicitly considered in the training data given they are from the same event. The proposed framework is limited in its application in that this type of model can only be applied in cases where, right after the occurrence of a particular event, an interpolation model is trained using the available response data from instrumented buildings and used to infer responses in un-instrumented buildings, all within the same event. In other words, the statistical models developed in this study can only be applied to the same event that generated the training data. As a result, comparison between earthquakes is to test its prediction performance given different events rather than validate its generality as shown in Figure 1.4 that EDPs from each event have very different statistics such as maximum and mean. A more generalized model incorporates earthquake characteristics should be used when it comes to a general model.

Figure 3.4 and 3.5 shows mean MARD over building height for each method for Chi-Chi and Tottori respectively. Comparing across different events, MARD over height does not have a visually clear trend. However, trend across different methods is very much consistent with that nonlinear models have highest prediction accuracy followed with linear model with regularization and linear models except for Tottori earthquake in which KRR with linear Kernel has worst performance. Observing that data from Tottori earthquake has the highest coverage area compared to the other two, the relative poor performance from regularized model might be due to insufficient sample amount to represent actual data distribution.

In general, the nonlinear models for PSDR has better prediction performance than PFA

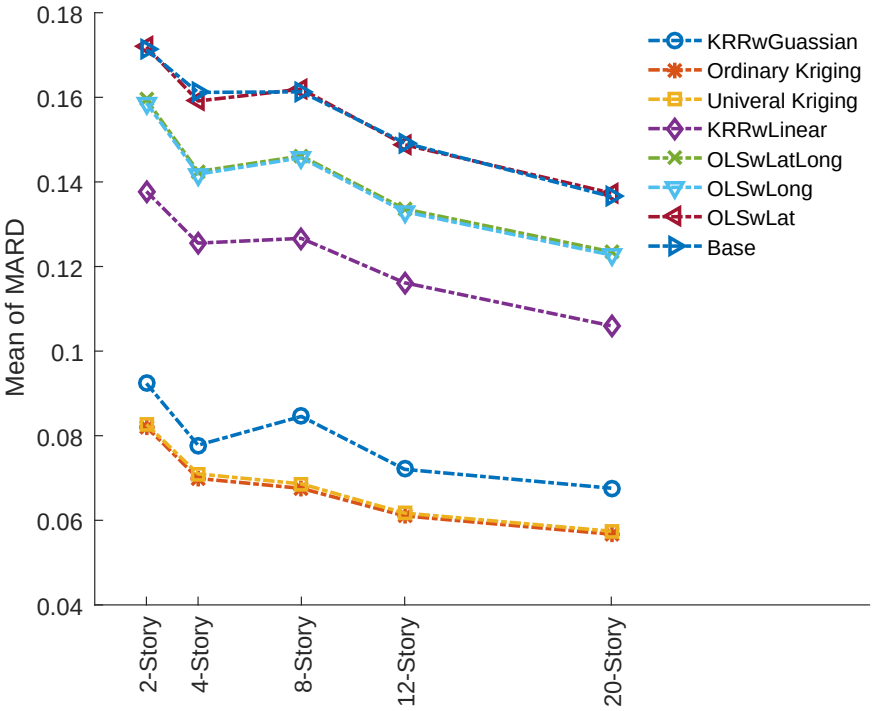


(a) PSDR

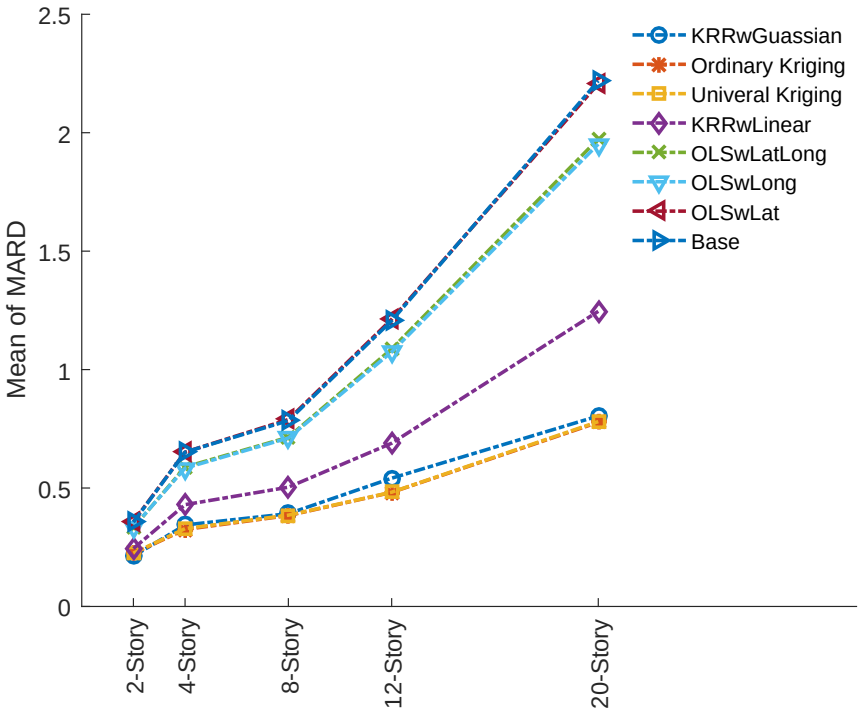


(b) PFA

Figure 3.2: MARD Distribution of 2-Story Building Subjected to Northridge Earthquake

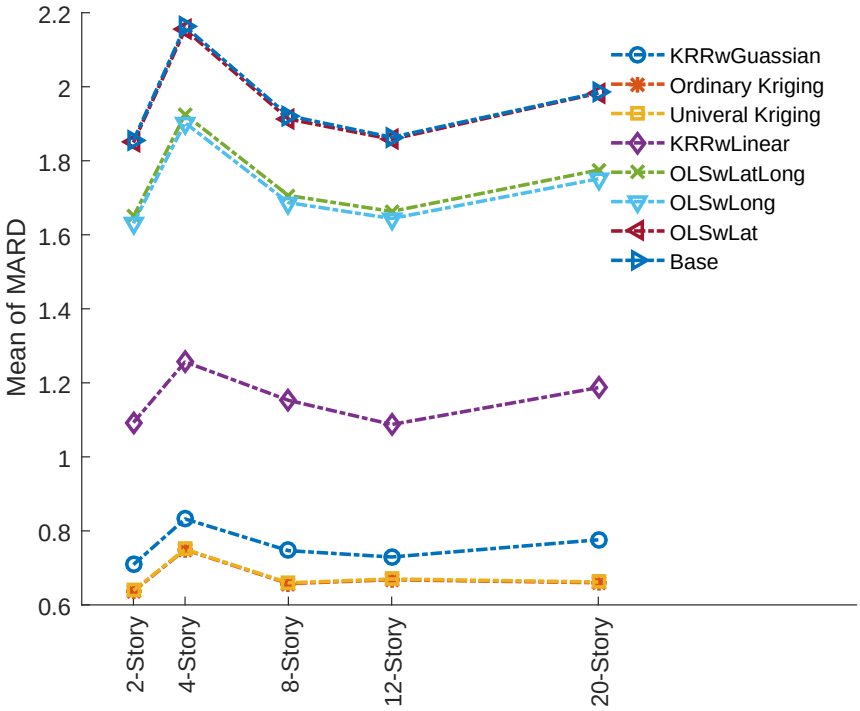


(a) PSDR

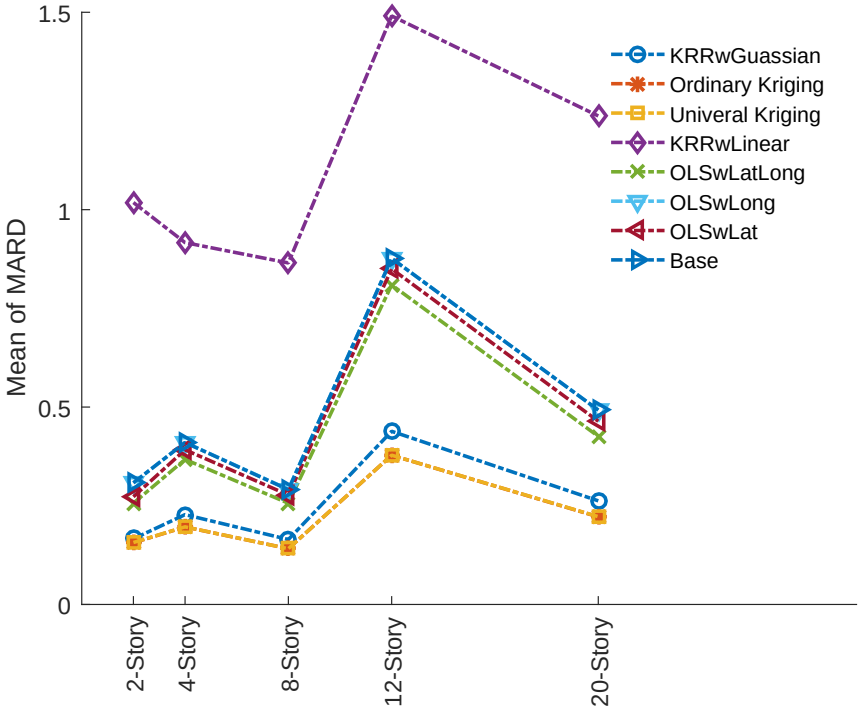


(b) PFA

Figure 3.3: Mean MARD Trend of All Buildings Subjected to Northridge Earthquake



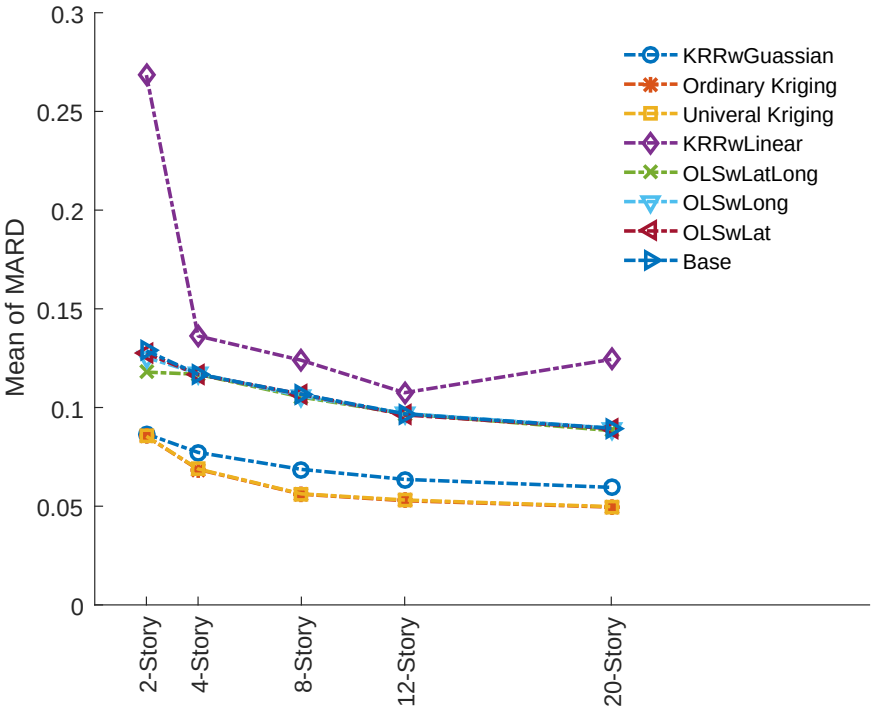
(a) PSDR



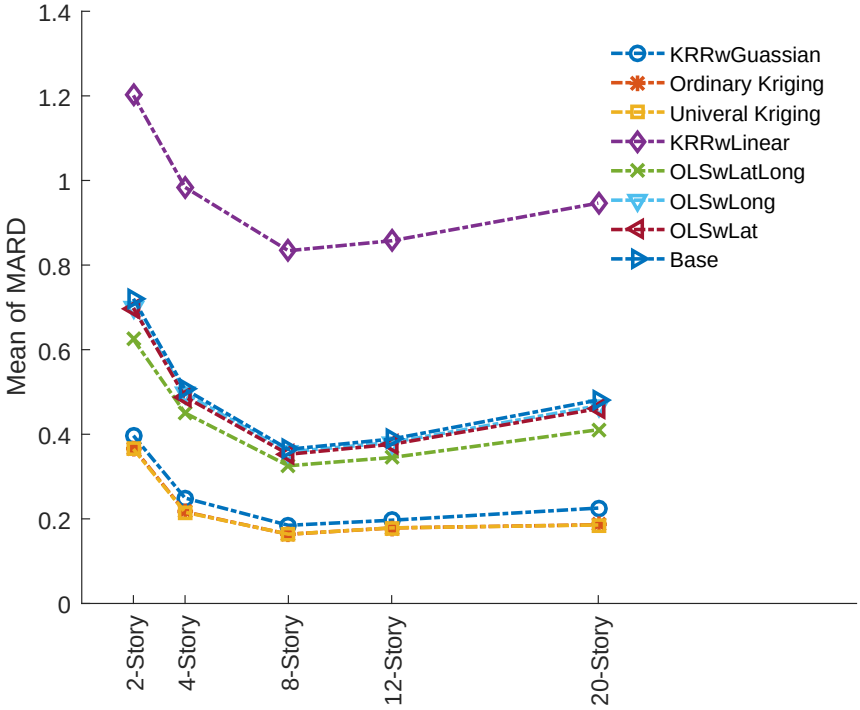
(b) PFA

Figure 3.4: Mean MARD Trend of All Buildings Subjected to Chi-Chi Earthquake

while several cases error is very high, i.e., about 40% MARD of PSDR from 12-story biulding subjected to Chi-Chi earthquake and 60% of PFA from 20-story building subjected to Northridge earthquake. Although different earthquake scenarios are unique from each other considering the differences in epicenter location, soil and rock properties that affect wave transfer, and fault types which causes different frequency characteristics, the general trend of building response data are expected to be quite consistent due to buildings severing as physical filters. Therefore, the observed inconsistencies of prediction performance between the three earthquakes are largely dependent on available sample distributions. The prediction result is expected to be more consistent in terms of accuracy levels given more representative data.



(a) PSDR



(b) PFA

Figure 3.5: Mean MARD Trend of All Buildings Subjected to Tottori Earthquake

CHAPTER 4

Summary and Conclusions

This study investigates several point estimator interpolation schemes for building seismic responses subjected to an earthquake over a geo-region. Data is retrieved from nonlinear response history analysis corresponding to five concrete moment frame buildings from 2-story to 20-story subjected to three historical recorded earthquakes, 1994 Northridge, 1999 Chi-Chi and 2000 Tottori respectively. Two types of building response EDPs are considered, PFA and PSDR which are both highly correlated with building damage.

The exponential fit trend of PSDR and PFA over epic distance indicates that both EDPs are strongly correlated with geo-spatial features while PSDR has a higher decreasing rate across all earthquakes. Seven different models are used to predict EDPs in addition to a baseline naive model. A frequentist approach, non-replacement Bootstrap, is applied to examine each model's prediction performance. Despite it is observed that OLS models are likely unbiased, it is shown to be outperformed in all earthquakes and building types by nonlinear models as a result of high variance in terms of prediction error. By adding regularization term to linear model, variance of prediction error decreases as observed in result of KRR with linear Kernel. The best performance comes from KRR with Gaussian Kernel and Kriging model which both includes nonlinear transformation of geo-spatial features which shows that response patterns are nonlinearly distributed in high dimension space formulated by geo-spatial features, latitude and longitude. KRR with Gaussian Kernel and Kriging model shows very similar result in all analysis. A gentle process is used to connect them through Gaussian process regression.

It is shown that nonlinear models have shorter tails in error distribution compared to linear models which indicates better stability and less variance. It can be concluded that

nonlinear models are superb compared to linear models.

Model performance is evaluated between buildings and earthquakes. Several discrepancies are observed across the three events, regularized linear model does not necessarily outperform OLS models. Prediction accuracy, measured by MARD, turns to have different trend over building height for each earthquake too. These observations are probably caused by uniqueness of each earthquake characteristics as well as site distribution over the geo-region. However, the general trend of building response are expected to be consistent due to buildings severing as filters. The current inconsistency between different earthquakes are mainly in accuracy levels which should be overcome by training with more representative data. And the proposed method to interpolate building seismic responses should only be applied to individual earthquake. Another solution is to develop a generalized model which potentially could be applied to any event.

REFERENCES

[1] null null. *Minimum Design Loads for Buildings and Other Structures*. American Society of Civil Engineers, asce/sei 7-10 edition, 2013.

[2] Michel Bruneau and Andrei Reinhorn. Overview of the resilience concept. In *Proceedings of the 8th US National Conference on Earthquake Engineering*, volume 2040, pages 18–22, 2006.

[3] AK Pandey and M Biswas. Damage detection in structures using changes in flexibility. *Journal of sound and vibration*, 169(1):3–17, 1994.

[4] John B Kosmatka and James M Ricles. Damage detection in structures by modal vibration characterization. *Journal of Structural Engineering*, 125(12):1384–1392, 1999.

[5] Farzad Naeim, H Lee, H Bhatia, S Hagie, and K Skliros. Csmip instrumented building response analysis and 3-d visualization system (csmip-3dv). In *Proceedings of the SMIP-2004 Seminar*, 2004.

[6] MP Limongelli. Seismic health monitoring of an instrumented multistory building using the interpolation method. *Earthquake Engineering & Structural Dynamics*, 43(11):1581–1602, 2014.

[7] Christophe Loth and Jack W Baker. A spatial cross-correlation model of spectral accelerations at multiple periods. *Earthquake Engineering & Structural Dynamics*, 42(3):397–417, 2013.

[8] D Jared DeBock, Jack W Garrison, Kevin Y Kim, and Abbie B Liel. Incorporation of spatial correlations between building response parameters in regional seismic loss assessment. *Bulletin of the Seismological Society of America*, 104(1):214–228, 2014.

[9] Timothy D Ancheta, Robert B Darragh, Jonathan P Stewart, Emel Seyhan, Walter J Silva, Brian S-J Chiou, Katie E Wooddell, Robert W Graves, Albert R Kottke, David M Boore, et al. Nga-west2 database. *Earthquake Spectra*, 30(3):989–1005, 2014.

[10] Silvia Mazzoni, Frank McKenna, Michael H Scott, and Gregory L Fenves. The open system for earthquake engineering simulation (opensees) user command-language manual. 2006.

[11] Curt B Haselton and Gregory G Deierlein. *Assessing seismic collapse safety of modern reinforced concrete moment frame buildings*. PhD thesis, Stanford University, 2007.

[12] Noel Cressie. Spatial prediction and ordinary kriging. *Mathematical geology*, 20(4):405–421, 1988.

[13] Noel AC Cressie. *Statistics for spatial data*. Wiley Online Library, 1993.

[14] Han Sun, Henry Burton, Yu Zhang, and John Wallace. Interbuilding interpolation of peak seismic response using spatially correlated demand parameters. *Earthquake Engineering & Structural Dynamics*.