UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
STATISTICAL FRAMEWORKS FOR IMPROVED SEISMIC DEMAND ANALYSIS OF ORDINARY STANDARD BRIDGE STRUCTURES

Permalink
https://escholarship.org/uc/item/4q75r7zp

Author
Fayaz, Jawad

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Performance-Based Earthquake Engineering Lab,
Department of Civil and Environmental Engineering

STATISTICAL FRAMEWORKS FOR IMPROVED SEISMIC DEMAND ANALYSIS OF ORDINARY STANDARD BRIDGE STRUCTURES

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in

Civil Engineering

by

Jawad Fayaz

Dissertation Committee:

Dr. Farzin Zareian (Chair)
Dr. Farzad Naeim
Dr. Joel Lanning
Dr. Mo Li

2021

DEDICATION

To my parents and brothers in recognition of the incredible amount of effort they have employed to support my graduate dream, especially to my mother, Shahzana, who has been an unconditional source of encouragement through all transitions of my life.

"Be uncommon among uncommon." – David Goggins

"He who has a 'why', can bear almost any 'how'." – Friedrich Nietzsche

"PhD is a journey from not knowing that you don't know anything to knowing that you don't know anything."

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

CURRICULUM VITAE

EDUCATIONAL BACKGROUND

University of California- Irvine, USA	**January 2018 –**
Doctor of Philosophy (PhD) – Structural and Earthquake Engineering	**June 2021**
University of California- Irvine, USA	**September 2016 –**
Master of Science (MS) – Structural Engineering	**March 2018**
R.V. College of Engineering- Bengaluru, India	**August 2012 –**
Bachelor of Engineering (BE) – Civil Engineering	**July 2016**

PATENT

Jawad Fayaz, Adarsha C. Y., T. Raghavendra, K. Natarajan and B.C. Udayashankar. "Waste Plastic Incorporated Fiber Reinforced Polymer Bars", *Indian Patent # 354719, Granted 12/2020*

PUBLICATIONS

Journal Articles:

12. **Jawad Fayaz** and Farzin Zareian (2nd round of review). "An Efficient Algorithm to Simulate Site-Based Ground Motions that match a Target Spectrum". *Earthquake Engineering and Structural Dynamics.*
11. **Jawad Fayaz** and Farzin Zareian (accepted). "Response Prediction Model (RPM) for Column Drift Ratio of Box-Girder Non-Skewed Seat-Type Ordinary Bridge Structures". *Journal of Bridge Engineering.*
10. **Jawad Fayaz**, Sanaz Rezaeian, and Farzin Zareian (2021). "Evaluation of simulated ground motions using probabilistic seismic demand analysis: CyberShake (ver. 15.12) simulations for Ordinary Standard Bridges". *Soil Dynamics and Earthquake Engineering.* Vol. 141, 106533.
9. **Jawad Fayaz**, Sarah Azar, Mayssa Dabaghi, and Farzin Zareian (2021). "An Efficient Algorithm to Simulate Hazard-Targeted Site-Based Synthetic Ground Motions". *Earthquake Spectra*. Vol. 37, Issue 2.
8. **Jawad Fayaz**, Yijun Xiang, and Farzin Zareian (2020). "Generalized Ground Motion Prediction Model (GGMPM) Using Hybrid Neural Networks". *Earthquake Engineering and Structural Dynamics*, Vol. 50, Issue 6.
7. **Jawad Fayaz**, Sarah Azar, Mayssa Dabaghi, and Farzin Zareian (2020). "Methodology for Validation of Simulated Ground Motions for Seismic Response Assessment: Application to Cybershake Source-Based Ground Motions". *Bulletin of the Seismological Society of America*, Vol. 111, No. 1.
6. **Jawad Fayaz**, Mayssa Dabaghi, and Farzin Zareian (2020). "Utilization of Site-Based Simulated Ground Motions for Hazard-Targeted Seismic Demand Estimation: application for Ordinary Bridges in Southern California". *Journal of Bridge Engineering*, Vol. 25, Issue 11.

5. **Jawad Fayaz**, Miguel Riquelme, and Farzin Zareian (2020). "Sensitivity of The Response of Box-Girder Seat-Type Bridges to the Duration of Ground Motions arising from Crustal and Subduction Earthquakes". *Engineering Structures*, Vol. 219, 110845.
4. Pablo Torres, **Jawad Fayaz**, and Farzin Zareian (2020). "Strength Resistance Factors for Seismic Design of Exposed Based Plate Connections in Special Steel Moment Resisting Frames". *Earthquake Spectra*, Vol. 36, Issue 2.
3. **Jawad Fayaz** and Farzin Zareian (2019). "Reliability Analysis of Steel SMRF and SCBF Structures Considering the Vertical Component of Near-Fault Ground Motions". *Journal of Structural Engineering*, Vol. 145, Issue 7.
2. **Jawad Fayaz**, Adarsha C. Y, T. Raghavendra, K. Natarajan, and B.C. Udayashankar (2019). "Structural Behavior of R.C. Members Reinforced with Waste Plastic incorporated Glass Fibre Reinforced Polymer (WPGFRP) Rebars". *Journal of Building Engineering*, Vol. 23.
1. T. Raghavendra, H. Siddanagouda, **Jawad Fayaz**, Adarsha C.Y, and B.C. Udayashankar (2018). **"**Mitigation of H2S Emissions by Recycling Discarded Gypsum Wall Boards in CLSM". *Global Journal of Researches in Engineering*, Vol. 18, Issue 1-E.

Conference Proceedings:

8. **Jawad Fayaz** and Farzin Zareian (2022). "Response Prediction Models for Standard Ordinary Bridges for the Assessment and Design of the Transportation Lifelines". ASCE Lifelines. UCLA, USA.
7. **Jawad Fayaz**, Yijun Xiang, and Farzin Zareian (2020). "Ground Motion Spectral Estimation using Recurrent Neural Networks (RNN)". 17th World Conference on Earthquake Engineering (17WCEE). Sendai, Japan.
6. **Jawad Fayaz**, Mayssa Dabaghi, and Farzin Zareian (2020). "Guidelines for Utilization of Site-Based Simulated Ground Motions for Performance-Based Earthquake Engineering of Ordinary Bridges". 17th World Conference on Earthquake Engineering (17WCEE). Sendai, Japan.
5. Alessandro Cardoni, **Jawad Fayaz**, Gian P. Cimellaro, and Farzin Zareian (2020). "Utilizing Simulated Ground Motions to Quantify Highway Network Performance during Moderate Seismic Events". 17th World Conference on Earthquake Engineering (17WCEE). Sendai, Japan.
4. **Jawad Fayaz**, Yijun Xiang, and Farzin Zareian (2019). "Performance Assessment of Bridges Under A Sequence of Seismic Excitations". *Computational Methods in Structural Dynamics and Earthquake Engineering* (COMPDYN 2019), Crete Island, Greece.
3. **Jawad Fayaz**, Mayssa Dabaghi, and Farzin Zareian (2019). "Probabilistic Analytical Benchmarking of Source-Based and Site-Based Ground Motion Simulation Models". *13th International Conferences on Applications of Statistics and Probability in Civil Engineering* (ICASP13), Seoul, South Korea.
2. **Jawad Fayaz**, Miguel Riquelme, and Farzin Zareian (2019). "Sensitivity of The Response of Box-Girder Seat-Type Bridges to the Duration of Ground Motions Arising from Subduction Earthquakes". *XII Chilean Conference on Seismology and Earthquake Engineering* (ACHISINA), Valdivia, Chile.
1. T. Raghavendra, H. Siddanagouda, **Jawad Fayaz**, Adarsha C.Y, and B.C. Udayashankar (2016). "Performance of ternary binder blend containing cement, waste gypsum wallboards and blast furnace slag in CLSM". *Procedia Engineering*, Vol. 145.

Technical Reports:

5. "Validation and Utilization of Physics-Based Simulated Ground Motions for Bridge Performance Assessment" (under-review). Pacific Earthquake Engineering Research (PEER) Award No. 1146-NCTRFZ.
4. "Critical assessment of probabilistic seismic demand analysis of bridge structures using CyberShake simulations" (2020). Southern California Earthquake Center (SCEC) Award No. 118062867, SCEC ID 19114.
3. "Validation of Caltrans Ordinary Bridge Modeling Approach using Bayesian State and Parameter Estimation Method" (2020). California Strong Motion Instrumentation Program (CSMIP) Agreement No. 1018-568.
2. "Guidelines for Ground Motion Modelling for Performance-Based Earthquake Engineering of Ordinary Bridges" (2019), Caltrans Final Report No. 65A0647.
1. "Preliminary Report on Engineering and Geological Effects of the July 2019 Ridgecrest Earthquake Sequence" (2019). Reconnaissance Report GEER-064NSF.

AWARDS AND HONORS

i. Recipient of **2020 Medhat Haroun Engineering Fellowship** awarded by Department of Civil and Environmental Engineering, University of California- Irvine, USA.
ii. Recipient of **2020 Phi Beta Kappa International Scholarships** awarded by Phi Beta Kappa Alumni Association, USA.
iii. Recipient of **2020 Graduate Dean's Dissertation Fellowship** awarded by Graduate Division, University of California- Irvine, USA.
iv. Recipient of **2019 Henry Samueli Endowed Fellowship** awarded by Henry Samueli School of Engineering, University of California- Irvine, USA.
v. Recipient of **2019 Structural Engineering Association of Southern California (SEAOSC) Foundation Scholarship** awarded by Structural Engineering Association of Southern California (SEAOSC), California, USA.
vi. Recipient of **2019 Travel Grant Award** awarded by Graduate Student Affairs Office, University of California- Irvine, USA.
vii. Felicitated with **Achievement Award** by the Department of Civil Engineering at R.V. College of Engineering- Bengaluru, India in May 2016.
viii. Recipient of **2016 Most Innovative Project of RVCE** among all undergrad projects at R.V. College of Engineering- Bengaluru, India.

CONTACT LINKS

Website: https://jfayaz.github.io/
GitHub: https://github.com/jfayaz
LinkedIn: https://www.linkedin.com/in/jawadfayaz/
Research Gate: https://www.researchgate.net/profile/Jawad-Fayaz-2

ABSTRACT OF THE DISSERTATION

Statistical Frameworks for Improved Seismic Demand Analysis of Ordinary Standard Bridge Structures

by

Jawad Fayaz

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2021

Professor Farzin Zareian

Transportation networks of highway bridges are considered the lifelines of a community's infrastructure as they play a significant role in joining communities and serving as the first outlet during a calamity such as earthquakes. They are expected to sustain minor damage and maintain their functionality after major natural and human-made disasters. Observations from seismic events of the last three decades reveal that bridges designed according to seismic design codes demonstrate poor performance and occasionally undergo significant damages leading to major consequence on the affected societies. In the light of these effects, numerous studies have showcased that one of the primary reasons for the unexpected performance of the bridge structures is the improper estimation of the expected seismic demands during the design and analysis phase.

With the evolution of risk, reliability, and hazard analysis in quantifying the seismic vulnerability of structures, the seismic structural design procedures are continuously updating to develop methodologies that achieve more accurate estimations of the structural demands corresponding to the target hazard levels. The most widely used conventional procedure is to conduct Incremental Dynamic Analysis (IDA) by selecting and scaling seismic ground motion records to attain a scalar Intensity Measure (IM), such as spectral acceleration (S_a), associated with a target hazard level of the IM hazard curve. The scaled ground motions are then used to conduct the Non-Linear Time-History Analysis (NLTHA) of finite-element models of the structures, and the obtained response value, *i.e.,* Engineering Demand Parameters (EDP), are then utilized for developing EDP hazard curves by integrating the EDP-IM data over the IM hazard curve. Numerous studies are conducted worldwide describing the limitations of this type of analysis, such as sufficiency and efficiency of

single scalar IM as the target, unrealistic scaling of recorded ground motions, type of IDA, ground motions not being site-specific, etc.

The research effort presented herein firstly proposes a supplementary Generalized Ground Motion Prediction Model (GGMPM) that can be used to construct a vector-based (29 x 1, representing intensity-, duration- and frequency- the content of ground motions) IM hazard curves which can be then used to select and scale ground motions targeted to a vector of correlated IMs. The proposed GGMPM consists of Recurrent Neural Network (RNN) and inter-event and intra-event covariance functional forms developed using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). Though the proposed GGMPM framework provides a swift tool to the engineering community to tackle some of the deficiencies of current IDA methods, this research work further develops EDP hazard curves by conducting large-scale NLTHA of the four bridge structures using synthetic ground motions of *Site-Based* (*i.e.,* DRD model) and *Physics-Based* (*i.e.,* CyberShake 15.12) simulation models. For the five sites in the southern California region, rupture variations and their respective probabilities are obtained from Uniform California Earthquake Rupture Forecast, Version 2 (UCERF2) database, which are used to simulate *Site-Based* synthetic ground motions using the DRD simulation model for a time-span of 100,000 years. Similarly, *Physics-Based* synthetic ground motions simulated for the CyberShake 15.12 study representing a time-span of 200,000 years are selected. This leads to around ~20,000 *Physics-Based* ground motions and ~10,000 *Site-Based* ground motions for each site. These simulated ground motions are then used to conduct NLTHA of the four OSBs to obtain simulation-based- *Site-Based* and *Physics-Based* EDP hazard curves. The two types of simulation-based EDP hazard curves are compared against the conventional *IDA-Based* EDP hazard curves and various regression models are proposed to transform *IDA-Based* EDPs to simulation-based EDPs. Finally, the three types of EDP hazard curves (*i.e., Site-Based*, *Physics-Based,* and *IDA-Based*) for the four bridges and five sites are provided for comparison.

CHAPTER 1

INTRODUCTION

1.1 Overview

This dissertation proposes different statistical frameworks that can be utilized to conduct an improved structural seismic demand assessment. In this research, comprehensive studies are conducted to facilitate engineers with statistical tools to obtain more accurate "site-specific" estimations of the seismic demands of structures, in particular bridge structures. The current methodologies of structural seismic demand assessment do not explicitly include the regional seismicity and are highly generalized. This study is a step towards conducting "site-specific" assessment of structures which implies that the possible seismic conditions and events of the site are considered while conducting structural demand analysis. This research aims at providing the engineering community with the frameworks for better seismic hazard modeling in the context of Performance-based Earthquake Engineering for standard ordinary bridges, and a complete set of testbeds to demonstrate the key features of the proposed technology. In general, the research study presented herein have three-fold goals: 1) improve the methods of selecting hazard-targeted ground motions, 2) utilize *Site-Based* synthetic ground motions for bridge demand analysis, and 3) utilize *Physics-Based* synthetic ground motions for bridge demand analysis. Furthermore, to expedite the engineering community various open source softwares are provided along with the frameworks.

To attain the first goal, a data-driven non-parametric Generalized Ground Motion Prediction Model (GGMPM) is developed that predicts an internally correlated vector of 29 Intensity Measures (denoted as **IM**) including geomean of Arias Intensity (I_a), geomean of Cumulative Absolute Velocity (*CAV*), geomean of Significant Duration (D_{5-95}) (denoted as $I_{a_{geom}}$, CAV_{geom}, and $D_{5-95_{geom}}$, respectively), and *RotD50* Spectral Acceleration (S_a) at 26 periods (for 5% damped oscillator) (Fayaz *et al.,* 2020d). To incorporate the higher-order dependencies among the components of **IM**, the framework of Recurrent Neural Network (RNN) is adopted along with optimization techniques of the evolution strategy. In particular, Long-Short-Term-Memory (LSTM) network is used

1

to estimate **IM** using the seismic source and site parameters which describe the physics of the rupture and the site characteristics, as inputs. A sensitivity analysis is conducted in order to choose a subset of seismic source and site parameters that can be used as inputs to the RNN framework. Furthermore, the residuals of the RNN structure are carefully calibrated in terms of two 29×29 covariance matrices that estimate the intra-event and inter-event variabilities of the **IM**. This is done by using functional forms to estimate the variances and correlations whose coefficients are calibrated by maximizing the log-likelihood computed from the discrepancy between the **IM** estimated from the RNN framework and the recorded data. The predictions of the developed GGMPM are then compared against the current state-of-art Ground Motion Prediction Models (GMPMs).

To achieve the second goal, *Site-Based* stochastic models of Rezaeian and Der Kiureghian (2012) and Dabaghi and Der Kiureghian (2018) - denoted as the DRD simulation model, is utilized. Site-specific synthetic ground motion catalogs representing a time span of 100,000 years are simulated for five sites located in Southern California with diverse site and local seismicity conditions. This leads to around 10,500 ground motion simulations for each site with $M_w > 6.0$ occurring within a vicinity of 200 km. The simulated *Site-Based* ground motions are then used to conduct Non-Linear Time-History Analysis (NLTHA) of four Ordinary Standard Bridges (OSBs). The Engineering Demand Parameters (EDPs) obtained from the bridge structures are used to develop *Site-Based* EDP hazard curves for each bridge at each site. The developed *Site-Based* EDP hazard curves are compared against the EDP hazard curves obtained using conventional Incremental Dynamic Analysis (IDA) and a regression model is proposed that can be used to convert *IDA-Based* EDPs to *Site-Based* EDPs as a function of return period, site soil condition, and site basin depth. Furthermore, to assist design engineers, a reduced sample number of ground motions and increment angles that can statistically replicate the simulation set at the IM corresponding to the desired hazard level λ_{IM} (= 1/975 years) is proposed for the four bridge structures. This will assist engineers in making informed decisions in selecting an adequate number of ground motions and intercept angles for conducting the NLTHA of bridge structures (Fayaz *et al.,* 2020b).

For the third goal, a study is conducted using *Physics-Based* simulated ground motions of CyberShake ver. *15.12*. For the five sites located in Southern California, catalogs of *Physics-Based* simulated ground motions representing 200,000 years are obtained from the CyberShake database. CyberShake (Graves *et al.,* 2011) is a ground motion simulation tool that contains

2

ground-motion waveforms representing scenarios present in the Uniform California Earthquake Rupture Forecast, Version 2 (UCERF2) (Field *et al.*, 2009) in Southern California. At each site, the selected catalog contains roughly 20,000 ground motions from events with $M_w > 6.0$ occurring within a vicinity of 200 km. For each site, the catalog of simulated ground motions is used to conduct NLTHA of the four Ordinary Standard Bridges (OSBs). For each combination of bridge and site, column drift demands obtained from NLTHA data are used to generate EDP hazard curves. This data is compared with the results obtained from conventional methods where recorded ground motions are used to conduct Incremental Dynamic Analysis (IDA) of the bridge and the results are integrated over a ground motion Intensity Measure (IM) hazard curve for the site. Finally, site-specific predictive relations are proposed that correlate the ratio between the two types of EDPs with hazard level, site soil condition, and site basin depth. The proposed relations can assist engineers to scale the EDPs as per the effects of site and basin conditions (Fayaz *et al.,* 2021a).

1.2   Background

Performance-based Earthquake Engineering (PBEE) has been the focus of much research in the past two decades. Researchers have faced up to challenges such as: identifying performance objectives that are quantifiable and meaningful for engineers and stakeholders alike, identifying representations of ground motion intensity that well relate to seismic hazard at the site and the global bridge response characteristics, and many other advances whose enumeration does not fit into this brief introduction. The new challenge facing researchers is to transfer these sophisticated developments to the practicing engineers through applied but comprehensive procedures that facilitate implementation.

Despite fundamental advances in analytical modeling of bridge components, bridge structures, and transportation networks, current methods for design of Ordinary Standard Bridge (OSB) structures (e.g., SDC ver. 1.7, 2013; ver. 2.0, 2019, AASHTO, 2011) follow conventional methods and utilize tools that are incapable of addressing issues related to infrastructure resiliency directly. Current OSB design methods mainly focus on bridge component design with a holistic outlook on bridge structural behavior. On the other hand, design for infrastructure resiliency is founded on the concepts of Performance-based Earthquake Engineering (PBEE) and is capable of tackling issues such as bridge functionality, durability, lifecycle cost, and transportation network resiliency. As promoted by the Pacific Earthquake Engineering Research (PEER) center's PBEE framing

equation (see Equation 1.1, *G* functions represent complementary cumulative distribution functions, and *d* represents the derivative function), PBEE can be achieved by stringing four random variables of Intensity Measure (IM), Engineering Demand Parameter (EDP), Damage Measure (DM), and Decision Variable (DV). Conversely, current bridge design methods utilize conventional methods and tools to proportion bridge components for the mean value of EDP for an IM representing a target hazard level (i.e., λ_{IM}). Embracing PBEE concepts for the design of bridge structures requires a more in-depth characterization of EDP such that exercising Equation 1.1 is possible from λ_{IM} to λ_{DV} (i.e., an annual frequency of the desired realization of the DV). This study is a step in facilitating the use of PBEE in the design of bridge structures.

$$\lambda(DV) = \int \int \int G(DV|DM).dG(DM|EDP).dG(EDP|IM).d\lambda(IM) \tag{1.1}$$

$$\lambda_{EDP} = \sum G(EDP|IM)\Delta\lambda_{IM} \tag{1.2}$$

In Equation 1.1, λ(DV) is the desired realization of the DV (e.g., mean annual frequency of exceedance) and the *G* functions represent complementary cumulative distribution functions. For instance, for downtime performance (i.e., average annual downtime for a given bridge) one can complete the PBEE methodology as follows: intensity measures, IMs, (e.g., spectral acceleration at the first mode period of the structure, $S_a(T_1)$), are determined from seismic hazard analysis; relevant engineering demand parameters, EDPs, (e.g., deck rotation, column drift ratio) are predicted from structural analysis for given values of IMs (and representative ground motions); component/system damage states are developed from repair strategies and DMs fragility curves are developed for each component/system; and finally, predictions are made on DVs (i.e., total amount of downtime losses of the bridge). For engineering analysis, Equation 1.1 is only implemented up to EDP level to obtain EDP hazard curves (λ_{EDP}). This is shown in Equation 1.2 where the close form integral has been replaced by discrete summation for practical computation of the integral.

Engineering of bridges in accordance with Equation 1.1 up to the level of DV was originally tackled by Mackie *et al.* (2008). They utilized several simplifying assumptions on how to treat the correlation among damages in bridge components to evaluate DVs for typical bridges in California. Similar studies were pursued by Moschonas *et al.* (2009) who estimated seismic fragility curves for typical modern bridges in Greece; and by Kotsoglou and Pantazopoulou (2010) who developed a methodology for simulation and assessment of common over crossings subject

to transverse seismic excitations while considering soil-structure interaction effects. The common theme among these studies – and many more that are not reviewed here – is that they utilize nonlinear (dynamic or static) analyses using a set of ground motions to estimate the EDP|IM relationships (*a.k.a.* the EDP-IM curves). The variability in the EDP-IM curves is typically large. For example, Dutta and Mander (1998) calibrated the dispersion of statistical data obtained from actual bridge damage in the Loma Prieta (1989) and Northridge (1994) earthquakes and observed the dispersion in the EDP|IM relationship to be as high as 0.6. Kaviani *et al.* (2012) presented similar results based on simulations carried out over a wide variety of bridge model. Characterizing EDP|IM curves and reducing their variability are fundamental in developing accurate estimates of DVs using Equation 1.1. Accurate estimation of EDP|IM curves for a bridge structure requires two basic ingredients: (1) a rigorous finite-element model capable of capturing component-, and system-, level response from elastic to highly inelastic regimes, and (2) proper set of ground motion that represent seismic hazard at the location of the bridge (Mackie *et al.*, 2008; Kotsoglou and Pantazopoulou, 2010; Yoon *et al.*, 2019; Fayaz *et al.*, 2020a; Fayaz *et al.*, 2020b). The latter particularly requires selection and scaling of ground motion records, or using synthetic motions, for estimation of EDP-IM curves. Equation 1.1 is currently used up to the EDP level and integrated over the IM hazard curve to obtain the EDP hazard curves. The current methods of estimating EDPs by depending on "proper" IMs (Luco and Cornell, 2007) can be updated with the current increase in the computational resources. The overdependence on IMs representing the seismic hazard for developing EDP hazard curves (through integration) often leads to large variations in response analysis. Also, though there have been numerous research articles on what can be considered a good IM (Jalayer *et al.*, 2012; Eads *et al.*, 2015; Lin and Baker, 2015, Ebrahimian and Jalayer, 2020), no single IM has been developed that is independent of the structural application and can completely describe the behavior of the structure. *RotD50 S_a* (Boore 2010) is the current state-of-art IM that is adopted by all current GMPMs and is expressed as *RotDpp*, where *Rot* indicates the rotation of the two orthogonal components of the ground motion, *D* indicates the period dependency, and *pp* corresponds to the percentile value (mainly limited to the 50th percentiles, that is, the median value). For the sake of brevity, in this study, *RotD50* spectral acceleration at first mode period is used as the primary IM of ground motions and used interchangeably with $S_a(T)$ or S_a.

Current Caltrans design practice is based on utilizing design Acceleration Response Spectrum (ARS) with 5% probability of occurrence in 50 years (*i.e.* Return Period = 975 years). Conditioned

on using nonlinear response history analyses for assessing the behavior of a bridge structure, engineers select 7 actual or synthetic records and adjust them to the desired, and adjusted, ARS curves using linear scaling or spectral matching. Such matching has few shortcomings: (1) it does not directly incorporate near field effects such as directionality and incident angle, (2) fails to accurately address ground motion time-domain characteristics such as strong motion duration and velocity pules in the scaling/adjustment routine, and (3) does not directly discuss matching of two orthogonal components of ground motions in the proposed adjustments. These scaled ground motions are then applied in three orientations (0, 45, and 90 degrees). For each orientation (and for each of the seven sets of time histories) the peak response at each pertinent Degree of Freedom (DOF) is recorded. Then for each of the seven time-histories the maximum response among the three orientations is obtained, which leads to 7 peak responses at each pertinent DOF. The bridge is then designed for the average of the 7 recorded peak responses at each degree of freedom of interest.

The improvements in the procedures of seismic demand analysis of structures have highly increased the need for acceptable ground motion time series. In this context, a set of earthquake ground motions recorded during past earthquakes are selected, scaled, and utilized to evaluate the seismic demands on structures. It is assumed that this set of ground motions represents the seismic hazard at the location of the structure. Two commonly used methods to obtain hazard consistent ground motions for NLTHA are: 1) selecting a set of ground motion recordings that have a proper range of moment magnitude and distance (M_w and R) obtained from seismic hazard deaggregation at the site, and 2) scaling a number of ground motions to match the Intensity Measure (IM) of a target Hazard Spectrum of the site, which is usually a Uniform Hazard Spectrum (UHS) or a Conditional Mean Spectrum (CMS) (Baker 2011). These two methods are frequently applied in a sequence where recorded ground motions arising from events that match the deaggregated M_w and R, are selected and then scaled to match the target IM. However, these methods have logical limitations; they are based on the assumption that ground motion recordings at other locations can be utilized to represent the seismic hazard at the site of interest. Despite the expansion of the recorded ground motions databases, there remains a lack of usable records especially for scenarios with high magnitude and close distances, which makes any ground motion selection and scaling method to be approximate. Also, scaling processes can highly modify the characteristics of the recorded ground motions and, therefore, may result in unrealistic ground motion characteristics

6

and inaccurate estimation of the structural seismic demands. This has led to an increased interest in developing methods for the generation of synthetic ground motions for design scenarios.

Ground motion simulation methods can be broadly classified into two groups: 1) *Physics-Based* (also known as *Source-Based*) and 2) *Site-Based*. The *Physics-Based* simulation approach is a deterministic technique (randomness is added at an earlier stage for rupture generation, and at a later stage to augment the record with high-frequency content) that uses the three-dimensional (or one-dimensional) representation of the Earth structure to spatially propagate the seismic waves from the rupture source to a site. *Physics-Based* models can produce realistic accelerograms at low frequencies (typically <1 Hz), but often need to be augmented for high frequencies by combining with a stochastic or empirical component, resulting in "hybrid" models (Douglas and Gehl, 2008). In general, these models tend to heavily employ seismological principles to describe the source mechanism and wave travel path, as pointed out by Stafford *et al.* (2009). They depend on physical parameters that vary significantly from region to region. This limits their use in regions where seismological data are lacking–exactly in places where there is an increased need for generation of synthetic ground motions. On the other hand, the *Site-Based* simulation approach involves a stochastic method to generate a data-driven realization of a ground motion time series at a location. Compared to *Source-Based* models, *Site-Based* models are simpler, computationally efficient, and require input parameters that are easily accessible to engineers. Such models implicitly account for source, path, and site effects by developing predictive relations for the model parameters in terms of seismic event parameters (e.g., magnitude, distance, site conditions) that describe the source, the path, and the site. Using the predicted model parameters, *Site-Based* models then utilize a stochastic process to simulate ground motions.

In the light of the shortcomings of current Caltrans design practice in ground motion hazard representation and demand analysis, sophisticated techniques are being proposed to upgrade design and analysis methodologies. In a recent study conducted by Caltrans research team Yoon *et al.,* 2019, the concept of Probabilistic Damage Control Application (PDCA) was used to analyze the bridge structures using the *Site-Based* ground motion simulation model (Rezaeian and Der Kiureghian, 2012; and Dabaghi and Der Kiureghian, 2018; denoted as the DRD simulation model). The study proposed to use the event parameters of the top 3 contributing sources to simulate ground motions and then either conduct point scaling to match UHS $S_a(T_1)$ or use range scaling method to match UHS $S_a(T_1\pm1$ sec). However, this method has some limitations as it adjusts the amplitude

of simulated ground motion time histories by applying linear scaling methods. Due to increase in the computational power and these types of research works, the current engineering community is gradually moving away from generalized design and analysis methodologies. The concept of conducting site-specific analysis within the framework of PBEE is being extensively studied and has been the topic of interest.

Probabilistic Seismic Hazard Analysis (PSHA) has played a major role in promoting probabilistic thinking in structural seismic design and assessment. PSHA forms a critical module in various types of analysis from individual structures to regional seismic risk assessment (Abrahamson, 2006, Abrahamson and Bommer, 2005, Abrahamson *et al.,* 2019, Gerstenberger, 2020). The implementation of PSHA for a given site resides on the execution of the triple integral described in Equation 1.3 where λ represents the rate of exceedance in one year, $G(.)$ represents the Survival Function (*i.e.,* 1 – Cumulative Density Function, F), $P(.)$ represents the probability density function, IM represents the Intensity Measure used (such as Spectral Acceleration (S_a), M represents the magnitude of the event, R represents the distance to the rupture, and N_s, N_M, and N_R represent the number of potential sources, magnitudes and rupture distances. The integration process aggregates the total hazard at a location which leads to the final rate of exceedance of the IM (λ_{IM}). The integration process of the PSHA integral requires two primary ingredients: 1) An Earthquake Rupture Forecast (ERF) model to forecast the potential events along with their respective probabilities and rates, and 2) Ground Motion Model (GMM) that relates the seismic event parameters to an IM and can be used to estimate the probability of exceedance of an IM given the event parameters (M, R etc.).

$$\lambda_{IM} = \sum_{i}^{N_S} \sum_{j}^{N_M} \sum_{k}^{N_R} \lambda_i G(IM|M_j, R_k) P(M_j) P(R_k) \qquad (1.3)$$

The probabilities included in an ERF are based on four layers of modeling: 1) Model of physical geometry of fault, 2) Model for characterizing the deformation of fault in terms of slip rates and related factors, 3) Earthquake rate model of the region, and 4) Probability model to estimate earthquake probability during a time-interval. These are used to produce both time-independent and time-dependent forecasts of earthquake probabilities. Among these one such series of forecasting models includes Uniform California Earthquake Rupture Forecast (UCERF); three models, namely, UCERF1 (1988), UCERF2 (Field *et al.* 2009), and UCERF3 (Field *et al.* 2014)

are released. Apart from the added complexity in the rupture generation and updated probabilities, one of the main differences is the number of possible fault sections. While UCERF1 and UCERF2 contained 16 and 200 faults in California, UCERF3 relies on 350 faults for its computation. Owing to the additional fault segmentation and calculation of the "grand inversion," results require heavy computational resources to cover a broad range of models (>5,000), each considering more than 250,000 fault-based earthquake possibilities (including multi-fault ruptures) throughout California. However, due to all these added complexities of UCERF3, UCERF2 remains as the popular choice and most of the PSHA frameworks continue to utilize it for conducting PSHA.

Ground Motion Models (GMMs) are used to estimate the levels of ground motion Intensity Measures (IMs) using the earthquake source and site parameters (e.g., Magnitude M, Rupture Distance R_{rup}, Site Soil-Shear-Wave Velocity V_{s30}). In general, the most used IMs in the field of Seismic Engineering include Arias Intensity (I_a), Cumulative Absolute Velocity (CAV), Significant Duration (D_{5-95}), Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), and Spectral Acceleration (S_a). Conventionally, GMMs are developed as parametric functional forms using empirical data. Since the ground motions are recorded in two or three orthogonal directions, various measures have been used to quantify the intensity of ground motions on single-degree-of-freedom (SDOF) systems, such as maximum S_a among the two components, average S_a of the two components, Square-Root-of-Sum-of-Squares (SRSS) S_a of the two components, Geomean of S_a of the two components, $GMRotI50\ S_a$, $RotD50\ S_a$ etc. $RotD50\ S_a$ (Boore 2010) is the current state-of-art IM that is adopted by all current GMPMs and is expressed as *RotDpp*, where *Rot* indicates the rotation of the two orthogonal components of the ground motion, *D* indicates the period dependency, and *pp* corresponds to the percentile value (mainly limited to the 50th percentiles, i.e., the median value). For the sake of brevity, $RotD50\ S_a$ is referred as S_a in this study. Some of the commonly used GMMs include ASK14 (Abrahamson *et al.*, 2014), CY14 (Chiou and Youngs, 2014), CB14 (Campbell and Bozorgnia, 2014), AS16 (Afshari and Stewart, 2016), CB19 (Campbell and Bozorgnia, 2019) etc. Apart from the parametric models, various other non-parametric data-driven GMPMs have been proposed in the current studies (e.g., Tezcan and Cheng, 2012; Dhanya and Raghukanth, 2017).

Though the discussed GMMs are primarily developed to estimate linear responses of an SDOF oscillator and are based on past studies (Bazzurro and Luco, 2004; Luco *et al.*, 2005; Luco and

Cornell, 2007) it has been observed S_a does not correlate well with the inelastic response of multi-degree-of-freedom (MDOF) structures, resulting in poor seismic-demand prediction. To tackle this, studies such as Tothong and Cornell (2006) and Bozorgnia *et al.* (2010) have developed GMMs that estimate the inelastic response spectrum conditioned on earthquake scenario parameters. These studies define the inelasticity of the SDOF system in terms of explanatory factors like Strength Reduction Factor (R), Normalized Yield Capacity (V_y/W), Ductility Capacity (μ_{cap}), and Initial Period (T). Though these GMMs provide an advancement to estimate the behavior of nonlinear structures, the use of idealized SDOFs to represent MDOF systems still doesn't incorporate the required level of sophistication in modeling nonlinear behavior. Hence the estimates of Engineering Demand Parameters (EDPs) obtained from the PSHA of such type of GMMs are not adequate to estimate the real structural demands for design and analysis. Also, none of the IMs have completely settled the debate on the efficiency and sufficiency to describe structural response (Jalayer *et al.* 2012).

1.3   Problem Statement

Caltrans' seismic design philosophy for proportioning standard ordinary bridge system and components is moving towards embracing the concepts of Performance-Based Earthquake Engineering (PBEE). Although PBEE concepts have been extensively investigated, such research has not become standardized and transformed into an applied format for Caltrans design practice for standard ordinary bridges. Specifically, the methodologies utilized in the characterization of IM and then obtaining corresponding EDPs are overly generalized and do not explicitly incorporate the tendencies of regional seismicity. This research aims at providing frameworks for the first step of implementing site-specific PBEE in the design of standard ordinary bridges. The proposed effort will encompass aspects of ground motion modeling, structural hazard analysis, ground motion simulation techniques and their utilization.

Current Caltrans Seismic Design Criteria (SDC, ver. 1.7 – ver. 2.0), and state-of-practice, in general, are not well-suited to address site-specific Performance-based Earthquake Engineering (PBEE) concepts in the design of standard ordinary bridges. In essence, current SDC (SDC ver. 1.7) states that standard ordinary bridges are "expected to remain standing but may suffer significant damage requiring closure" at the design level seismic hazard. Such a design philosophy fails to address issues such as: (1) safety, functionality, and durability of bridge structures, (2)

service life optimization and inclusion of lifecycle cost in decision making on selection of "best" structural system and proportioning its components, (3) a holistic view of the transportation network and its performance as a whole, and (4) no explicit consideration of regional seismicity. The current methodologies of design and analysis utilized by SDC (ver.1.7 and ver. 2.0) of selecting and scaling previously recorded ground motions from all over the globe to analyze bridge structures give no consideration to the region-specific seismicity. Apart from numerous issues associated with the scaling of ground motion records (such as alteration of frequency content, non-realistic motions, bias of scaling etc.), the premise of using ground motion records from other sites with different source and site conditions to conduct analysis of structures is fundamentally unnatural. These issues, as addressed in California Bridges and Structures Strategic Direction (2014), are foundational issues to any update to SDC and other design guidelines and tools for proportioning standard ordinary bridges. Such updates require that PBEE concepts become the cornerstone of Caltrans practice in the design of standard ordinary bridges.

PBEE has been the focus of much research in the past two decades. Within PBEE framework, researchers have faced up to challenges such as: identifying performance objectives that are quantifiable and meaningful for engineers and stakeholders alike, identifying representations of ground motion intensity that well relate to seismic hazard at the site and the global bridge response characteristics, and many other advances whose enumeration does not fit into this brief introduction. The new challenge facing researchers is to transfer these sophisticated developments to the practicing engineers through relatively simple but comprehensive procedures that facilitate implementation. The first step for such implementation is the development of guidelines for proper selection of hazard-targeted ground motions and conducting site-specific structural demand analysis. This dissertation tries to propose different statistical frameworks to tackle these issues.

1.4   Dissertation Outline

This dissertation is organized into six chapters, including this introduction. As a first measure to select proper hazard-targeted ground motions, Chapter 2 presents a novel framework of Generalized Ground Motion Prediction Model (GGMPM) developed using hybrid Recurrent Neural Network (RNN) which can be used to estimate an internally correlated 29×1 vector of ground motion intensity measures. Chapter 3 gives an overview of the four bridge structures and five southern California sites used in this study and discusses the developed analytical models of

the bridge structures. Chapter 4 explains the response EDP used in this study and describes the development of EDP hazard curves through conventional Incremental Dynamic Analysis (IDA). Chapter 5 describes the *Site-Based* ground motion simulation tool (DRD model) and its utilization to develop EDP hazard curves of the four bridge structures at the five sites. Similarly, Chapter 6 explains the *Physics-Based* ground motion simulations (CyberShake ver. *15.12* study) and its application to develop EDP hazard curves of the four bridge structures at the five sites.

CHAPTER 2

GENERALIZED GROUND MOTION PREDICTION MODEL

2.1 Introduction

In the field of structural and hazard analysis, Ground Motion Prediction Models (GMPMs) form an essential tool that is used to estimate the levels of ground motion intensity using the earthquake event parameters (e.g., magnitude M, source- to-site R_{rup}). In general, GMPMs provide statistical predictions of Intensity Measure (IMs) such as Arias Intensity (I_a), Cumulative Absolute Velocity (CAV), Significant Duration (D_{5-95}), Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), and Spectral Acceleration (S_a) of a single-degree-of-freedom-system for various periods. GMPMs are used along with an earthquake rupture forecast (ERF) database such as UCERF2, which provides the GMPMs with the necessary event parameters to estimate ground motion intensity at the sites of interest. These are widely useful in an extensive range of research areas of seismic engineering, such as regional seismic analysis, structural loss estimation, hazard map. Over the years, there have been numerous GMPMs developed across the world for both global and regional bases. Conventionally, GMPMs are developed as parametric functional forms using empirical data. Since the ground motions are recorded in two or three orthogonal directions, various measures have been used to quantify the intensity of ground motions on single-degree-of-freedom systems such as maximum S_a among the two components, average S_a of the two components, square-root-of-sum-of-squares S_a of the two components, geomean of S_a of the two components, *GMRotI50 S_a*, *RotD50 S_a*. *RotD50 S_a* is the current state-of-art IM that is adopted by all current GMPMs and is expressed as *RotDpp*, where *Rot* indicates the rotation of the two orthogonal components of the ground motion, *D* indicates the period dependency, and *pp* corresponds to the percentile value (mainly limited to the 50th percentiles, that is, the median value). To avoid any confusion and for the sake of brevity, in this dissertation, the *RotD50* spectral acceleration at bridge's first mode period, which is used as the primary IM of ground motions, is termed as $S_a(T)$ or S_a. The rotations that lead to *RotDpp* depend on period, whereas a single- period-independent rotation is used for *GMRotIpp*; the angle is chosen to minimize the spread of the rotation-dependent geometric mean (normalized by *RotDpp*) over the usable range of oscillator periods.

The GMPMs mainly express the IM as a function of predictor variables such as M, R_{rup}, Shear-Wave

Velocity over the upper 30 m of the site V_{s30}, etc. Abrahamson *et al.* (2014) proposed the ASK14 ground motion model for shallow crustal earth- quakes in active tectonic regions. It includes regional differences between California, Japan, China, and Taiwan. Also, Chiou and Youngs (2014) updated their 2008 version of the model and incorporated regional differences in far-source distance and site effects between active tectonic regions in their CY2014 model. Similarly, Campbell and Bozorgnia (2014) includes regionally independent geometric attenuation, regionally dependent anelastic attenuation, and magnitude-dependent aleatory variability in their CB14 model and corrected the bias of their 2008 version. The growing interest of using intensity- based IMs (such as I_a, CAV) (also termed as integral IMs), and duration-based IMs (such as D_{5-95}) alongside spectrum- based IM (such as S_a) as described by Kiani and Pezeshk (2017), Du and Wang (2018), and Yeow *et al.* (2018) has led to the development of various GMPMs. Afshari and Stewart (2016) developed a GMPM for geometric mean (geomean) of Significant Duration (D_{5-95}), while Campbell and Bozorgnia (2019) recently proposed GMPMs for geomean Arias Intensity (I_a) and CAV. Other investigators have developed relationships correlating the amplitude- and spectrum-based IMs (e.g., Bradley 2012, Du and Wang 2013, Liu *et al.,* 2016, Xu *et al.,* 2016) and ground motion duration measures. As the parametric models with functional forms require pre-informed physical interpretation of seismological and geotechnical modeling, researchers can lean towards data-driven non-parametric modeling techniques for developing GMPMs. These models do not require any predefined formulas and make fair use of advanced statistical techniques to develop predictive models for the ground motion IMs. Tezcan and Cheng (2012) used Support Vector Regression (SVR), which maps data points into a high dimensional feature space to build their non-parametric model to predict response spectra for 13 periods between 0-4s. Recently, Dhanya and Raghukanth (2017) adopted an Artificial Neural Network (ANN) in combination with a Genetic Algorithm (GA) to train a data-driven model to predict PGA, PGV, and spectral accelerations at 26 periods between 0.01 and 4s.

However, one of the drawbacks of these GMPM models is that they predict the target IMs (I_a, *CAV*, S_a at different periods, D_{5-95}) independently from each other. In other words, using the same event parameters (M, R_{rup}, V_{s30}, etc.), the values estimated for the different IMs may or may not be correlated with each other. Particularly for S_a at different periods, Baker and Jayaram (2008) measured the correlations between the epsilons and concluded that the observed correlations are not

sensitive to the choice of GMPMs. They proposed a functional form that estimates the cross-correlations between S_a for periods between 0.01s and 10s. Using this correlational structure, a different form of hazard spectrum was introduced, known as conditional spectrum (CS). Furthermore, Eads *et al.* (2016) developed a metric for quantifying the ground motion intensities called *SaRatio*, which is the ratio between S_a at the first period and the average spectral value over a period range. The primary limitation of such an approach is that the characteristics of seismic excitation are only represented in terms of spectral accelerations. It is well acknowledged that the severity of seismic excitation is not sufficiently described by spectral acceleration(s), which does not explicitly account for other important features that are mainly time-domain signal characteristics. To overcome this limitation, Bradley (2012) proposed a generalized conditional IM (GCIM) approach to construct the multivariate distribution of any set of IMs conditioned on the occurrence of a specific ground-motion IM obtained from probabilistic seismic hazard analysis. Bradley (2012) developed a GM selection algorithm based on the random realization from the conditional multivariate distribution of IMs rather than implicit causal parameters such as source magnitude and source-to-site distance. Although other researchers (Du and Wang 2013, Liu *et al.,* 2016) have also tried to propose empirical correlations between different IMs, the general representation of correlations can be improved. For example, some of the IMs do not necessarily follow the assumption of linear correlation with each other, and the assumption of multivariate lognormality also needs more sophisticated evaluation.

As a first measure to select proper hazard-targeted ground motions, this chapter aims at developing a data-driven non-parametric Generalized Ground Motion Prediction Model (GGMPM) that predicts an internally correlated vector of 29 Intensity Measures (denoted as **IM**) including geomean of Arias Intensity (I_a), geomean of Cumulative Absolute Velocity (*CAV*), geomean of Significant Duration (D_{5-95}) (denoted as $I_{a_{geom}}$, CAV_{geom}, and $D_{5-95_{geom}}$, respectively), and *RotD50* Spectral Acceleration (S_a) at 26 periods (for 5% damped oscillator). To incorporate the higher-order dependencies among the components of **IM**, the framework of Recurrent Neural Network (RNN) is adopted along with optimization techniques of the Evolution Strategy. Long-Short-Term-Memory (LSTM) network is used to estimate **IM** using the seismic source and site parameters which describe the physics of the rupture and the site characteristics, as inputs. A sensitivity analysis is conducted in order to choose a subset of source and site parameters as input to the RNN framework. Furthermore,

the residuals of the RNN structure are carefully calibrated in terms of two 29×29 covariance matrices that estimate the intra-event and inter-event variabilities of the **IM**. This is done by using functional forms to estimate the variances and correlations whose coefficients are calibrated by maximizing the log-likelihood computed from the discrepancy between the **IM** estimated by the RNN framework and the recorded data. The predictions of the developed GGMPM are then compared against the current state-of-art GMPMs.

2.2   Ground Motion Database
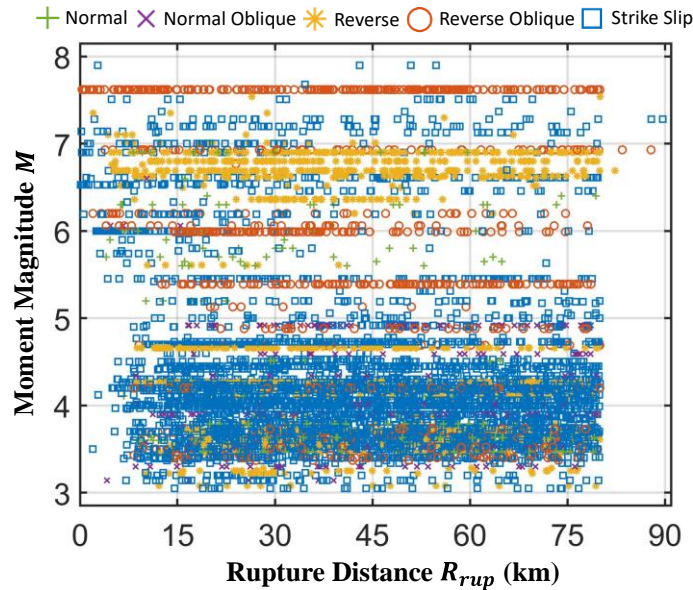
Figure 2.1 – Magnitude (*M*) and Source-to-Site Distance (R_{rup}) of the selected Ground Motions

A subset of the PEER NGA-West2 (Timothy *et al.*, 2014) database containing the bi-directional ground motion acceleration records along with the site information, source information, event parameters, and ground motion intensity measures are selected as the ground motion database for this study. Some of the records from the database are eliminated based on the exclusion criteria given below:

1) Earthquakes that lack any event or site information such as Rake Angle (λ), Strike Angle (ϕ), V_{s30} etc.
2) Earthquakes with a hypocentral depth greater than 20 km or less than 1 km.
3) Earthquakes do not belong to shallow active crustal regions.
4) Recordings that are not representative of free field ground motion.

16

5) Earthquakes with fewer than 5 recordings when $M_w \leq 5$ and with fewer than 3 recordings when $5 \leq M_w$.

6) Recordings that lack any one horizontal component.

7) Recordings having R_{rup} greater than 90 km.

8) Recordings from aftershocks.
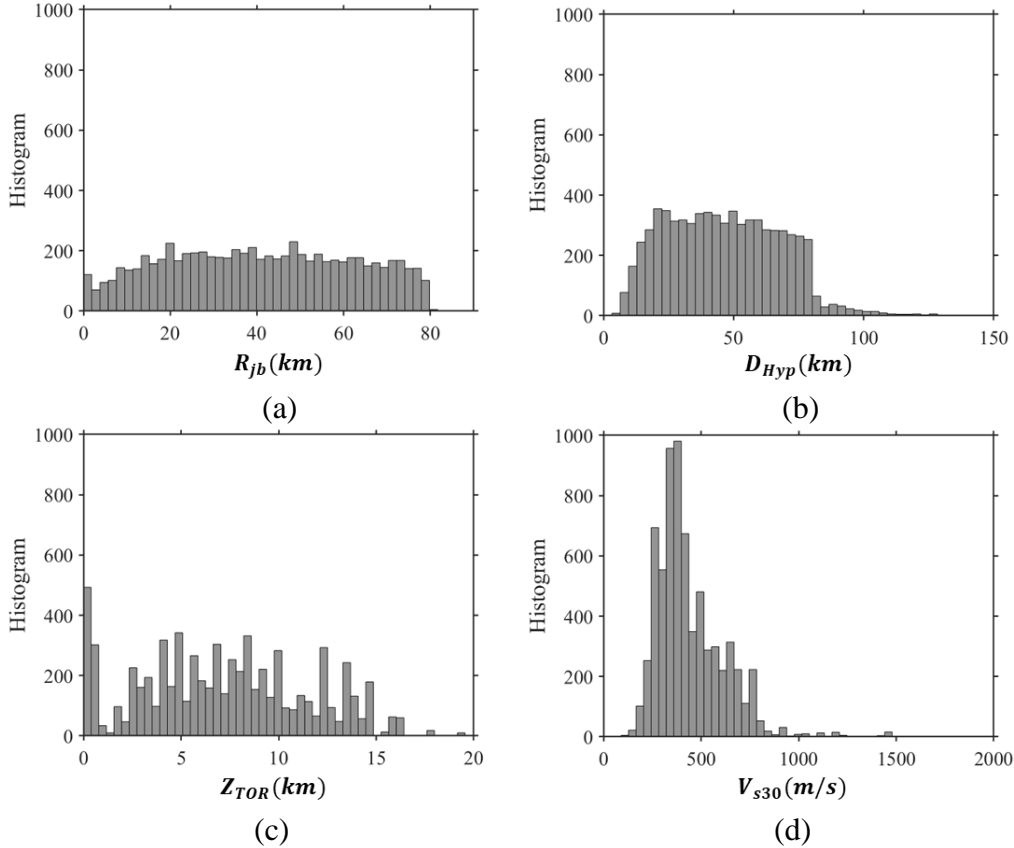
Figure 2.2 – Histograms of: (a) R_{jb} (km), (b) D_{Hyp} (km), (c) Z_{TOR} (km), and (d) V_{s30} (m/s)

Based on the above-mentioned exclusion criteria, 6958 recordings with 277 earthquakes are finally selected. The earthquakes are classified into 5 Fault mechanisms (F) which include Normal (17 earthquakes with 185 recordings), Normal Oblique (12 earthquakes with 245 recordings), Reverse (34 earthquakes with 1,057 recordings), Reverse Oblique (23 earthquakes with 910 recordings), and Strike-Slip (191 earthquakes with 4,575 recordings). The magnitude and distance details of the selected motions are shown in Figure 2.1. The statistical details of the other parameters obtained from the metadata are listed below, a few of which are presented in Figure 2.2:

1) Moment magnitude (M) ranges from 3.05 to 7.9.

17

2) Closest distance to the fault rupture plane (R_{rup}) ranges from 0.07 to 90 km.

3) Joyne-Boore distance to the surface projection of fault rupture plane (R_{jb}) ranges from 0.02 to 90 km.

4) Closest distance to the surface projection of the top of the fault rupture plan (R_x) ranges from -79.93 to 98.57 km.

5) Depth to the top of the fault rupture plane (Z_{TOR}) ranges from 0 to 19.54 km.

6) The average dip angle (δ) of the fault rupture plane ranges from 10 to 90 degrees.

7) Hypocentral depth measured from sea level (Z_{HYP}) ranges from 0.02 to 20.23 km.

8) Distance to Hypocenter (D_{HYP}) ranges from 2.6 to 160.45 km.

9) Rupture length (L) and Rupture width (W) range between 0.4 to 305 km and 0.3 to 70.4 km, respectively.

10) Shear wave velocity in the top 30m of the profile (V_{s30}) ranges from 116.35 to 2016.13 m/s

2.3   Parameter Sensitivity Analysis

Before developing the RNN framework, it is necessary to identify which source and site parameters possess the highest predictive power to estimate the vector of IMs (**IM**). Non-parametric *Random Forests* algorithm (Breiman *et al.,* 1984) is primarily used to conduct the sensitivity analysis of $I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$, and *RotD50 S_a* at three periods, including 0.5, 1.0, and 2.0 secs with respect to all the source and site parameters. The main benefit of using this algorithm as compared to traditional methods of gradient-based sensitivity analysis is that it does not require any closed functional forms to compute the partial derivatives. Due to their non-parametric nature, the *Random Forests* algorithm has a high power of handling large data sets with higher dimensionality, handling the outliers, and deducing the highly non-linear relationships among the features and target variable. The method of Bagging is used to create randomized decision trees in *Random Forests*, and the bootstrap dataset is then created from random sampling (with replacement). Bootstrap datasets that do not contain a particular record from the original dataset are called Out-Of-Bag (OOB) examples (Zhu *et al.,* 2015), and Out-Of-Bag (OOB) estimate for the generalization error is the aggregation of errors of the OOB examples. By first fitting a random forest to the data, the OOB error for each data point is recorded and averaged over

the forest. The importance of the jth feature is measured by permuting the values of the jth feature among the data and computing the OOB error on this perturbed data set. The importance score for
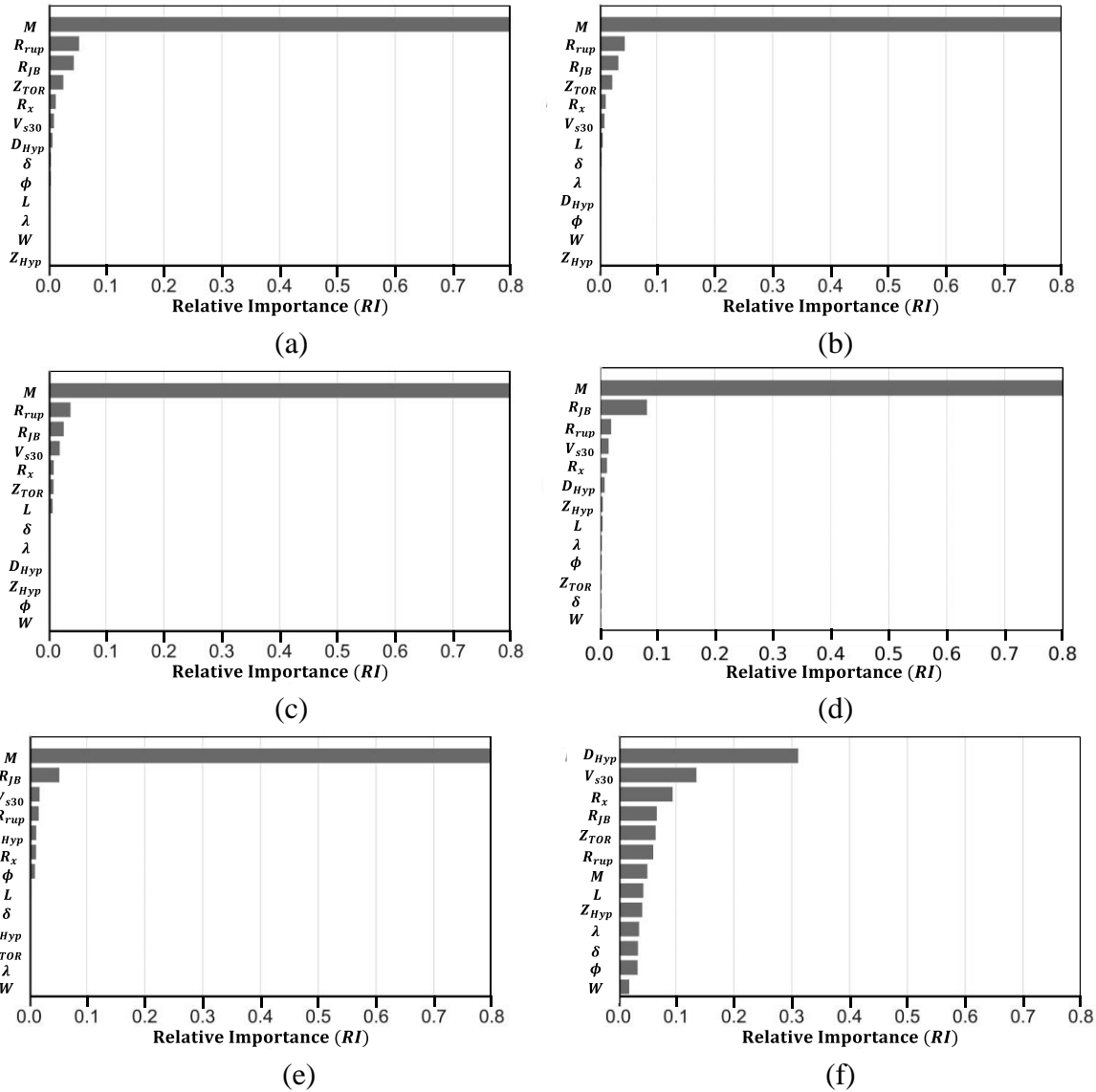
Figure 2.3 – Relative Importance (*RI*) of seismic source and site parameters to predict: (a) $RotD50S_a$ (*T*=0.5s), (b) $RotD50S_a$ (*T*=1.0s), (c) $RotD50S_a$ (*T*=2.0s), (d) $I_{a_{geom}}$, (e) CAV_{geom}, (f) $D_{5-95_{geom}}$

the jth feature is computed by averaging the difference in OOB error before and after the permutation over all trees. Features that produce large values for the normalized (by the standard deviation) score are ranked as more important than features that produce small values. Using the

Random Forests algorithm, the relative importance (*RI*) of each seismic source and site parameter in predicting the target IMs ($I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$, and *RotD50* S_a at T = 0.5s, 1.0s, and 2.0s) is obtained and compared. The results are presented in Figure 2.3. As can be observed from Figure 2.3, parameters including M, R_{rup}, R_{JB}, R_x, Z_{TOR}, and V_{s30} are deemed significant for all IMs. Specifically, it is observed that the D_{Hyp} becomes primarily important for estimating D_{5-95}. Source parameters of M, R_{rup}, and R_x are consistently observed to be rated significantly by the *Random Forests* algorithm. Also, the physics of the fault mechanism (*F*) is widely known to cause differences in the ground motion characteristics, hence finally 8 parameters (including F, M, R_{rup}, R_x, R_{JB}, Z_{TOR}, D_{Hyp}, and V_{s30}) are selected to be used as inputs to the RNN framework. All the selected parameters are consistent with the current-state-of-practice parametric and non-parametric GMPMs (e.g., CB14; CY14; Dhanya and Raghukanth, 2017).

2.4   Recurrent Neural Networks
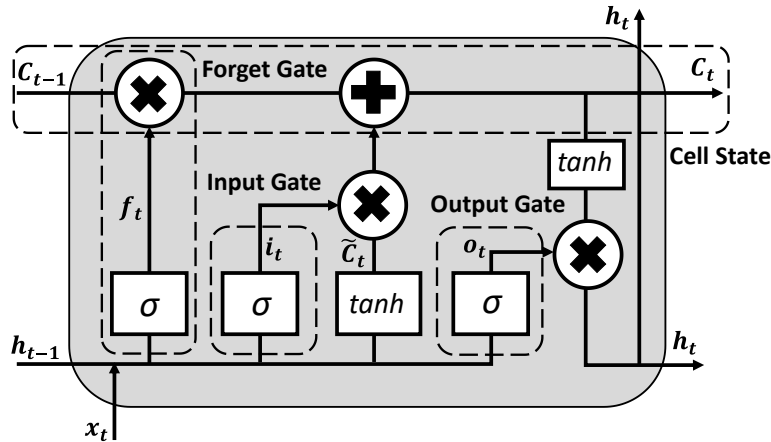
Figure 2.4 - Long Short-Term Memory (LSTM) cell structure

Recurrent Neural Network (RNN) is a class of artificial neural network (ANN) that attempts to model time or sequence-dependent behavior. This is performed by feeding back the output of a neural network layer at time *t* to the input of the same network layer at time *t* + 1. Hence, RNN possesses connections between nodes that form a directed graph along a temporal sequence, which allows it to exhibit temporal dynamic behavior. RNN not only feeds forward but keeps an internal memory to process the sequences of inputs so that all input vectors are related to each other. Therefore, RNN is one of the best candidates to train a data-driven model for the prediction of

sequential processes, in this case of the correlated vector of IMs. Since these IMs represent the characteristics of the same ground motion, the values are dependent on each other. This can be viewed as a sequence of IMs where one value depends on the other values in the sequence. The recurrent nature of RNN allows it to perform the same function for each input, copying and sending the data back to the network while producing the output simultaneously.

Although RNN is capable of tackling dependencies between the steps of the sequences, RNNs are known to have problems of short-term memory and vanishing gradients. To tackle this issue, the framework of Long-Short-Term-Memory (LSTM) (Hochreiter and Schmidhuber, 1996) is used, which consists of internal mechanisms called 'gates' that regulate information flow. A typical LSTM structure is a cell state consisting of three gates that are explained below and illustrated in Figure 2.4.

1) *Forget Gate*: The Forget Gate determines what information should be discarded or retained from the previous steps in the sequence. Information from the previously hidden state and information from the current input is passed through the sigmoid function (σ). The function of the gate is shown in Equation 2.1, where the current input x_t and the previous output h_{t-1} are combined using weights W_f and bias b_f with a sigmoid layer.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \qquad (2.1)$$

2) *Input Gate*: This determines how the cell state (C_t) will be updated, and memory will be modified based on the input. Firstly, the previous hidden state (h_{t-1}) and the current input (x_t) are combined through the sigmoid function (σ) as shown in Equation 2.2 using weights W_i and bias b_i to obtain the sigmoid output (i_t). Then using Equation 2.3, the hidden state (h_{t-1}) and current input (x_t) are passed to the *tanh* function with weights W_C and bias b_C to regulate the network. In Equation 2.4, the *tanh* output ($\tilde{C}_t$) is multiplied by the sigmoid output (i_t) and the previous cell state (C_{t-1}) is multiplied by the forget vector (f_t). If f_t is close to 0, C_{t-1} may be discarded in new cell state (C_t). Then the addition of the two products updates the cell state (C_t).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \qquad (2.2)$$

21

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{2.3}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{2.4}$$

3) *Output Gate*: This decides what the next hidden state (h_t) should be, and which information will be produced by the state. First, the previous hidden state (h_{t-1}) and current input (x_t) are passed to the sigmoid function with weights (W_o) and bias (b_o) to obtain output (o_t) using Equation 2.5. Then the newly modified cell state (C_t) is passed to the *tanh* function and the output is multiplied with the sigmoid output (o_t) using Equation 2.6 to determine the information to be carried by the hidden state (h_t). The new hidden state (h_t) is transferred to the next time step.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.5}$$

$$h_t = o_t \times tanh(C_t) \tag{2.6}$$
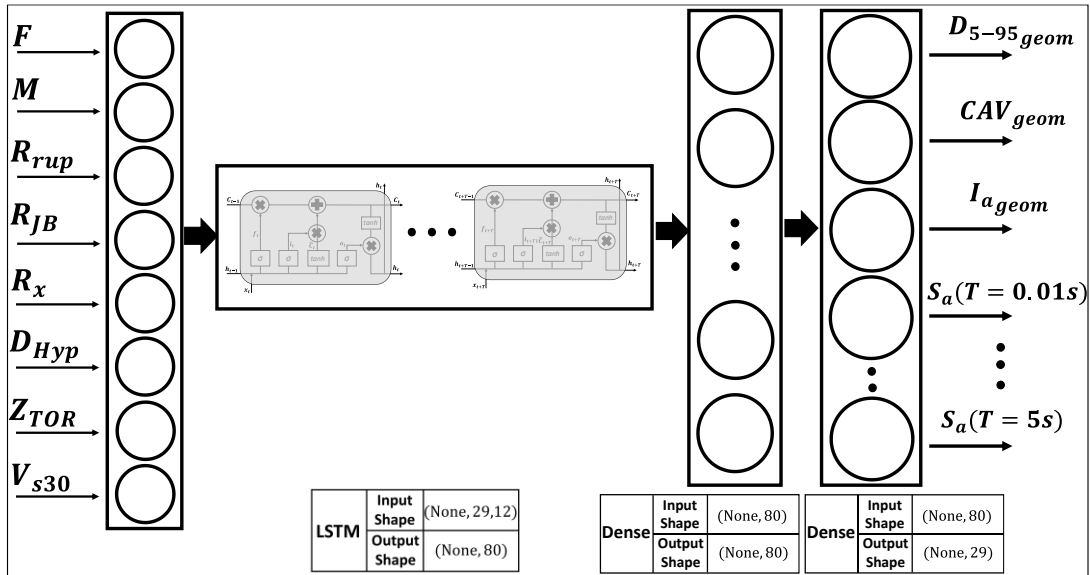
Figure 2.5 – Proposed Recurrent Neural Network (RNN) Architecture

The selected database of the PEER NGA-West2 consisting of 6,947 recordings from 277 events are randomly split into Train and Test sets with 70% of events in the Train set and 30% of events in the Test set. Since the database is heavily dominated by the smaller magnitude events, the Train and Test splitting is conducted in discretized sets of $3 \leq M_w \leq 4$, $4 < M_w < 5$, and $5 \leq M_w$. This means that 70:30 (Train: Test) random split is done independently for the events in each discretized set and then combined to form a single Train set and Test set. 10% of the train set is then further used to train the neural network with cross-validation. To further prevent overfitting, the neural

22

network framework is trained in epochs using Adam optimizer (Kingma and Ba, 2014) along with the Early Stopping (Prechelt, 2002) callback. The final proposed Neural Network architecture is presented in Figure 2.5. The source and site inputs to the RNN framework include a vector of 12 values of $\{M, R_{rup}, R_x, R_{JB}, D_{Hyp}, Z_{TOR}, V_{s30}, F\}$. F denotes the Fault mechanism represented by a one-hot encoded vector for 5 fault mechanisms. Unlike the conventional methods of using discretized values for the discretized classes (as done by Dhanya and Raghukanth, 2017), the 5 classes of fault mechanisms are represented by one-hot vector given in Table 2.1. Using discretized values for the classes is usually not the preferred method to differentiate between the classes since the classes that receive arbitrary greater values prioritize the gradient slope in the backpropagation algorithm. Hence to avoid this, the fault mechanisms are differentiated in terms of one-hot vector as given in Table 2.1. The framework then processes the correlated vector of IMs (**IM**) using the LSTM layer, which estimates the values of IMs while maintaining their internal dependencies. The LSTM layer is then connected to two Dense ANN layers that update the estimations made by the LSTM layer to obtain better final predictions. The output of the RNN framework is the median **IM**, which includes $I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$, and $RotD50S_a$ values corresponding to periods of 0.01 sec to 1 sec with the interval of 0.1 sec, and 1.2 sec to 3.0 sec with the interval of 0.2 sec, and 3.4 sec to 5 sec with the interval of 0.4 sec.

Table 2.1 - One-Hot Vectors for Fault Mechanisms

Mechanism (F)	One-Hot Vector			
Strike Slip	1	0	0	0
Normal	0	1	0	0
Reverse	0	0	1	0
Reverse Oblique	0	0	0	1
Normal Oblique	0	0	0	0

2.5 Estimation of Covariance Matrices for Residuals

2.5.1 Methodology

The proposed RNN structure is used to estimate 29×1 vectorized **IM**. The residuals of the RNN model are carefully tackled in two aspects. First, the error term, which measures the discrepancy

between the **IM** from the database and the predicted **IM** value through the RNN structure, is carefully partitioned into two parts: inter-event and intra-event effects. The hierarchical model is expressed in a functional form shown in Equation 2.7, where y_{ij} represent the vector of target IMs (**IM**) and $RNN(M_i, r_{ij}, \dots)$ is the estimated **IM** from implementing the RNN structure. The random effects of the ith event, namely $\pmb{\eta}_i$ represent inter-event variation, and ε_{ij} represents the intra-event variation of the jth recording from the ith event. For each target vector, the components of $\pmb{\eta}_i$ and ε_{ij} are assumed to be normally distributed with zero means and variances τ^2 and σ^2, respectively. Secondly, since the 29 components of y_{ij} represent the same ground motion recording, they are not treated as independent variables. The correlations in the residuals are fetched by estimating the variances τ^2 and σ^2 as 29×29 covariance matrices **T** and **Σ**, as shown in Equations 2.8 and 2.9.

In a nutshell, $\pmb{IM_{ij}}$ and $\pmb{RNN}(M_i, r_{ij}, \dots)$ are the true and predicted vectors of **IM** and **T** and **Σ** are the estimated 29×29 covariance matrices for inter-event (η_i) and intra-event (ε_{ij}) variation for j^{th} ground motion record of ith seismic event. The diagonal elements of **T** and **Σ** represent the inter-event (τ^2) and the intra-event (σ^2) variances of the 29 components of **IM**, including variances of $D_{5-95_{geom}}$ ($\sigma_D{}^2$, $\tau_D{}^2$), variances of CAV_{geom} ($\sigma_C{}^2$, $\tau_C{}^2$), variances of $I_{a_{geom}}$ ($\sigma_I{}^2$, $\tau_I{}^2$) and variances of $RotD50S_a$ at 26 periods ($\sigma_{Sa_{0.01}}{}^2$ to $\sigma_{Sa_{5.0}}{}^2$, $\tau_{Sa_{0.01}}{}^2$ to $\tau_{Sa_{5.0}}{}^2$). The off-diagonal element at the mth row and the nth column represent the covariances $\tau_m\tau_n\rho_{m,n}$ and $\sigma_m\sigma_n\rho_{m,n}$ for **T** and **Σ**, respectively, where $\rho_{m,n}$ is the correlation between the residual ($\pmb{IM_{ij}} - \pmb{RNN}(M_i, r_{ij}, \dots)$) of the mth IM and residual ($\pmb{IM_{ij}} - \pmb{RNN}(M_i, r_{ij}, \dots)$) of the nth IM, respectively. The ρ consists of three types of correlations: 1) correlation of the residuals among $D_{5-95_{geom}}$, CAV_{geom} and $I_{a_{geom}}$ (*i.e.* $\rho_{D,C}$, $\rho_{D,I}$, $\rho_{C,I}$), 2) correlation of the residuals between $D_{5-95_{geom}}$ and $RotD50S_a$ at 26 periods (*i.e.* $\rho_{D,Sa_{0.01}}$ to $\rho_{D,Sa_{5.0}}$), CAV_{geom} and $RotD50S_a$ at 26 periods (*i.e.* $\rho_{C,Sa_{0.01}}$ to $\rho_{C,Sa_{5.0}}$), and $I_{a_{geom}}$ and $RotD50S_a$ at 26 periods (*i.e.* $\rho_{I,Sa_{0.01}}$ to $\rho_{I,Sa_{5.0}}$), and 3) correlation of the residuals among the $RotD50S_a$ pairs $\rho_{i,j}$, where i and j represent the ith and j^{th} periods among the 26 periods (*i.e.* $\rho_{Sa_{0.01},Sa_{0.1}}$ to $\rho_{Sa_{4.6},Sa_{5.0}}$).

$$ln\, \boldsymbol{IM_{ij}} = \boldsymbol{RNN}(M_i, r_{ij}, \dots) + \boldsymbol{\eta_i} + \boldsymbol{\varepsilon_{ij}} \tag{2.7}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_D{}^2 & \rho_{D,C}\sigma_D\sigma_C & \cdots & \rho_{D,Sa_{5.0}}\sigma_D\sigma_{Sa_{5.0}} \\ \rho_{D,C}\sigma_D\sigma_C & \sigma_C{}^2 & & \rho_{C,Sa_{5.0}}\sigma_C\sigma_{Sa_{5.0}} \\ \vdots & & \ddots & \vdots \\ \rho_{D,Sa_{5.0}}\sigma_D\sigma_{Sa_{5.0}} & \rho_{C,Sa_{5.0}}\sigma_C\sigma_{Sa_{5.0}} & \cdots & \sigma_{Sa_{5.0}}{}^2 \end{bmatrix} \tag{2.8}$$

$$\boldsymbol{T} = \begin{bmatrix} \tau_D{}^2 & \rho_{D,C}\tau_D\tau_C & \cdots & \rho_{D,Sa_{5.0}}\tau_D\tau_{Sa_{5.0}} \\ \rho_{D,C}\tau_D\tau_C & \tau_C{}^2 & & \rho_{C,Sa_{5.0}}\tau_C\tau_{Sa_{5.0}} \\ \vdots & & \ddots & \vdots \\ \rho_{D,Sa_{5.0}}\tau_D\tau_{Sa_{5.0}} & \rho_{C,Sa_{5.0}}\tau_C\tau_{Sa_{5.0}} & \cdots & \tau_{Sa_{5.0}}{}^2 \end{bmatrix} \tag{2.9}$$
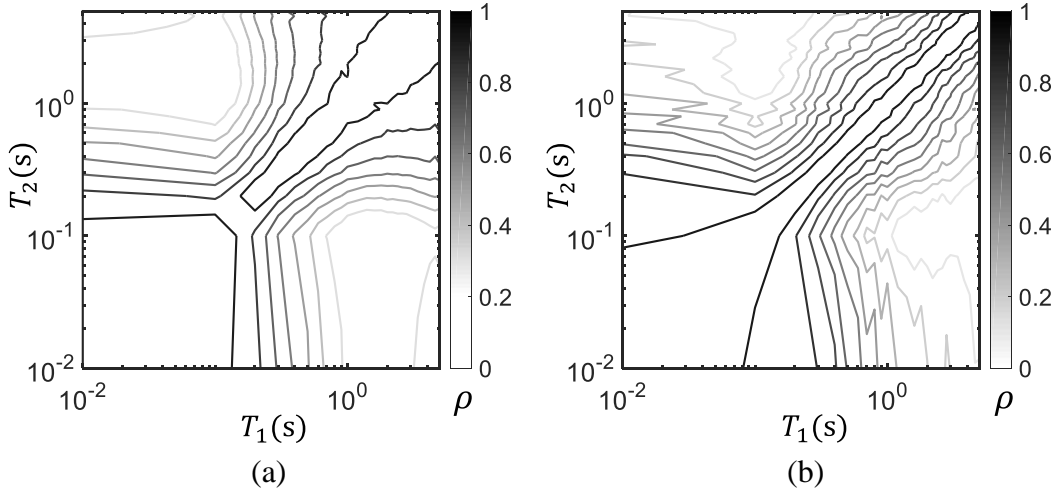
Figure 2.6 - Correlation of the residuals between *RotD50S$_a$* at 26 periods for: (a) $M \leq 4.5$, and (b) $M \geq 5.5$

Following the suggestions of Campbell and Bozorgnia (2014), the variances τ^2 and σ^2 for *RotD50S$_a$* at each period show significant differences when grouped by the earthquake magnitude M. Their work proposed two sets of τ^2 and σ^2, one for $M \leq 4.5$, one for $M \geq 5.5$, and linear interpolation is proposed for M between the 4.5 and 5.5. However, the issue remains unsolved whether the correlation between the residuals of the target IMs differs with earthquake magnitude M. To test the dependency of correlation coefficients on M, the dataset in this study is separated into groups of $M \leq 4.5$ and $M \geq 5.5$. For both groups, the correlation coefficient matrix is calculated for the residuals of *RotD50S$_a$* at 26 periods and are plotted in Figure 2.6a and Figure 2.6b, respectively.

25

By visual inspection, it can be observed that the correlation coefficient structures of the two groups are different. To statistically assess the similarity of the two correlation structures, chi-squared (χ^2) test (Srivastava and Yanagihara, 2010) for equality of covariance matrices is carried out. The test resulted in a *p-value* ≤ 0.001, which implies that the two correlation coefficient structures cannot be considered as the same. Therefore, in this study, the inter-event and intra-event covariance matrices **T** and **Σ** are estimated separately for $M \leq 4.5$ and $M \geq 5.5$, and linear interpolation is suggested for $4.5 < M < 5.5$.

To estimate parameters of the covariance matrices **T** and **Σ** (for various ranges of *M*), the log-likelihood function shown in Equation 2.10 is used. In Equation 2.10, y and μ are vectors of true and RNN-predicted **IM**s, respectively, and C is the covariance matrix, and $|C|$ and C^{-1} are its determinant and inverse, respectively. As an $p \times p$ covariance matrix has $p(p + 1)/2$ distinct elements, the total number of parameters that are required to be optimized for the four matrices is equal to $4 \times 29 \times 30/2 = 1740$. Since there are a large number of parameters to be optimized, which can be computationally expensive, the elements contained in the covariance matrices are not directly treated as unknowns. Instead, functional forms are fitted to preliminary estimation of the variances (both inter-event and intra-event) and correlations, and then the coefficients of the fitted functional forms are optimized to minimize the negative of the likelihood function. The optimized functional forms are then used to compute the variances and correlations of the residuals.

$$ln\,L = \frac{N}{2}ln(2\pi) - \frac{1}{2}ln|C| - \frac{1}{2}(y - \mu)^T C^{-1}(y - \mu) \tag{2.10}$$

In this study, the variances (inter-event (τ^2) and intra-event (σ^2)) and correlations (ρ) are estimated separately due to their nature. An aleatory variability model can be easily developed using the combination of τ^2, σ^2 and ρ, where τ^2 and σ^2 together form the diagonal elements of covariance matrices **T** and **Σ**, while the correlations (ρ) are parts of the off-diagonal elements. The following stepped procedure is followed to estimate the covariance matrices (steps are not necessarily to be followed sequentially). The computation of variances τ^2 and σ^2 is tackled in Steps 1 and 3, and the correlations ρ are dealt in Steps 2, 4, and 5. Finally, the three types of estimates are combined in Step 6. The methodology is also illustrated in Figure 2.7.
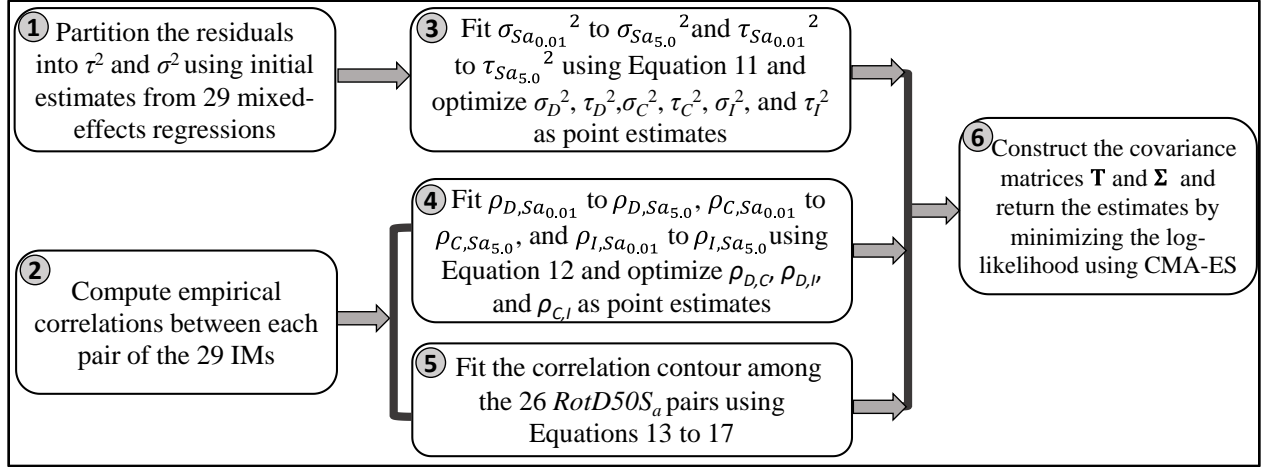
Figure 2.7 - Illustration for estimating correlation matrix

Step 1:

Since the total variability in the residuals ($\boldsymbol{IM_{ij}} - \boldsymbol{RNN}(M_i, r_{ij}, \dots)$) can be expressed as $\tau^2 + \sigma^2$, the residuals are partitioned into two parts: τ^2 and σ^2, using the ratios of the initial estimates of τ^2 and σ^2 (termed as τ^{ini^2} and σ^{ini^2}, respectively). The initial estimates τ^{ini^2} and σ^{ini^2} are obtained by conducting 29 independent linear mixed effect regressions using $D_{5-95_{geom}}$, CAV_{geom}, Ia_{geom}, and *RotD50S$_a$* at 26 periods as the targets and M_w, R_{rup}, R_{jb}, R_x, Z_{TOR}, V_{s30}, and F as the predicting variables. The 29 mixed-effects regressions return the initial estimates of inter-event variances ($\sigma_D^{ini^2}$, $\sigma_C^{ini^2}$, $\sigma_I^{ini^2}$, $\sigma_{Sa_{0.01}}^{ini^2}$ to $\sigma_{Sa_{5.0}}^{ini^2}$) and the intra-event variances ($\tau_D^{ini^2}$, $\tau_C^{ini^2}$, $\tau_I^{ini^2}$, $\tau_{Sa_{0.01}}^{ini^2}$ to $\tau_{Sa_{5.0}}^{ini^2}$) for the 29 targets. The superscript *ini* is used to denote these as initial estimates of the variances. Then the pairwise ratios of $\tau^{ini^2}/\sigma^{ini^2}$ for the 29 targets are used to partition the RNN residuals ($\boldsymbol{IM_{ij}} - \boldsymbol{RNN}(M_i, r_{ij}, \dots)$) into two parts: τ^2 and σ^2. While arguably the error terms of different statistical models are dependent on their mathematical formulations, for this study, it is postulated that the ratio of τ^2/σ^2 only depends on the seismic events and follows a pattern for the various target IMs. This partitioning of the residuals, however, doesn't affect the pairwise correlation of residuals of the target IMs. For this reason, correlations ρ computed from residuals will not differ for the inter-event and intra-event effects.

Step 2:

The Pearson product-moment correlations (Garren, 1988) of the residuals between each pair of the 29 targets are computed. This results in a 29×29 correlation coefficient matrix, whose off-diagonal elements include: $\rho_{D,Sa_{0.01}}$ to $\rho_{D,Sa_{5.0}}$, $\rho_{C,Sa_{0.01}}$ to $\rho_{C,Sa_{5.0}}$, $\rho_{Ia,Sa_{0.01}}$ to $\rho_{I,Sa_{5.0}}$, $\rho_{Sa_{0.01},Sa_{0.1}}$ to $\rho_{Sa_{4.6},Sa_{5.0}}$ and $\rho_{D,C}$, $\rho_{D,I}$, $\rho_{C,I}$, and the diagonal elements are equal to 1. This matrix represents the "true" computed pairwise correlations between the 29 targets.

Step 3:

The variances of *RotD50Sa* at 26 periods $\sigma_{Sa_{0.01}}{}^2$ to $\sigma_{Sa_{5.0}}{}^2$ and $\tau_{Sa_{0.01}}{}^2$ to $\tau_{Sa_{5.0}}{}^2$ calculated in *Step 1* are fitted using a 3rd order polynomial function, as shown in Equation 2.11. To quantify the feasibility of the polynomial models, 2nd, 3rd, and 4th order polynomial functions were evaluated using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Although AIC and BIC of 4th order polynomial functions were observed to be slightly (~ 0.5 points) less than those of 3rd order polynomial functions, the improvement was observed to be negligible and 3rd order polynomial function form was selected for the sake of dimensionality reduction of fitting parameters in the optimization process. While variances $\sigma_{Sa_{0.01}}{}^2$ to $\sigma_{Sa_{5.0}}{}^2$ and $\tau_{Sa_{0.01}}{}^2$ to $\tau_{Sa_{5.0}}{}^2$ are fitted as functions of the periods (Equation 2.11), other variances for $D_{5-95_{geom}}$, CAV_{geom}, $I_{a_{geom}}$ ($\sigma_D{}^2$, $\tau_D{}^2$, $\sigma_C{}^2$, $\tau_C{}^2$, $\sigma_I{}^2$, $\tau_I{}^2$) are considered as point estimates and optimized as single variables.

$$v(T) = p_{3,v}T^3 + p_{2,v}T^2 + p_{1,v}T^1 + p_{0,v} \qquad v \in \{\sigma, \tau\}, \text{ and } 0.01\text{s} \le T \le 5.0\text{s} \qquad (2.11)$$

Step 4:

Similarly, the correlations: $\rho_{D,Sa_{0.01}}$ to $\rho_{D,Sa_{5.0}}$, $\rho_{C,Sa_{0.01}}$ to $\rho_{C,Sa_{5.0}}$, $\rho_{I,Sa_{0.01}}$ to $\rho_{I,Sa_{5.0}}$ are fitted using 3rd order polynomial functions shown in Equation 2.12. It should be noted that the dimensional reduction not only benefits in the optimization scheme, but the functional form of the fitting can also be used for the generalized prediction of variances and correlations. Similar to *Step 3*, the correlations between $D_{5-95_{geom}}$, CAV_{geom}, $I_{a_{geom}}$ ($\rho_{D,C}$, $\rho_{D,I}$, $\rho_{C,I}$) are considered as point estimates and optimized as single variables.

$$\rho_{u,S_a}(T) = q_{3,u}T^3 + q_{2,u}T^2 + q_{1,u}T^1 + q_{0,u} \qquad u \in \{D_{5-95_{geom}}, CAV_{geom}, \text{and} \atop I_{a_{geom}}\}, \text{and } 0.01s \le T \le 5.0s \qquad (2.12)$$

Step 5:

To account for the correlation among the residuals of all *RotD50S$_a$* pairs, a functional form of correlation contour is utilized to fit the correlation coefficient matrix computed in *Step 2*. This represents the 26×26 submatrix from the 29×29 correlation coefficient matrix computed in *Step 2*. Due to its versatility, the functional form developed by Baker and Jayaram (2008) is adopted in this study with 5 selected parameters (s_1, s_2, s_3, t_1, t_2) shown in Equations 2.13 to 2.17. These parameters are optimized to obtain the contour that fits best to the computed correlation coefficient matrix.

$$C_1 = 1 - \cos\left(\frac{\pi}{2} - s_1 \ln\left(\frac{T_{max}}{\max(T1_{min}, t_1)}\right)\right) \qquad (2.13)$$

$$C_2 = \begin{cases} 1 - s_2\left(1 - \dfrac{1}{1 + e^{100T_{max}-5}}\right)\left(\dfrac{T_{max} - T_{min}}{T_{max} - 0.0099}\right), T_{max} < t_2 \\ 0, otherwise \end{cases} \qquad (2.14)$$

$$C_3 = \begin{cases} C_2, T_{max} < t_1 \\ C_1, otherwise \end{cases} \qquad (2.15)$$

$$C_4 = C_1 + s_3\left(\sqrt{C_3} - C_3\right)\left(1 + \cos\left(\frac{\pi T_{min}}{t_1}\right)\right) \qquad (2.16)$$

$$\rho = \begin{cases} C_2, T_{max} < t_1 \\ C_1, T_{min} > t_1 \\ \min(C_2, C_4), T_{max} < t_2 \\ C_4, otherwise \end{cases} \qquad \begin{matrix} s_1, s_2, s_3 \text{ are multiplication factors. } t_1, \\ t_2 \text{ are the threshold of periods that} \\ \text{control the shape of the correlation} \\ \text{matrix.} \end{matrix} \qquad (2.17)$$

Step 6:

Given the specific set of parameters, 29x29 covariance matrices **T** and **Σ** (Equation 2.8 and Equation 2.9) are formed with three submatrices:

   i. Point estimates σ_D^2, τ_D^2, σ_C^2, τ_C^2, σ_I^2, $\rho_{D,C}$, $\rho_{D,I}$, $\rho_{C,I}$ are used to form 3×3 covariance matrices among the residuals of $D_{5-95_{geom}}$, CAV_{geom}, $I_{a_{geom}}$.

ii. Equation 2.12 is used to construct correlations $\rho_{D,Sa_{0.01}}$ to $\rho_{D,Sa_{5.0}}$, $\rho_{C,Sa_{0.01}}$ to $\rho_{C,Sa_{5.0}}$, and $\rho_{I,Sa_{0.01}}$ to $\rho_{I,Sa_{5.0}}$, while Equation 2.11 is used for variances $\sigma_{Sa_{0.01}}{}^2$ to $\sigma_{Sa_{5.0}}{}^2$ and $\tau_{Sa_{0.01}}{}^2$ to $\tau_{Sa_{5.0}}{}^2$. This results in 26x3 (or 3x26) matrices representing covariance of $D_{5-95_{geom}}$, CAV_{geom}, and $I_{a_{geom}}$ with $RotDS_a$ at 26 periods.

iii. The coefficients (s_1, s_2, s_3, t_1, t_2) from Equations 2.13 to 2.17 are used to form a 26×26 correlation coefficient matrix. Combining the correlation coefficient matrix with constructed variances $\sigma_{Sa_{0.01}}{}^2$ to $\sigma_{Sa_{5.0}}{}^2$ and $\tau_{Sa_{0.01}}{}^2$ to $\tau_{Sa_{5.0}}{}^2$, 26×26 covariance matrices of S_a at 26 periods are formed. The three submatrices are then merged to develop 29×29 covariance matrices **T** and **Σ**, which are finally utilized to form the covariance matrix C in Equation 2.10 using Equations 2.18 and 2.19.

$$\mathbf{E_i} = \begin{bmatrix} \mathbf{\Sigma + T} & \cdots & \mathbf{T} \\ \vdots & \ddots & \vdots \\ \mathbf{T} & \cdots & \mathbf{\Sigma + T} \end{bmatrix}$$

where $\mathbf{E_i}$ is a $29n_i \times 29n_i$ sparse matrix accounts for the covariance of the ith event with diagonal matrices of $\mathbf{\Sigma + T}$ and off-diagonal matrices of $\mathbf{T}$. n_i is the number of recordings in the ith event. (2.18)

$$\mathbf{C} = \begin{bmatrix} \mathbf{E_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{E_n} \end{bmatrix}$$

where **C** is a $29N \times 29N$ sparse matrix with diagonal matrices of $\mathbf{E_i}$ and off-diagonal matrices of **0**. N is the total number of recordings from all events. (2.19)

The negative of log-likelihood function is then minimized by optimizing the parameters discussed in *Steps 1-5* using Covariance Matrix Adaption Evolution Strategy (CMA-ES) (Hansen, 2006). CMA-ES is a widely used genetic algorithm to solve non-convex optimization problems. In this algorithm, several particles are randomly generated, with each particle representing a particular setting of parameters that are random variables of the objective function to be minimized. During the search steps of the algorithm, the mean of the distribution of the particles as well as the covariance matrix of the distribution of the particles are both updated until the algorithm converges at the global minimum. The boundary conditions of the algorithm are fixed to keep the correlation values between -1 and 1, standard deviations ≥ 0, and the covariance matrices as positive-definite (eigenvalues ≥ 0).

2.5.2 Results

The final solution of the optimal parameters obtained from this procedure is presented in Tables 2.2 to 2.4, the optimal standard deviations of the 29 targets are presented in Figure 2.6 and Figure 2.9, and the optimally fitted correlations are compared against the ones directly computed from the residuals of the RNN structure in Figures 2.6, 2.7, 2.8, and 2.10. In Figures 2.6, 2.7, 2.8, and 2.10, '*computed*' represents the empirical values directly calculated from the residuals of the RNN, while '*fitted*' indicates the values obtained from the optimized function forms.

Table 2.2 - Coefficients of Polynomial Function (Equation 2.11)

v	$p_{3,v}$	$p_{2,v}$	$p_{1,v}$	$p_{0,v}$
$\sigma_{Sa}, M \leq 4.5$	-0.0075	0.0828	-0.2422	0.7544
$\sigma_{Sa}, M \geq 5.5$	0.0020	-0.0178	0.0568	0.5728
$\tau_{Sa}, M \leq 4.5$	-0.0009	-0.0204	0.1205	0.3530
$\tau_{Sa}, M \geq 5.5$	-0.0039	0.0148	0.0180	0.2422

Table 2.3 - Coefficients of Polynomial Function (Equation 2.12)

u	$q_{3,u}$	$q_{2,u}$	$q_{1,u}$	$q_{0,u}$
$D, M \leq 4.5$	0.0295	-0.2298	0.5709	-0.5045
$D, M \geq 5.5$	0.0102	-0.1488	0.4898	-0.5556
$C, M \leq 4.5$	-0.0253	0.2244	-0.6016	0.8326
$C, M \geq 5.5$	-0.0134	0.1129	-0.3099	0.5250
$I, M \leq 4.5$	-0.0308	0.2674	-0.7287	0.9466
$I, M \geq 5.5$	-0.0221	0.2097	-0.5183	0.8787

Figure 2.8 shows the intra-event and the inter-event standard deviations of the residuals at 26 periods $\sigma_{Sa_{0.01}}$ to $\sigma_{Sa_{5.0}}$ and $\tau_{Sa_{0.01}}$ to $\tau_{Sa_{5.0}}$ with Figure 2.8a (left column) showing the results of $M \leq 4.5$ and Figure 2.8b (right column) showing the results of $M \geq 5.5$. From Figure 2.8a, it can be observed that the relationship between σ_{Sa} and periods tend to show a decreasing trend at periods less than 2 sec and an increasing trend at periods larger than 2 sec. Overall, σ_{Sa} ranges from 0.55 to 0.75. In Figure 2.8b, σ_{Sa} remains consistent from 0.55 to 0.65. For τ_{Sa} in Figure 2.8a, the standard deviation has an increasing trend at periods smaller than 2 sec and a decreasing trend at periods larger than 2 sec, and τ_{Sa} ranges from 0.3 to 0.5 in general. For τ_{Sa} in Figure 2.8b, σ_{Sa} remains consistent from 0.2 to 0.35. To summarize, for $M \leq 4.5$, the trends of intra-event and

the inter-event standard deviations change after a period of 2 sec, while for $M \geq 5.5$, the minimum variations are observed throughout the periods.

Table 2.4 - Parameters of the correlation coefficient matrix for Equations 2.13-2.17

	s_1	s_2	s_3	t_1	t_2
$M \leq 4.5$	0.3869	0.1512	0.7174	0.0885	0.1989
$M \geq 5.5$	0.4241	0.0606	0.9033	0.0468	0.3408
Baker & Jayaram (2008)	0.366	0.105	0.5	0.109	0.2
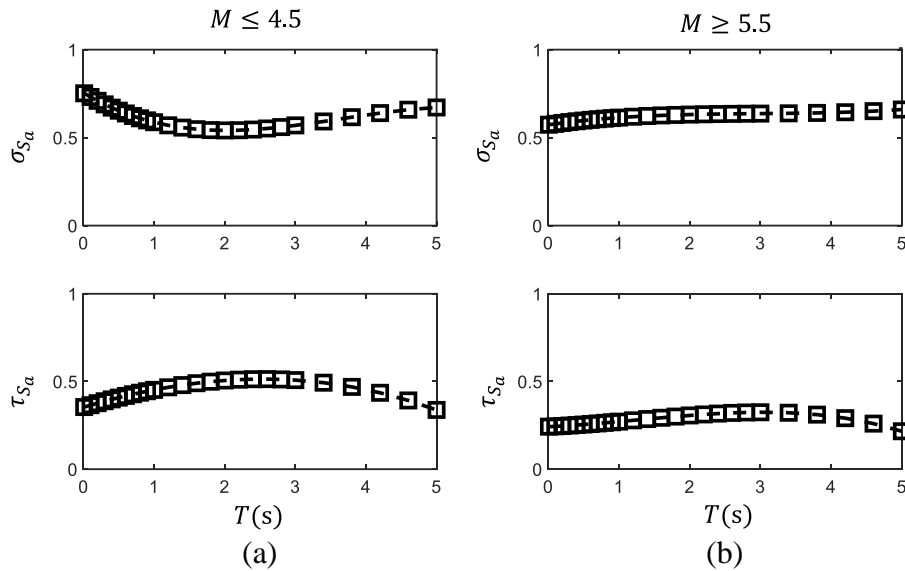
Figure 2.8 - Intra-event standard deviations ($\sigma_{Sa_{0.01}}$ to $\sigma_{Sa_{5.0}}$) and inter-event standard deviations ($\tau_{Sa_{0.01}}$ to $\tau_{Sa_{5.0}}$) of residuals of S_a with (a) $M \leq 4.5$, and (b) $M \geq 5.5$

Figure 2.9 illustrates the correlation between the residuals of $D_{5-95_{geom}}$, CAV_{geom}, and $I_{a_{geom}}$ and $RotD50S_a$ at 26 periods $\rho_{D,Sa_{0.01}}$ to $\rho_{D,Sa_{5.0}}$, $\rho_{C,Sa_{0.01}}$ to $\rho_{C,Sa_{5.0}}$, and $\rho_{I,Sa_{0.01}}$ to $\rho_{I,Sa_{5.0}}$ where Figure 9a (left column) shows the results for $M \leq 4.5$ and Figure 2.9b (right column) shows the results for $M \geq 5.5$. Generally, the residuals of $D_{5-95_{geom}}$ are observed to display slightly negative correlations with the residuals of $RotD50S_a$ while the residuals of CAV_{geom} and $I_{a_{geom}}$ show positive correlations with the residuals of $RotD50S_a$. The correlation of residuals is observed to be strong at the smaller periods (≤ 1.5 sec) and weak at the longer periods (≥ 1.5 sec). For $M \leq 4.5$, all three correlations show a good match between the computed and fitted values, except that the fitted values for $\rho_{D,Sa}$ are slightly larger than the computed ones when the period is larger than 4 sec.

32

For $M \geq 5.5$, the computed values of $\rho_{D,Sa}$ show a slight positive correlation while the fitted ones show a negative correlation when the period is larger than 3 sec. $\rho_{C,Sa}$ displays a relatively good match between fitted and computed values for all periods. The fitted values of $\rho_{I,Sa}$ show a slightly higher correlation than the computed ones when the period is larger than 3 sec.
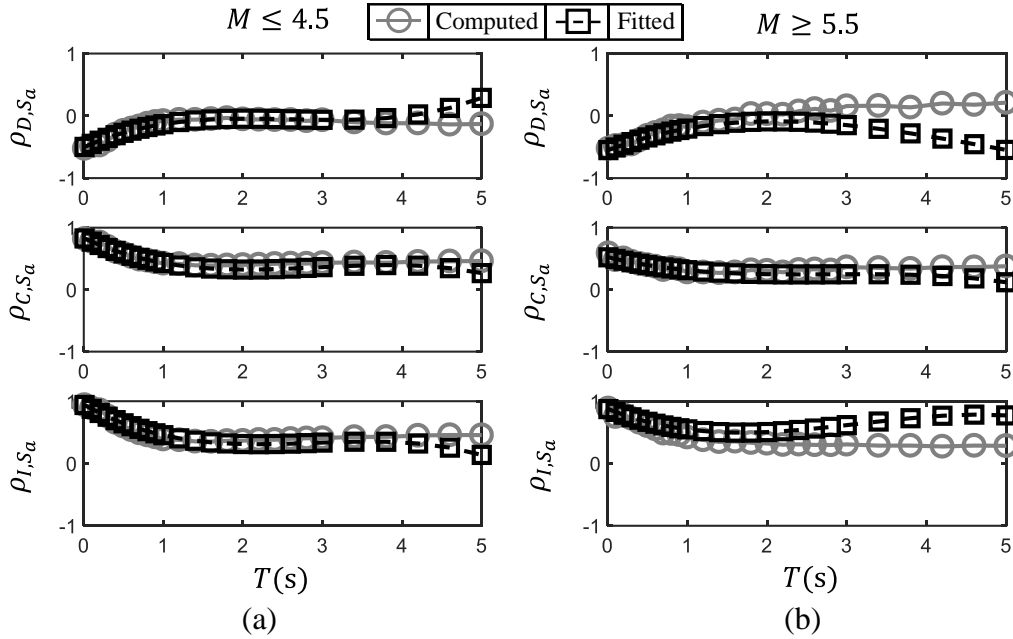
Figure 2.9 - Correlation between the residuals of $D_{5\text{-}95}$, CAV and I_a with S_a ($\rho_{D,Sa_{0.01}}$ to $\rho_{D,Sa_{5.0}}$, $\rho_{C,Sa_{0.01}}$ to $\rho_{C,Sa_{5.0}}$, $\rho_{I,Sa_{0.01}}$ to $\rho_{I,Sa_{5.0}}$) with (a) $M \leq 4.5$, and (b) $M \geq 5.5$

Figures 2.10 and 2.11 show the correlation coefficient structure of the residuals of $RotD50S_a$ at 26 periods with $M \leq 4.5$ and $M \geq 5.5$, respectively. It can be observed that for both magnitude groups, the fitted and the computed correlations match relatively well when periods are larger than 0.5 sec, and the fitted correlations are observed to be slightly higher than the computed one for periods smaller than 0.5 sec. However, the results of $M \geq 5.5$ show a better match between the fitted and computed correlations than the results of $M \leq 4.5$, especially when the periods > 0.5 sec.
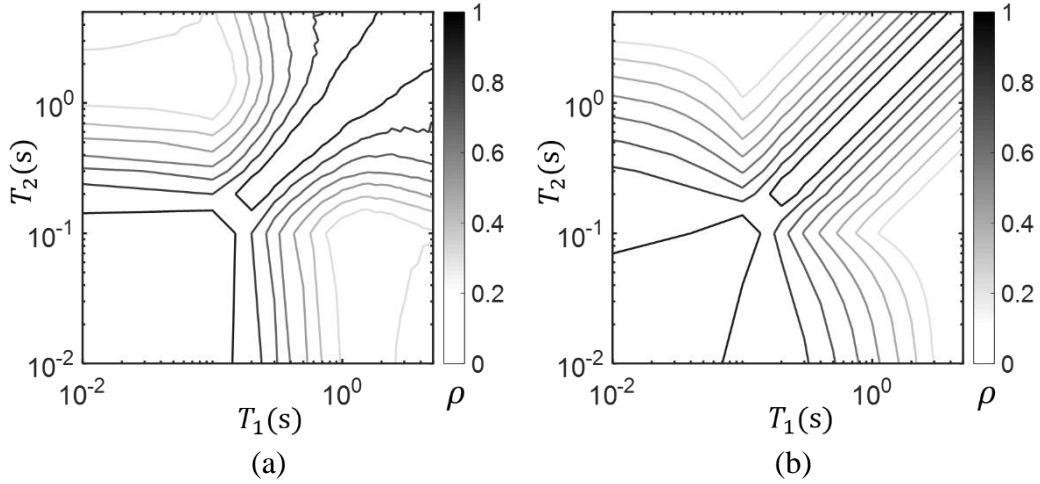
Figure 2.10 - Correlation coefficient of residuals of S_a at 26 periods with $M \leq 4.5$, (a) computed from data, (b) fitted through optimization
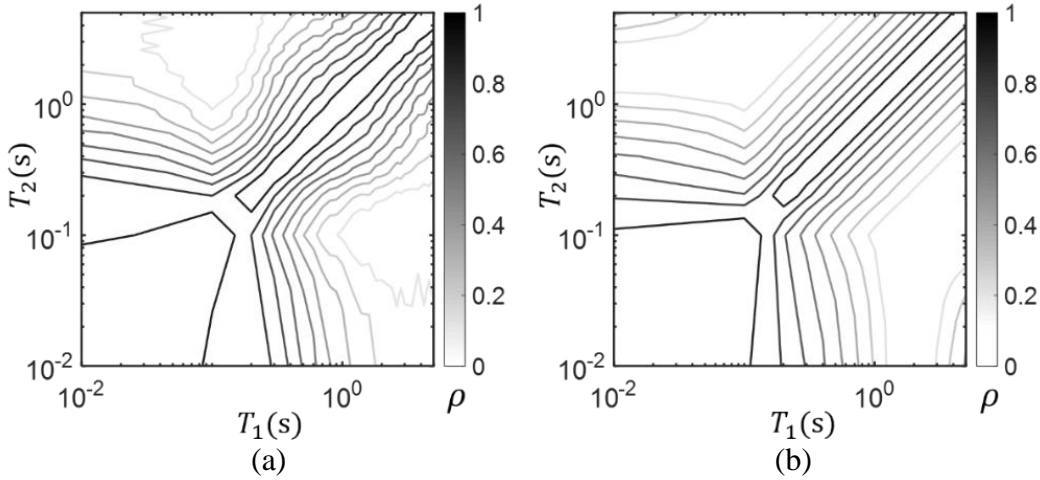
Figure 2.11 - Correlation coefficient of residuals of S_a at 26 periods with $M \geq 5.5$, (a) computed from data, (b) fitted through optimization

Figure 2.12 demonstrates the fitted standard deviation of residuals of $D_{5-95_{geom}}$ (σ_D, τ_D), CAV_{geom} (σ_C, τ_C), and $I_{a_{geom}}$ (σ_I, τ_I). It can be observed that the intra-event variability is generally smaller than the inter-event variability and the variability is lesser in the group with $M \geq 5.5$ compared to the $M \leq 4.5$ group. $I_{a_{geom}}$ is observed to possess higher variability compared to $D_{5-95_{geom}}$, CAV_{geom}, and the latter two are much more stable regarding the variation in intra- and inter-event as well as different magnitude groups. Lastly, Figure 2.13 shows the computed and fitted correlations between residuals of $D_{5-95_{geom}}$, CAV_{geom}, and $I_{a_{geom}}$ ($\rho_{D,C}, \rho_{D,I}, \rho_{C,I}$). In general, there

is a strong positive correlation between the residuals of CAV_{geom}, and $I_{a_{geom}}$, while the correlations between residuals of $D_{5-95_{geom}}$ and the other two are less stable, as the sign of the correlation changes from different magnitude groups, and the absolute value shows weak correlations. Overall, the fitted values match well to the computed values for CAV_{geom}, and $I_{a_{geom}}$.
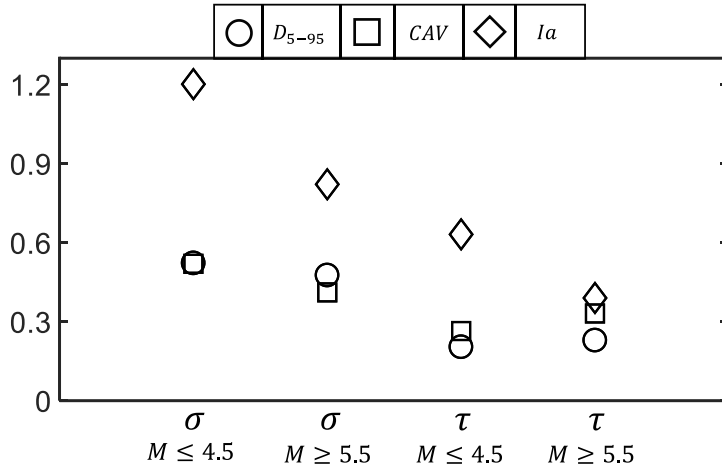
Figure 2.12 - Standard deviations of $D_{5-95_{geom}}$ (σ_D, τ_D), CAV_{geom} (σ_C, τ_C), $I_{a_{geom}}$ (σ_I, τ_I) for both $M \leq 4.5$ and $M \geq 5.5$
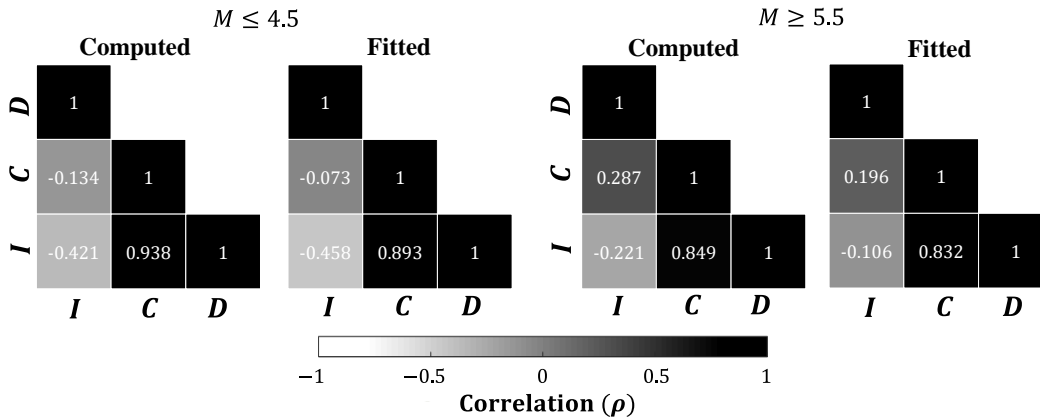
Figure 2.13 - Computed and fitted correlations between residuals of $D_{5-95_{geom}}$, CAV_{geom} and $I_{a_{geom}}$ ($\rho_{D,C}$, $\rho_{D,I}$, $\rho_{C,I}$) for both $M \leq 4.5$ and $M \geq 5.5$

2.6   Model Performance

The proposed RNN framework is evaluated firstly by comparing the recorded values of the components of **IM** with the median prediction of **IM** made by the RNN framework (median predictions of the proposed GGMPM). This is presented in Figure 2.14 for both Train and Test sets.

35

As mentioned in Section 2.3, the Train and Test sets are split in discretized based on Magnitudes. Figure 2.14a shows the predicted *RotD50S$_a$* plotted against the measured values of *RotD50S$_a$* for both Train and Test sets for all periods combined. As can be observed from the figure, the RNN framework shows good predictions for all ranges of *RotD50S$_a$* for both Train and Test sets. It can be observed from the figure that the scatter between Measured *vs.* Predicted *RotD50S$_a$* for both Train and Test sets tend to follow a 1:1 line, which leads to the conclusion that the statistical performance of the RNN framework is satisfactory. Similar observations are made in Figure 2.14b and 2.14d, that show the trend between the measured and predicted values of CAV_{geom}, and $I_{a_{geom}}$, respectively. An exception to this is observed in Figure 2.14c, due to the highly complex nature of duration and the fact that duration parameters such as $D_{5-95_{geom}}$ cannot be expected to grow with higher magnitude, leads to less accurate predictions of $D_{5-95_{geom}}$. It is mainly seen that the predictions are good for the $D_{5-95_{geom}}$ with values between 4.5 secs to 54.6 secs (logged values between 1.5 to 4), where most of the data lie. Furthermore, the performance of the RNN structure is tested by checking its goodness-of-fit using the statistical measure of R^2 in Figure 2.15. Figure 2.15a shows the values of R^2 for both Train and Test sets for *RotD50S$_a$* at 26 periods. As can be observed from the figure, the values of R^2 for all periods tend to be greater than 0.85, which shows the high predictive power of the proposed RNN framework. Also, as the R^2 from the Test set is observed to be very close to the R^2 of the Train set, it can be concluded that the RNN framework is not overfitted to the database and can be effectively used for the future predictions and hazard analysis. Figure 2.15b shows the values of R^2 for $I_{a_{geom}}$, CAV_{geom}, and $D_{5-95_{geom}}$. As can be seen, the RNN framework demonstrates good prediction power for $I_{a_{geom}}$ and CAV_{geom} with R^2 of around 0.9; however, as expected, the R^2 for $D_{5-95_{geom}}$ is observed to be much lower. But the $R^2 \sim 0.45$ is still statistically sound prediction power. Note, the **IM** is predicted such that their internal dependencies are maintained, and still, the RNN framework is observed to perform satisfactorily.

The performance of the model is further evaluated by checking the normality of the residuals obtained from the RNN framework. This is done by checking for any noticeable trends in the residuals with respect to *M* and *R$_{rup}$,* as shown in Figures 2.16 and 2.17, respectively. Figures 2.16a, 2.16b, 2.16c, 2.16d, 2.16e, and 2.16f show the *M vs.* residual plots of *S$_a$* for periods of 0.5, 1.0, and

2.0 secs, $D_{5-95_{geom}}$, CAV_{geom}, and $I_{a_{geom}}$, respectively. Figures 2.17a, 2.17b, 2.17c, 2.17d, 2.17e, and 2.17f show the R_{rup} vs. residual plots of S_a for periods of 0.5, 1.0, and 2.0 secs, $D_{5-95_{geom}}$, CAV_{geom}, and $I_{a_{geom}}$, respectively. In both Figures 2.16 and 2.17, to show any trends in the residuals, discretized trend lines are drawn showing the 25th, 50th, and 75th percentile of the residuals. It can be observed from the figures that no noticeable trends are observed in the residuals and except for $D_{5-95_{geom}}$ residuals of all other components of **IM** tend to be small.

(a)

(b)

(c)

(d)
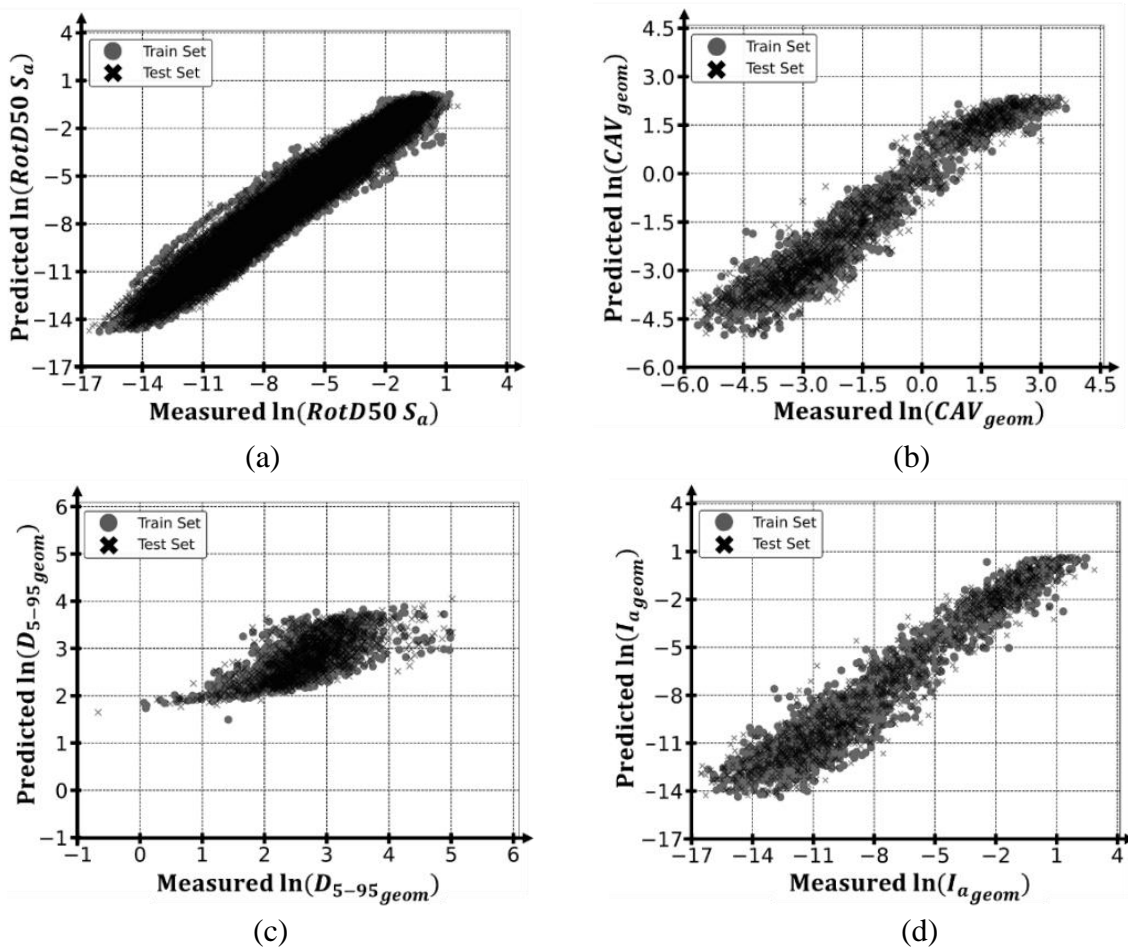
Figure 2.14 – Predicted vs Measured for Train and Test sets: (a) S_a for the 26 periods, (b) CAV, (c) $D_{5\text{-}95}$, and (d) I_a

(a)

(b)
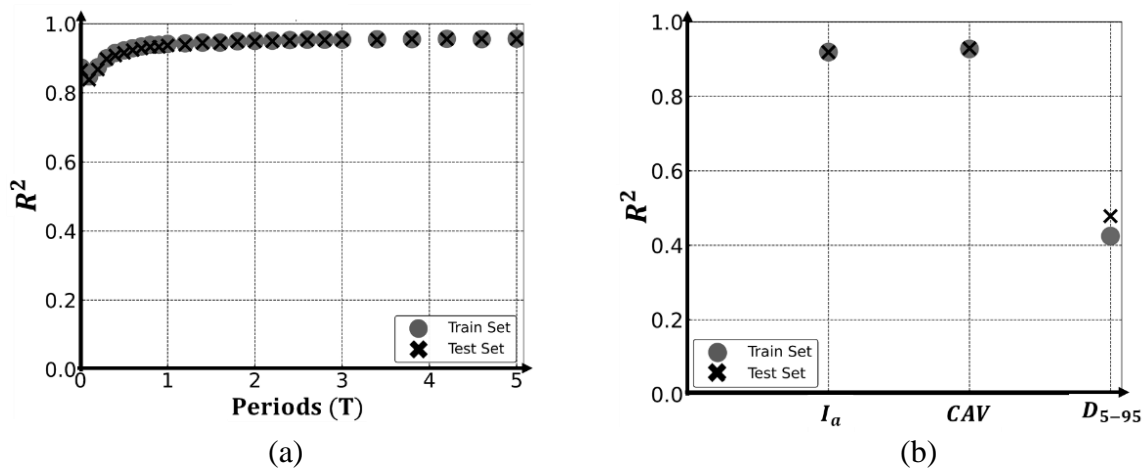
Figure 2.15 – Values of R^2 for Train and Test sets: (a) *RotD50S$_a$* for all 26 periods, (b) $I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$

(a)

(b)

(c)

(d)

(e)

(f)
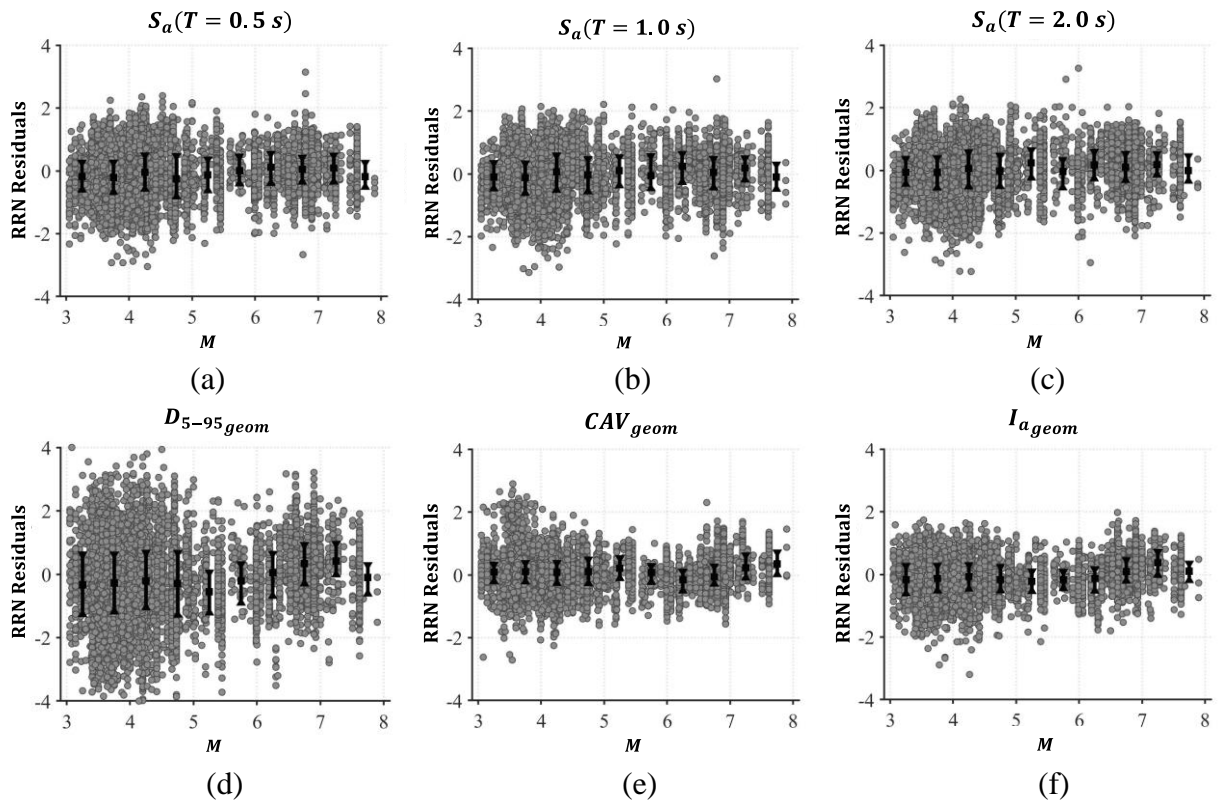
Figure 2.16 – Residuals vs *M*: (a) S_a(*T*=0.5s), (b) S_a(*T*=1.0s), (c) S_a(*T*=2.0s), (d) $D_{5-95_{geom}}$ (e) CAV_{geom}, and (f) $I_{a_{geom}}$
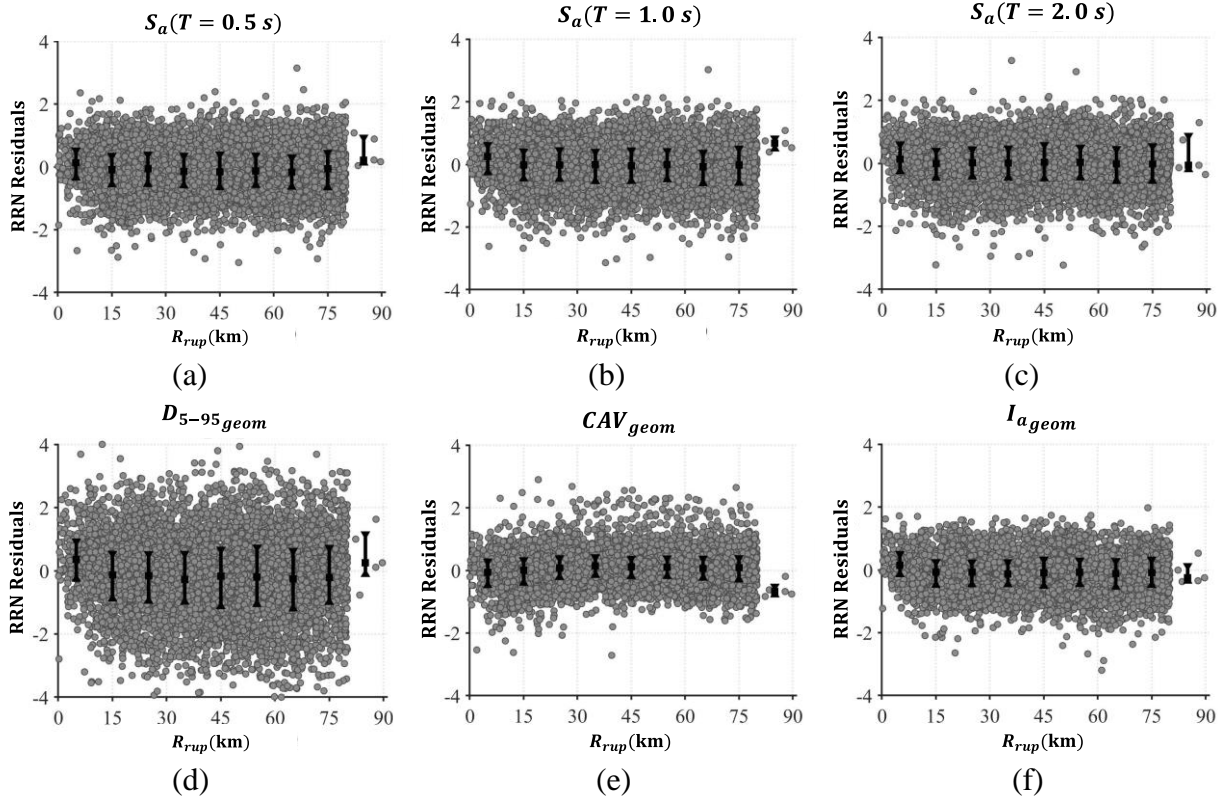
Figure 2.17 – Residuals vs R_{rup}: (a) S_a(T=0.5s), (b) S_a(T=1.0s), (c) S_a(T=2.0s), (d) $D_{5-95_{geom}}$ (e) CAV_{geom}, and (f) $I_{a_{geom}}$

2.7 Comparison against other Ground Motion Prediction Models

2.7.1 Comparison against Campbell and Bozorgnia (2014) and Conditional Mean Spectrum (CMS)

In this section, two earthquake scenarios are selected from the Test set of the NGAWest2 database; ground motion spectra generated from the proposed RNN framework (median predictions of the GGMPM) are compared against the spectra generated from CB14 (Campbell and Bozorgnia, 2014) GMPM, recorded spectra, and Conditional Mean Spectra (CMS) (Baker, J.W. (2011)) developed using CB14, conditioned on four periods ($T = 0.2$ s, $T = 0.5$ s, $T = 1$ s, and $T = 3$ s) and four epsilons ($\varepsilon = -2$, $\varepsilon = -1$, $\varepsilon = 1$, and $\varepsilon = 2$). Figure 2.18 shows the spectral comparisons for ground motion arising from Reverse-Oblique mechanism with $M = 6.9$, $R_{rup} = 18.33$ km and V_{s30}= 663.3 m/s. Figures 2.18a, 2.18b, 2.18c, and 2.18d show the recorded spectrum along with the spectrum generated from the RNN framework with four CMS conditioned on $T = 0.2$ s, $T = 0.5$ s, $T = 1$ s, and $T = 3$ s, respectively. Since the CMS are conditioned on the period (T) and ϵ, the four
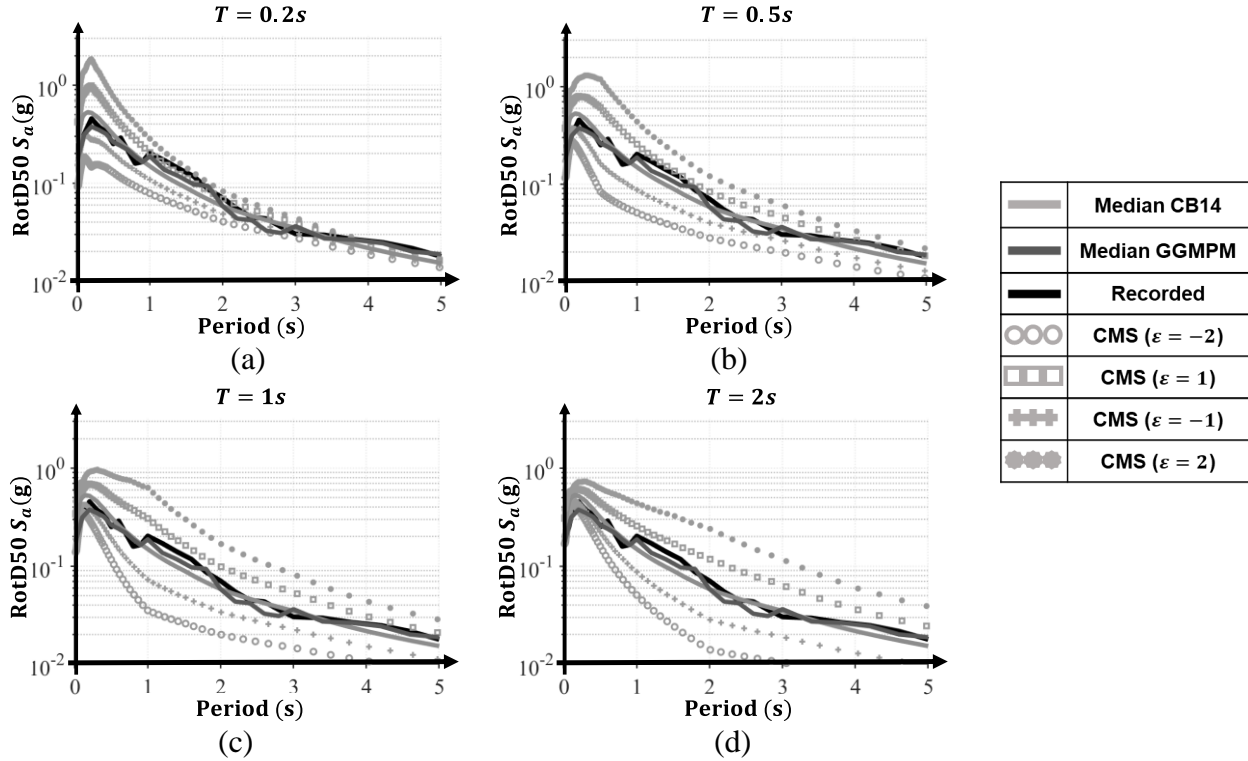
Figure 2.18 – Spectral Comparisons for ground motion from Reverse-Oblique mechanism with $M = 6.9$, $R_{rup} = 18.3$ km and $V_{s30} = 663.3$ m/s for CMS of: (a) $T = 0.2$ sec, (b) $T = 0.5$ sec, (c) $T = 1$ sec, and (d) $T = 2$ sec

sub-figures further contain the CMS conditioned on the four epsilons ($\varepsilon = -2$, $\varepsilon = -1$, $\varepsilon = 1$, and $\varepsilon = 2$) and the median estimate of CB14. As can be observed from the figures, the median spectrum generated from the GGMPM (*i.e* RNN predictions) lies very close to the recorded spectrum. Based on Figure 2.18, it can be observed that having additional constraints on the period (T) and ε can cause undesired variability in the estimation of the ground motion spectrum. It is observed that CB14 or CMS conditioned on $\varepsilon = 0$ show a closer match with respect to the recorded spectrum. It can be further observed that all four CMS and CB14 tends to be smoother than the recorded spectrum and differ from the recorded spectrum at several periods. Similar trends are observed for another scenario, shown in Figure 2.19, obtained from the Test set for the ground motion arising from a Strike-Slip mechanism with $M = 6.0$, $R_{rup} = 13$ km, and $V_{s30} = 349.9$ m/s. The spectrum generated from the RNN framework is estimated very close to the spectrum of the recorded ground motion, and high variability is observed in CMS due to variation across the periods (T) and ε. CB14 tends to estimate a smoothened version of the spectrum; however, it consistently underestimates the Sa for

mid-range periods (from $T = 1.5$ to $T = 3$ secs). In general, it is observed that CMS conditioned on various periods causes some variability in the predicted spectra. For example, comparing Figures 2.18a, 2.18b, 2.18c, and 2.18d or Figures 2.19a, 2.19b, 2.19c, and 2.19d, the generated CMS spectra show a wide range of predictions for different values of T and ε
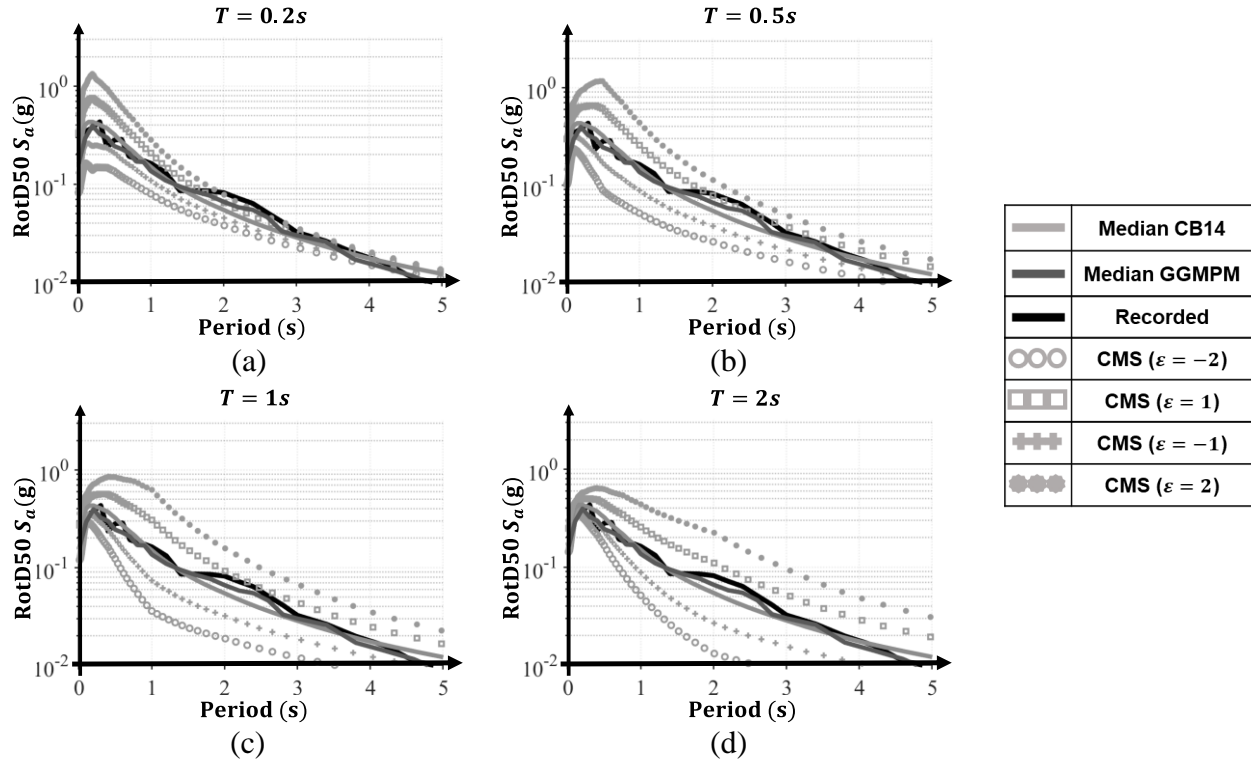
Figure 2.19 – Spectral Comparisons for ground motion from Strike-Slip mechanism with $M = 6.2$, $R_{rup} = 13$ km and $V_{s30} = 349.9$ m/s for CMS of: (a) $T = 0.2$ sec, (b) $T = 0.5$ sec, (c) $T = 1$ sec, and (d) $T = 2$ sec

2.7.2   Comparison against Campbell and Bozorgnia (2019) and Afshari and Stewart (2016)

Furthermore, the RNN predictions of $I_{a_{geom}}$, CAV_{geom}, and $D_{5-95_{geom}}$ of the two events selected in the previous section are compared against the recorded values in Figures 2.20a and 2.20b. Note that the spectra shown in Figure 2.18 and $I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$ shown in Figure 2.20a belongs to the same ground motion. Similarly, the spectra shown in Figure 2.19 and $I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$ shown in Figure 2.20b belong to the same scenario. The median predictions of the GGMPM (*i.e* RNN predictions) for $I_{a_{geom}}$, CAV_{geom} is further compared against the median predictions of CB19 (Campbell and Bozorgnia, 2019) and $D_{5-95_{geom}}$ is compared to AS16 (Afshari

and Stewart, 2016). As can be observed from Figure 2.20, even while maintaining the correlations between the IMs, the RNN framework tends to predict $I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$ much closer to the recorded IMs as compared to the CB19 and AS16 GMPMs.

| O | Recorded | □ | Median CB19 |
| X | Median GGMPM | ◇ | Median AS16 |

(a)

| O | Recorded | □ | Median CB19 |
| X | Median GGMPM | ◇ | Median AS16 |

(b)
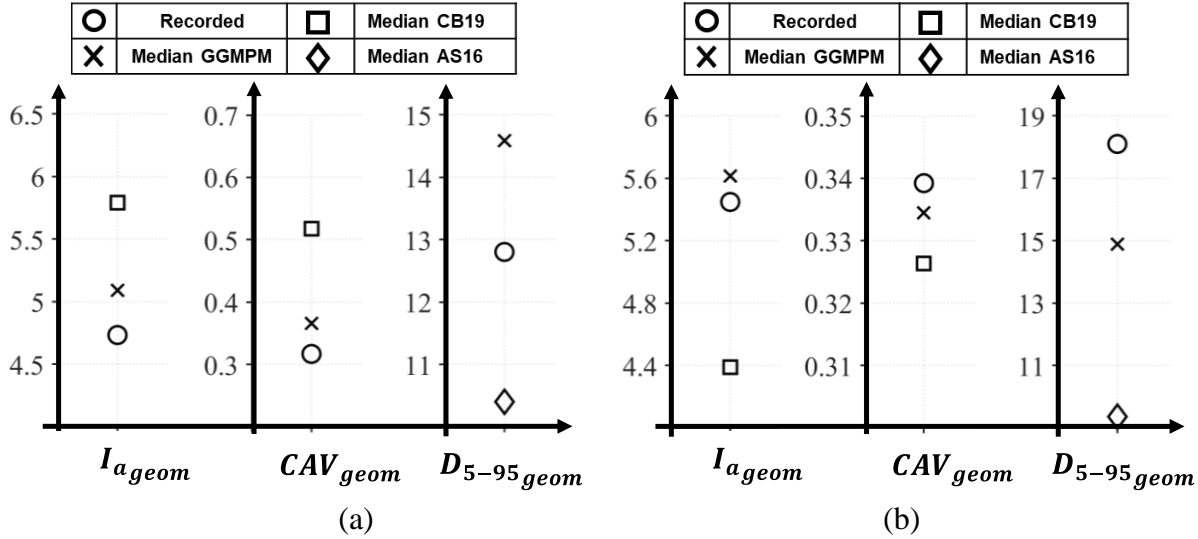
Figure 2.20 – $I_{a_{geom}}$, CAV_{geom}, $D_{5-95_{geom}}$ comparisons for: (a) Reverse-Oblique mechanism with $M = 6.9$, $R_{rup} = 18.3$ km and $V_{s30} = 663.3$ m/s, and (b) Strike-Slip mechanism with $M = 6.2$, $R_{rup} = 13$ km and $V_{s30} = 349.9$ m/s

2.8   Spectral Comparisons

The proposed GGMPM (i.e., RNN framework and covariance matrices) is compared against the Conditional Spectrum (CS) proposed by Baker, J.W. (2011) for the estimations of spectral accelerations. This is done by using the records of two seismic events, including Northridge (1994) ($M = 6.4$, 140 station recordings) and Landers (1992) ($M = 7.3$, 70 station recordings). For the two events, GGMPM and CS are used to construct the median and $\pm$ 1 standard deviation of S_a for all recording stations. Since the covariance matrices of the GGMPM describe the covariance between the 29 target IMs, the variability of all IMs conditioned on one IM can be estimated using Equation 2.20 as done by Baker, J.W. (2011). In Equation 2.20, $\gamma_{IM_i|IM^*}$ represents the conditional standard deviation of residual between the ith IM and IM^*, and ρ represents the correlation between them. In Equation 2.21, $\sigma_{IM_i|IM^*}$ represents the $\pm$ 1 standard deviation of the ith IM conditioned on $\pm$ 1 standard deviation of IM^*. The event parameters of the three seismic events are also used to construct Conditional Spectrum (CS) based on CB14 GMPM. For both GGMPM and CS, the

42

spectra are constructed for 3 periods ($T = 0.2$, 0.5 and 2.0 secs). It should be noted that the GGMPM framework does not require conditioning of T to obtain the median prediction; however, the estimation bands are conditioned on the variance of the selected *IM**. Since the construction of Conditional Spectrum (CS) also requires an appropriate epsilon, 40 CMS were constructed using epsilons of -2 to 2 with an interval of 0.1, and the one with a minimum sum of squared error (SSE) with respect to the recorded spectrum was used to construct the CS in the figures and compared against the true spectrum and estimate of GGMPM framework. Figure 2.21 shows some examples of the spectrum of recorded ground motion against the estimates of GGMPM and CS (i.e., CMS $\pm\sigma$) for the two events conditioned on T=0.2 secs. This is done for all stations of the three events for all three periods.

$$\gamma_{IM_i|IM^*} = \sqrt{1 - \rho^2(IM_i, IM^*)} \tag{2.20}$$

$$\sigma_{IM_i|IM^*} = \sigma_{IM_i}(\rho^2(IM_i, IM^*) + \sqrt{1 - \rho^2(IM_i, IM^*)}) \tag{2.21}$$

$$\ln L = \sum_{i=1}^{n=26} \left[-\frac{\ln(2\pi)}{2} - \frac{\ln(\sigma_i^2)}{2} - \frac{1}{2\sigma_i^2}(x_i - \mu_i)^2 \right] \tag{2.22}$$
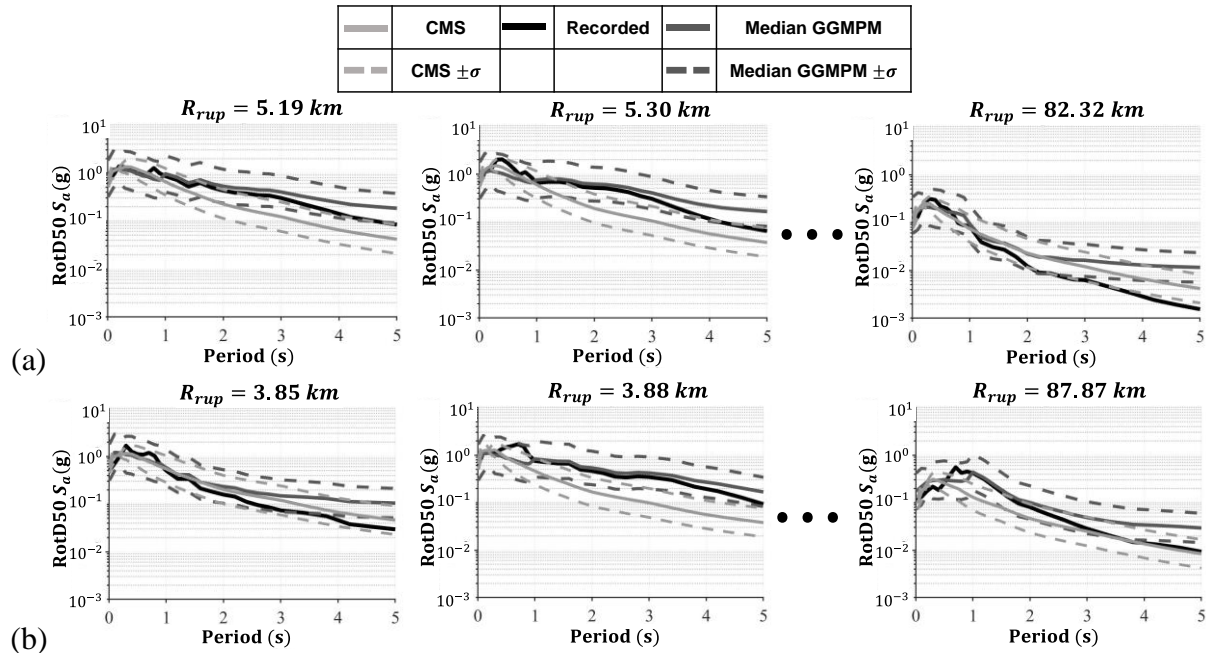
Figure 2.21 – Recorded spectra against estimates of GGMPM and CS conditioned on $T = 0.2$ secs for: (a) Northridge and (b) Landers seismic events.

43
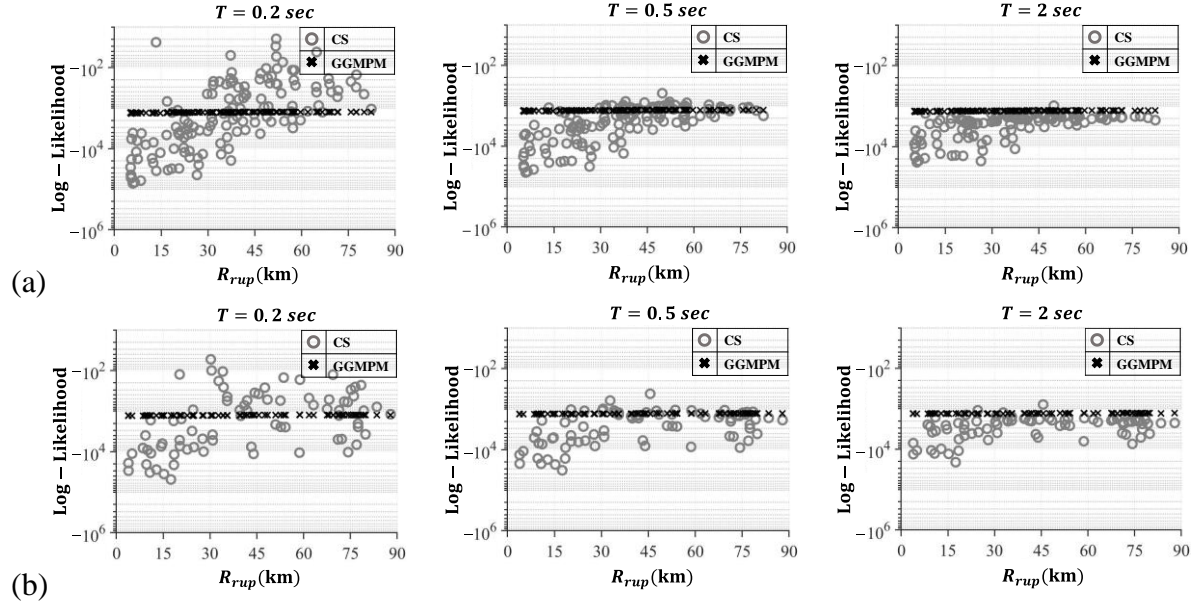
Figure 2.22 – Log-likelihood of the Recorded spectra to fall within the bands of estimation of GGMPM and CS for: (a) Northridge and (b) Landers, with IM^* as S_a at $T = 0.2$, 0.5, and 2.0 secs

The predictions of the GGMPM and CS are compared by computing the log-likelihood of the recorded spectrum at each recording station to fall within the bands of estimation of GGMPM and CS, using Equation 2.22. In Equation 2.22, the log-likelihood for each of the 26 periods is calculated (term inside the brackets) and then summed across the entire spectrum to achieve one estimate per recording. Hence higher log-likelihood represents better predictive power. The results of this analysis are presented in Figure 2.22 for the two seismic events for three periods. For all the cases, it is observed that the estimation of GGMPM has a relatively stable log-likelihood indicating that there is not relative bias in the predictions with respect to the M or R_{rup} of the events. For both scenarios, it is observed that for the period of 0.2 secs, the CS performs better for stations with large R_{rup}, while for the shorter distances, the GGMPM performs better. It is also observed from these figures that with the increase in the period, the estimation power of CS for larger R_{rup} is reduced as compared to the GGMPM. Specifically, for the period of 2 secs, the GGMPM framework consistently leads to higher log-likelihood as compared to CS. Also, it can be observed that the log-likelihoods of the CS between the two events are highly variable as compared to the proposed GGMPM. Hence, based on a brief comparison presented here, it can also be derived that the proposed GGMPM is an improved method for estimating the spectra of ground motions of an

44

earthquake without large variability in predictive power with respect to the distance to the fault rupture or magnitude of the seismic event.

2.9   Conclusions

This chapter presents a non-parametric Generalized Ground Motion Prediction Model (GGMPM) to predict a vector of ground motion intensity measure (denoted as **IM)** with 29 components using 8 inputs of earthquake event and site parameters (including M, R_{rup}, R_x, R_{JB}, D_{Hyp}, Z_{TOR}, V_{s30}, F) that describe the physics of the rupture. The suggested **IM** includes geomeans of I_a, CAV, and D_{5-95}, which comprise three out of 29 components of the suggested **IM** obtained from pairs of two horizontal components of ground motion records. The other 26 components of **IM** include 5% damped $RotD50S_a$ at 26 periods. Characterization of ground motion intensity as a vector was brought to the mainstream research arena after Baker and Jayaram (2008); the suggested GGMPM for estimation of **IM** is a step in that direction; it incorporates both time and frequency domain ground motion intensity measures with their correlation in a single stand-alone model.

GGMPM for the suggested **IM** is developed using the hybrid framework of Recurrent Neural Network (RNN) and Covariance Matrix Adaptation Evolution Strategy (CMA-ES); In particular, Long-Short-Term-Memory (LSTM) cells are used to incorporate the higher-order dependencies within **IM** and optimization technique of CMA-ES is further used to describe the aleatory variability among the components of **IM**. The aleatory variability of the framework is expressed in two components: inter-event variability and intra-event variability. The two types are variabilities are expressed in terms of covariance matrices, which are optimized using the CMA-ES algorithm, separately for $M \leq 4.5$ and $M \geq 5.5$, and linear interpolation can be performed for $4.5 < M < 5.5$. The developed GGMPM is carefully trained and methodologically tested with a subset of NGA-West2 ground motion database containing mainshock records with $3.05 \leq M \leq 7.9$. The used subset of ground motions has a significant overlap with the set used in Campbell and Bozorgnia (2014). Detailed validation exercises are carried out to demonstrate the adequacy of the developed GGMPM. The superiority of the GGMPM lies in its capability of estimating the vector of ground motion intensity measures with an acceptable representation of the correlation between the estimated intensity measures. By using illustrative examples, it is shown that only for short period systems in large distances, spectra developed using Conditional Spectrum is a better match

to recorded *RotD50 S_a* spectrum compared with the suggested GGMPM; otherwise, (i.e., for medium to short distances, with medium to long period structures) GGMPM provides a better match to the recorded ground motion spectra. Also, the GGMPM performs satisfactorily against the predictions of AS16 and CB19 for geomeans of D_{5-95}, I_a, and *CAV*. Based on all the results, it is observed that the proposed GGMPM demonstrates acceptable performance in predicting the vector of intensity measures while maintaining their internal correlations. Furthermore, another advantage of this framework is that it can be easily re-trained with other ground motions records or extended to a larger vector of IMs, that can include any additional IMs of interest. The proposed GGMPM can be readily utilized in any modern seismic hazard analysis toolbox. An executable application is developed to obtain estimations from the proposed GGMPM.

Chapter Citation:
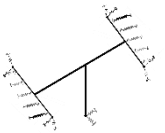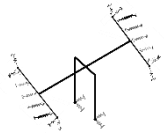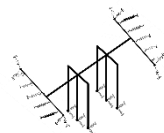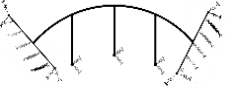
1) Fayaz J., Xiang Y., and Zareian F. (2020d). Generalized Ground Motion Prediction Model (GGMPM) Using Hybrid Neural Networks. *Earthquake Engineering and Structural Dynamics.* Volume 50, Issue 6, Pages 1539-1561.

CHAPTER 3

BRIDGE STRUCTURAL MODELS AND SITE DATABASE

3.1 Inventory of Bridges

Table 3.1 - Characteristics of Bridge Structures

Bridge	A	B	C	F
Name	Jack Tone Road Overcrossing	La Veta Avenue Overcrossing	Jack Tone Road Overhead	E22-N55 Connector Over-crossing
Total Length	220.6 ft	300.0 ft	418.0 ft	500.0 ft
Number of Spans	2	2	3	4
Column Bent	Single-column	Two-column	Three-column	Single-column
Column Radius	33.1 in	33.5 in	33.1 in	47.7 in
Column Height	22.0 ft	22.0 ft	24.1 ft	18.5 ft
Reinforcement of Column Section	Long: 44 #11 (bundles of 2) Trans: Spiral, #6 @ 3.34 in	Long: 44 #11 (bundles of 2) Trans: Spiral, #4 @ 6.00 in	Long: 34 #14 (bundles of 2) Trans: Spiral, #7 @ 4.52 in	Long: 42 #14 (bundles of 2) Trans: Spiral, #7 @ 2.95 in
Fundamental Period	0.61 sec	0.83 sec	0.79 sec	1.11 sec
Configuration				

This study is focused on Caltrans ordinary standard bridges. A bridge is considered as an ordinary standard bridge if it satisfies the following five conditions: (1) the span length should be less than 300 feet; (2) the bridge should be constructed with normal-weight concrete; (3) foundations must be supported on spread footings, pile caps with piles, or pile shafts; and (4) the soil is not susceptible to liquefaction, lateral spreading or scour. This chapter provides a brief description of the selected ordinary standard bridge structures and the related Opensees (McKenna *et al.* 2010) modeling technique based on the first generation of guidelines for modeling ordinary bridges.

Four representative Reinforced Concrete (RC) ordinary bridge structures are selected for the statistical analysis. Table 3.1 includes the details of the four ordinary bridges with seat-type abutments, which reflect the common bridge engineering practice in California. The selection was focused on modern structures (designed after 2000) and based on the need to have a manageable number of structures and, at the same time, to cover the most common bridge geometries encountered within the Caltrans highway bridge inventory. The first selected bridge is the *Jack Tone Road Overcrossing* (Bridge A) located at the city of Ripon, with two spans supported on a single column. The second bridge is the *La Veta Avenue Overcrossing* (Bridge B) located at the city of Tustin, with two spans supported on a two-column bridge bent. The third bridge is the *Jack Tone Road Overhead* (Bridge C) located at Ripon, with three spans and two three-column bridge bents. And the fourth bridge is the curved bridge *E22-N55 Connector Over-crossing* (Bridge F) located in Santa Ana, with four spans supported on single columns.

3.2 Analytical Modelling

Finite Element models of the bridges are developed in the OpenSees (McKenna *et al.* 2010). The seismic demand of a bridge is estimated by developing and analyzing a mathematical model of the superstructure and substructure of the bridge subjected to representative ground motions. The models represent the geometry, boundary conditions, mass distribution, energy dissipation as well as the interaction between elements. Since the bridge consists of many components that exhibit nonlinear behavior, a fully 3D nonlinear model is developed. The finite element models are comprised of: seat-type abutments, abutment piles, shear keys, column bents, elastomeric bearing pads, backfill soil, and superstructure. A detailed representation of the Opensees finite-element model of the bridge structures is given in Figure 3.1. The concrete and steel used in modeling possess a compressive strength $f_c' = 5.0$ *ksi* (34 *MPa*) with a modulus of elasticity $E_s = 4030.5$ *ksi* (27.8 *GPa*) and yield strength = 65 *ksi* (448 *MPa*) with a modulus of elasticity $E_s = 29000$ *ksi* (200 *GPa*), respectively.
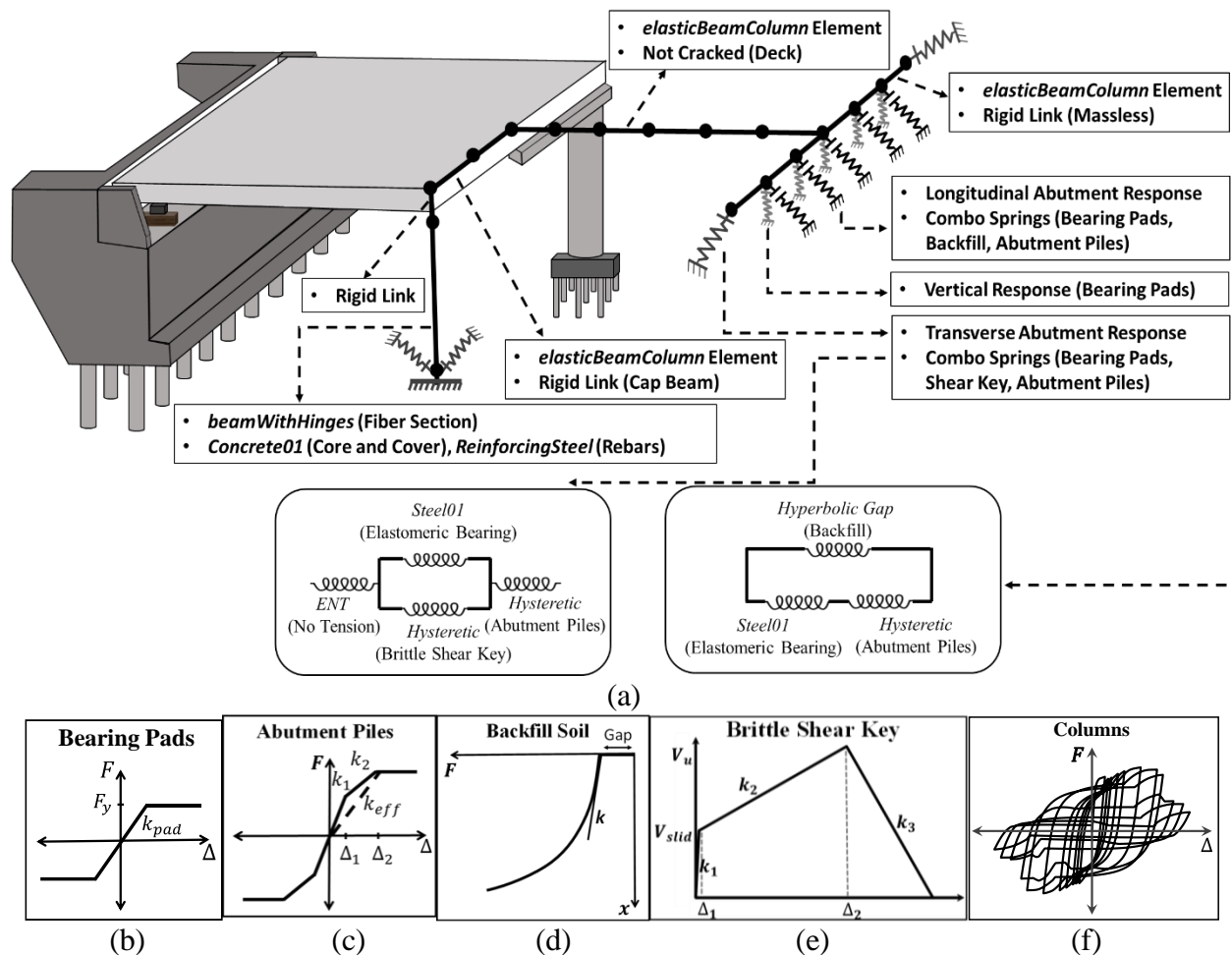
Figure 3.1 - a) Details of the finite-element model of bridges, b) Bearing pads response, c) Abutment pile response, d) Backfill soil response, e) Shear key response, and f) Column response

3.2.1 Superstructure

The superstructure is modeled using linear elastic elements since Caltrans SDC (2013) recommends, and designs, the superstructure to remain elastic under the seismic excitations. Dependent on the bridge type, the superstructure is either prestressed concrete beams or steel girders with a concrete slab. It should be noted that the stiffness of the superstructure does not have a significant effect on the seismic response of the bridge since the longitudinal response is typically governed by the bearings, columns, abutments, and foundation. As the superstructure is designed to remain elastic when undergoing seismic events, the superstructure is modeled with `elasticBeamColumn` elements using uncracked section properties of the deck of the bridge. To capture the dynamic response accurately, the mass of the superstructure is discretized and

distributed throughout the length of the deck with each span's mass being distributed in ten intervals connected by the `elasticBeamColumn` elements.

3.2.2 Column Bent

Single (SCB) and multi-column bents (MCB) with seat-type abutments are the most common types of bridges designed in California. The bent caps are modeled using `elasticBeamColumn` elements having high torsional, in-plane and out-of-plane stiffnesses, while the columns are modeled using `beamWithHinges` element (two Gauss integration points) with fiber-discretized cross-sections to model confined concrete for the core, unconfined concrete for the cover and steel rebars. Fiber-defined cross-sections have the unique advantage of allowing the specification of material properties specific to different locations in a member cross-section. For instance, unconfined concrete properties are assigned to the cover concrete while confined concrete properties are assigned to the core fibers. Further, the precise location of the longitudinal reinforcing bars and prestressing strands may be specified, and material properties assigned to them. Furthermore, rigid links are used to connect the top of the column to the bent cap beam and translation springs representing the behavior of foundations are applied at the base of the columns. Figure 3.2a presents the finite element discretization of the bents for the bridge structures. The details of the models of the structural components are presented in the sections that follow. The hysteretic behavior of columns is given in Figure 3.1f.
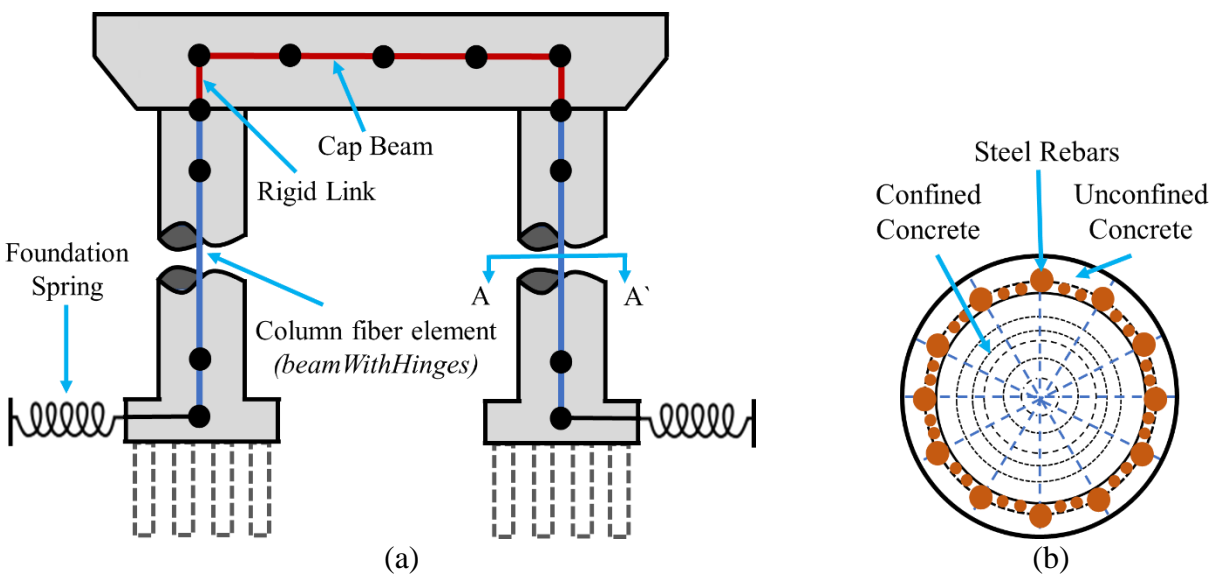
Figure 3.2 - a) Details of Bent, b) Column fiber section at AA`

3.2.2.1 Column Section and Bent Cap

The bridge columns are modeled using displacement-based beam-column elements for all the bridges. The modeling is done using the `beamWithHinges` element (two Gauss integration points) with fiber-discretized cross-sections to model confined concrete for the core, unconfined concrete for the cover and steel rebars. The plasticity of the columns is concentrated at two plastic hinges at the opposite ends connected by a linear elastic element. The length of the plastic hinge is determined based on Caltrans SDC (2013). Details such as column diameter, longitudinal and transverse reinforcement ratios vary across the four bridges. Three different constitutive rules are used simultaneously within a fiber-discretized cross-section: (i) confined concrete for the core concrete, (ii) unconfined concrete for the cover concrete, and (iii) steel rebar for the reinforcing bars. The Opensees `Concrete01` constitutive model is used for both the cover and core concrete, and steel rebars are modeled by the `ReinforcingSteel` material. A rigid element is attached to the top of the nonlinear beam-column element to model the portion of the column-bent embedded in the superstructure. Figure 3.2b shows a discretized fiber section for a bridge column, which consists of unconfined and confined concrete properties assigned to the fibers along with a precise location of the longitudinal reinforcement. Assuming a monolithic construction of the cap beam and columns, the cap beam is modeled as a rigid bent using an `elasticBeamColumn` element with high torsional, in-plane and out-of-plane stiffnesses. The concrete and steel are modeled using `Concrete01` and `ReinforcingSteel` materials, respectively, which are available in Opensees.

3.2.2.2 Column Bases and Piles

The boundary condition of the column base proves to introduce a significant impact on the seismic responses obtained from NLTHA. In current models, the single-column bridges (Bridge A and Bridge F) are modeled with a fixed base connection, while pinned base connections are assigned to the multi-column bridges (Bridge B and C). However, the flexibility of base connections arising from piles beneath is specifically modeled. To describe the translational behavior of pile foundations, a linear elastic spring model is used. The vertical movement is restricted since vertical ground motion are not applied to the bridge structures in this study. The response of piles laterally loaded (by horizontal forces) is independent of their length, in most practical situations. Only the

uppermost part of the pile experiences appreciable displacement. It is along this active length, l_c, that the imposed load is transmitted to the supporting soil. The length l_c is typically on the order of 5 to 10 pile diameters and is a function of the pile length with respect to the soil. Equation 3.1 presents simple algebraic expression, given by Fang (1999), for estimating l_c of a circular solid pile with diameter d, Young's modulus E_p, and $\widehat{E_s}$ is the reference Young's modulus of the soil at a depth, $z = d$.

Equation 3.2 is then used to calculate the horizontal stiffness, K_h, of a pile foundation, but it is valid only for piles with length $L > l_c$. In Equation 3.3, N represents the number of piles in a footing and $K_{h,total}$ is the total horizontal stiffness of a pile footing. The reference modulus of sand varies from 34.5 Mpa (5 ksi) to 68.9 Mpa (10 ksi). In this study, the minimum reference modulus of sand, 34.5 Mpa (5ksi), is used for conservatism. The piles are assumed to be made of concrete (E_p = 21994 Mpa, 3190 ksi). With the above information combined with the bridge data, the horizontal stiffness of the column foundation is calculated for all the bridges.

$$l_c = 2\, d\, \left(E_p \big/ \widehat{E_s}\right)^{0.22} \tag{3.1}$$

$$K_h = 0.8\, d\, \widehat{E_s}\, \left(E_p \big/ \widehat{E_s}\right)^{0.28} \tag{3.2}$$

$$K_{h,total} = N.K_h \tag{3.3}$$

3.2.3   Shear Keys

In the transverse direction, exterior and interior shear keys are designed to provide transverse support to bridge superstructure and transmit the lateral shear forces through vertical reinforcement between the shear key and the stem wall in minor-to-moderate earthquake events and service loads. These elements are designed to break off once subject to strong motions to save the abutment structure and piles. Shear-key design philosophy requires that these elements behave similarly to a fuse. In essence, shear-keys are designed to resist lateral service loads with minor deformation and break under extreme loads to save the stem wall and the supporting piles. Hence, the shear keys are designed and modeled in a brittle/isolated manner using the `hysteretic` spring model available in Opensees. The model is defined with a trilinear backbone curve as given in Figure

3.1e. The shear key is designed as per Caltrans SDC (2013) with an area of vertical reinforcement (A_{vsk}) calculated as per Equation 3.4, where $P_{dl}{}^{sup}$ is the superstructure dead load reaction at the abutment and f_{ye} is the yield strength of steel rebars.

Based on past experimental observations detailed in Kottari (2016), the sliding shear resistance of an isolated shear key can be associated with two states: i) shear resistance at first sliding (V_{slid}) and ii) ultimate sliding shear resistance (V_u) right before the rupture of the dowel bars. Assuming a smooth construction joint, the shear resistance due to the dowel action (F_d) of the vertical dowel bars is calculated using Equation 3.5 which leads to the calculation of V_{slid} as shown in Equation 3.9 through Equations 3.6, 3.7 and 3.8. Based on the equilibrium of the horizontal and vertical forces (Borzogzadeh *et al.* 2006), V_u is calculated as per Equation 3.10

$$A_{vsk} = \frac{\alpha \times P_{dl}{}^{sup}}{1.8 \times f_{ye}} \qquad 0.5 < \alpha < 1 \tag{3.4}$$

$$F_d = \sum_{i=1}^{\#\,of\,Vertical\,bars} \sqrt{2.\,M_{pl,i}.\,f_{cb,i}.\,d_{b,i}} \tag{3.5}$$

$$M_{pl,i} = \frac{f_y.\,d_{b,i}{}^3}{6} \tag{3.6}$$

$$f_{b,i} = a_i.\,f_c{}^{1.2} \tag{3.7}$$

$$a_i = 2.0 + \frac{0.5}{d_{b,i}} \tag{3.8}$$

$$V_{slid} = \frac{T + F_d}{\left(1 - \mu_f.\,tan\beta\right)} \tag{3.9}$$

$$V_u = \frac{\mu_f.\,cos\alpha + sin\alpha}{1 - \mu_f.\,tan\beta}.\,A_{vsk}.\,f_{su} \tag{3.10}$$

In these equations, $M_{pl,i}$ is the plastic moment capacity of bar i, $f_{cb,i}$ is the compressive strength of confined concrete, f_c is the uniaxial concrete compressive strength, $d_{b,i}$ is the diameter of bar i, β is the angle of the inclined face of the shear key with respect to a vertical plane, T is the cohesive force,

and μ_f is the coefficient of friction of the smooth construction joint ($= 0.36$) (Kottari, 2016). α is the angle of inclination of the vertical dowel bar with respect to the vertical (angle of kink). It is assumed that a bond breaker is applied on the construction joint, hence $T = 0$ in Equation 3.9. The value of α is obtained from Kottari (2016) through interpolation for the provided diameter of dowel bars. The initial stiffness (k_1) of the backbone curve is computed through the summation of shear and flexural responses of the concrete cantilever action of the shear key (Omrani *et al.* 2015), while the stiffnesses of the hardening (k_2) and softening (k_3) branches are expressed as a percentage of k_1 (ranging from 0.5 % to 2.5% for various rebar diameters) that is interpolated according to Kottari (2016)

3.2.4 Abutment Piles

As stated previously, piles are considered to provide longitudinal and transverse stiffness to the abutments. For the passive longitudinal response, piles act in parallel with the backfill soil, while piles alone account for the active resistance. The transverse resistance just like the active resistance is also provided solely by the piles. The abutment stiffness in active action is dependent on the resistance of piles. The Caltrans recommendation for pile stiffness (Caltrans, 1990), 119 kN/pile (40 kips/pile), is considered in this study. However, the behavior of abutments in active action is not unilinear to the ultimate strength. The initial stiffness in active action degrades with surface soil yielding. Therefore, in this study, a trilinear response stemming from the recommendations of Choi (2002) is used to model the response of the piles.

The model assumes that piles become plastic at a deformation of 25.4 mm (1 inch), first yielding occurs at a displacement equal to 30% of the ultimate deformation, and the yielding force is 70% of the ultimate force. The force deformation response of the pile along with the model parameters is presented in Figure 3.1c. The piles are modeled in Opensees through a trilinear `hysteretic` spring model whose parameters are given in Equations 3.11 to 3.15.

$$K_{eff} = 7.0 \text{ kN/mm/pile} \times \text{Number of Piles} \tag{3.11}$$

$$K_1 = 2.333 \times K_{eff} \tag{3.12}$$

$$\Delta_1 = 7.62 \; mm \; (0.3 \; in) \tag{3.13}$$

$$K_2 = 0.428 \times K_{eff} \tag{3.14}$$

$$\Delta_2 = 25.4 \ mm \ (1.0 \ in) \tag{3.15}$$

3.2.5 Backfill Soil

The hyperbolic soil model proposed by Shamsabadi and Kapuskar (2006) is used in the present study to capture the response of the abutment backwall soil in passive response. The model is based on experimental testing of bridge abutments with 1.67 m (5.5 ft) high backwalls and typical cohesionless and cohesive backfill soils conducted at the University of California Los Angeles. The test results were then extended to develop closed-form solutions for the abutment backfill soil response for a range of backwall heights based on a series of analyses using the limit-equilibrium method that implements mobilized logarithmic-spiral failure surfaces coupled with a modified hyperbolic soil stress-strain behavior. Figure 3.1d shows a typical abutment force-displacement backbone curve, where F_{ult} is the maximum abutment force developed at maximum displacement, y_{max}, and which possesses an initial stiffness k. Equation 3.16 presents the closed-form solution (used in this study) for the force-displacement response of the backfill soil, where parameters β, α, n, and η are expressed in Equations 3.17 to 3.20 as suggested by Khalili-Tehrani *et al.* (2010) in terms of the soil parameters of cohesion (c), friction angle (φ), unit weight (γ), strain corresponding to 50% ultimate stress (ε_{50}), backwall-soil friction angle (δ), and soil failure ratio (R_f), which is the ratio of deviatoric stress at failure to the theoretical ultimate deviatoric stress of soil. F and y are abutment passive resistance and deflection, respectively, and the term $\hat{H}$ is the ratio of the actual height of the abutment backwall to the reference height of 1 meter.

$$F(y) = \frac{\frac{1}{\beta}(\eta - 1)\alpha y}{\frac{H}{H_r} + \frac{1}{\beta}(\eta - 2)y} \tag{3.16}$$

$$\beta = [1703 - 683.4(\tan\varphi)^{1.23}]\varepsilon_{50} \tag{3.17}$$

$$\alpha = \begin{cases} [5.62(\tan\varphi)^2 + 0.53]\gamma + [10.58(\tan\varphi)^{1.79} + 2.86]c, & \varphi \neq 0, \ c \neq 0 \\ [5.62(\tan\varphi)^2 + 0.53]1.06\gamma, & c = 0 \\ 0.50\gamma + 2.63c, & \varphi = 0 \end{cases} \tag{3.18}$$

$$n = \begin{cases} 2, & c = 2 \\ \dfrac{0.91(\tan\varphi)^2 + 1.49}{\sqrt{c}} + 0.9, & c \neq 0 \end{cases} \tag{3.19}$$

$$\eta = \begin{cases} 15.47, & \varphi < 5, \ c \neq 0 \\ 18.10 - 9.38\sqrt{\tan\varphi}, & \varphi \geq 5, \quad c = 0 \\ 16.36 - 7.49\sqrt{\tan\varphi}, & c = 0 \end{cases} \tag{3.20}$$

Four classes of backfill material have been identified for California Highway bridges in a study by Earth Mechanics, Inc. (Earth Mechanics, 2005): (I) Dense to very dense sand with gravel; (II) Medium dense silty sands, some with gravel; (III) Medium dense clayey sands, some with gravel; and (IV) Stiff-hard clays with fine to coarse-grained sands, some with silts (see Table 3.2). In this study, category II is used as the soil type for all the bridges. Phase relationships are used to determine the total unit weight (γ) from the lower bounds of measured field dry density (ρ_d) and moisture content (W). The backwall-backfill interface friction angle, δ, has been set to two-third of the soil internal friction angle, φ, per recommendations of Shamsabadi *et al.* (2005, 2007, 2010). Zero-length springs characterized by nonlinear soil behavior are used to capture the response of the abutment soil. The backfill soil is modeled using the `HyperbolicGapMaterial` material with a Generalized Hyperbolic Force-Deformation (GHFD) backbone (Shamsabadi *et al.* 2005, 2007; Duncan and Mokwa, 2011). It must be noted that in the case of seat type abutment, there exists a gap between the deck and abutment backwall, which in this case is taken as 25.4 mm (1 inch).

Table 3.2 - Backfill soil type categories in California highway bridges

Category	ρ_d (kip/ft³)	W (%)	γ (kip/ft³)	φ (deg)	c (kip/ft²)	ε_{50}	R_f
I	1.2	3-6	0.12	38	0	0.0035	0.96
II	0.12-1.2	5-14	0.11	33	0.5	0.0035	0.96
III	0.1-0.11	7-15	12	23	2	0.0035	0.96
IV	0.11-0.08	14-29	0.09	6	3.5	0.007	0.96

3.2.6 Bearing Pads

Elastomeric bearings are the most commonly adopted bearing types in concrete bridges. These bearings typically transfer horizontal forces by friction and their behavior is characterized by

sliding which in turn depends on the initial stiffness. Once the coefficient of friction is exceeded, the stiffness of the rubber pads drops to zero and therefore, their response can be characterized by an elastic perfectly plastic material. The initial stiffness, k_{pad}, of the bearing pad is calculated using Equation 3.18, where, G is the shear modulus, A is the cross-sectional area, and h is the thickness of the bearing pad.

Figure 3.1b shows the force deformation response of an elastomeric bearing pad. The yield force, F_y, is calculated by multiplying the normal force, N, acting on the bearing with the coefficient of friction, μ, of the pad. Scharge (1981) presented an expression for the coefficient of friction, specific to elastomers on concrete, based on experimental tests, and is a function of the normal stress in MPa, σ_n, as presented in Equation 3.19. The response of the bearing pad is captured using the `Steel01` material provided by Opensees and is applied to a zero-length element to capture its force deformation response.

$$k_{pad} = \frac{GA}{h} \tag{3.18}$$

$$\mu = 0.05 + \frac{0.4}{\sigma_n} \tag{3.19}$$

Table 3.3 - Summary of variables used to describe Opensees springs

Parameter	Bridge A	Bridge B	Bridge C	Bridge F
Concrete, f_{ck}	5 ksi	5 ksi	5 ksi	5 ksi
Steel, f_y	65 ksi	65 ksi	65 ksi	65 ksi
Circular solid pile diameter, d	24 in	24 in	24 in	118.1 in
Number of Column piles, N	25	20	24	1
Young's modulus of soil, E_p	3190 ksi	3190 ksi	3190 ksi	3190 ksi
Reference Young's modulus of the soil at a depth= d, $\widehat{E_s}$	5 ksi	5 ksi	5 ksi	5 ksi
Diameter of Shear Key rebars, d_b	1 in	1.41 in	1.41 in	1 in
Number of Abutment Piles	12	24	34	14
Shear Modulus of Bearing Pads, G	0.108 ksi	0.108 ksi	0.108 ksi	0.108 ksi
Area of Bearing Pads, A	139.47 in^2	280.3 in^2	327.98 in^2	309.76 in^2
Height of Bearing Pads, h	2.56 in	2.46 in	4.52 in	4.92 in

Overall, the longitudinal behavior of the abutment is modeled using five springs in parallel connected by a rigid link while the transverse behavior is modeled using one spring on both ends of the abutment. The model of the longitudinal abutment response comprises of a combination of springs that represent the abutment piles, elastomeric bearing pads, and backfill soil. And the model of the transverse abutment response comprises of a combination of springs that represent the abutment piles, elastomeric bearing pads, and shear key. The arrangement of the transverse and longitudinal springs is shown in Figures 3.1a. Summary of all the variables used to describe the springs of the Opensees models for the four bridges are provided in Table 3.3.

3.3 Site Database
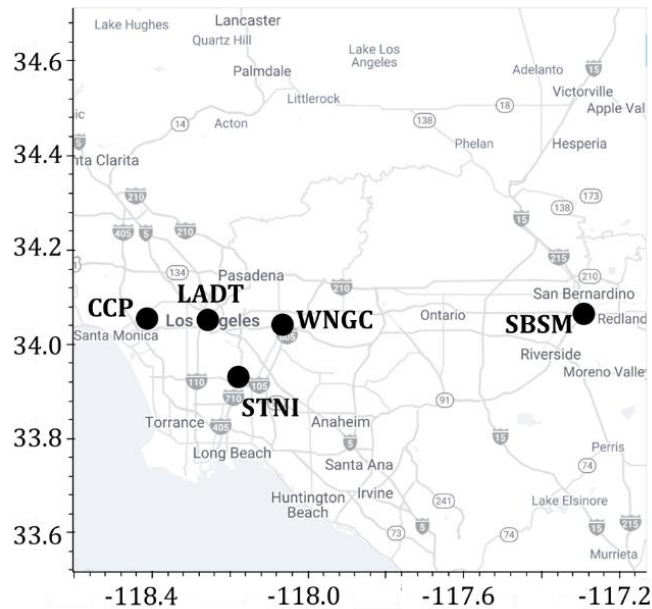
Figure 3.3 - Selected five sites in the greater Southern California region

A subset of 5 representative sites is selected for the research conducted in this study. The selected sites include Los Angeles downtown (LADT), San Bernardino strong motion (SBSM), Seven Ten-Ninety Interchange (STNI), Whittier Narrows Golf Course (WNGC), and Century City Plaza (CCP) sites. The sites LADT and CCP are located within the northern L.A. basin and are selected because of their proximity to a large inventory of buildings and bridge structures of societal importance. STNI is situated on a very deep part of the basin in this region, where the effects of the geologic basin on the resulting ground motions are highly pronounced (Graves *et al.*, 2011). The WNGC and SBSM sites are interesting because they exhibit coupling of basin and directivity effects in the ground motions (Graves *et al.*, 2011). Figure 3.3 shows the locations of the selected

five sites in the Southern California region. Table 3.4 includes the site information including shear-wave velocity averaged over the top-most 30 m of soil (V_{s30}), and the depth where V_{s30} is equal to 2.5 km/sec ($Z_{2.5}$) (aka basin depth) obtained from Lee *et al.* (2014).

Table 3.4 - Characteristics of the selected sites

Site Name	Site Abbreviation	V_{s30} (m/s) (CVM 4.26)	$Z_{2.5}$ (km) (CVM 4.26)	Location	
				Latitude	Longitude
Los Angeles Downtown	LADT	358.6	2.08	34.052	-118.257
San Bernardino Strong Motion	SBSM	354.8	1.77	34.064	-117.292
Century City Plaza	CCP	361.7	2.96	34.054	-118.413
Whittier Narrows Golf Course	WNGC	295.9	2.44	34.041	-118.065
Seven Ten-Ninety Interchange	STNI	268.5	5.57	33.930	-118.179

Chapter Citation:

1) Fayaz J., Dabaghi M., and Zareian F. (2020b). Utilization of Site-Based Simulated Ground Motions for Hazard-Targeted Seismic Demand Estimation: application for Ordinary Bridges in Southern California. *Journal of Bridge Engineering*, Vol. 25, Issue 11.

CHAPTER 4

SEISMIC DEMAND ANALYSIS USING INCREMENTAL DYNAMIC ANALYSIS

4.1 Introduction

Current seismic design procedures such as Caltrans SDC (2013, 2019) are continuously evolving to develop methodologies for conducting seismic analysis to achieve more accurate estimations of the demands corresponding to target hazard levels. In this chapter one of the conventional method of Incremental Dynamic Analysis (IDA) is implemented to obtain Engineering Demand Parameter (EDP) hazard curves. Sets of 20 recorded ground motions are scaled and selected using the hazard curves developed through Campbell and Bozorgnia (2014) (CB14) and the concept of Average S_a (Eads *et al.*, 2015) for 15 hazard levels. The selected ground motion records (20 ground motions ×15 hazard levels = 300 ground motions) are then used to conduct Non-Linear Time-History Analysis of the four bridge structures to obtain the EDP-IM data (IDA curves). The EDP-IM data (IDA curves) is then integrated over the S_a hazard curves obtained from Campbell and Bozorgnia (2014) (CB14) GMPM using Equation 1.2 to obtain the final EDP hazard curves. This is done for each bridge structure for all five sites. Due to asymmetric nature of the dynamic properties of the bridge structures in the two directions (longitudinal and transverse), the EDP is obtained by rotating the two orthogonal components of ground motions in incremental rotations from 0 to 180 degrees (excluding 180º for straight bridges) with 15º increments. Therefore, a total of approximately 78,000 NLTHA (= 4 bridges × 5 sites × 20 ground motions × 15 hazard levels × 13 intercept angles) are conducted in this chapter. The EDP hazard curves obtained in this chapter are used as the basis of comparison for EDP hazard curves for Chapters 5, 6, and 7.

4.2 Selection of recorded ground motions for IDA

The conventional method of Incremental Dynamic Analysis (IDA) uses recorded ground motions for developing Engineering Demand Parameters hazard curves. This is done by first obtaining hazard curves for various spectral periods generated using the Campbell and Bozorgnia (2014) (CB14) empirical Ground Motion Model (GMM) for each site with the ruptures based on

CyberShake-UCERF2, using the OpenSHA software (Field *et al.* 2003). Then the hazard curves are used to develop 15 Uniform Hazard Spectra (UHS) for 15 hazard levels (Return periods of: 5000, 3750, 2500, 1500, 1250, 1000, 900, 700, 500, 300, 200, 100, 75, 50, 25). For each UHS for
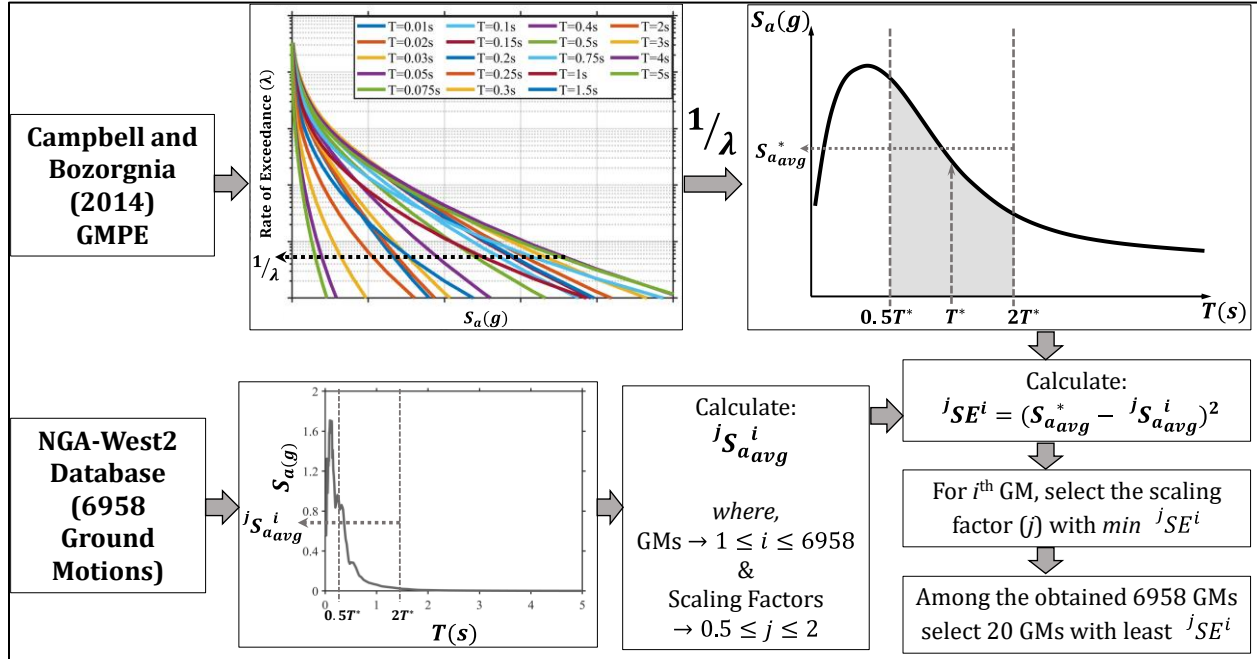
Figure 4.1 - Illustration of selecting hazard-representative recorded ground motions for IDA

a given site, average S_a ($S_{a_{avg}}^*$) (Eads *et al.*, 2015) is calculated between the period range of $0.5T^*$ to $2T^*$, where T^* represents the first mode period of each bridge structure. Then, for each UHS, 20 recorded ground motions that match the $S_{a_{avg}}^*$ with scaling factors of 0.5 to 2 are selected. To select the recorded ground motions, a set of 6958 mainshock recordings available in the NGA-West2 database (Timothy *et al.*, 2014) are used. Each ith recorded ground motion ($1 \leq i \leq 6958$) is scaled with a scaling factor j (in the range $0.5 \leq j \leq 2$) and its average S_a (${}^{j}S_{a_{avg}}^{i}$) is computed.

${}^{j}S_{a_{avg}}^{i}$ is compared against $S_{a_{avg}}^*$ by computing their squared error (${}^{j}SE^i = (S_{a_{avg}}^* - {}^{j}S_{a_{avg}}^{i})^2$). Then, for each ith recorded ground motion, the one with the minimum ${}^{j}SE^i$ is selected, which leads to 6958 scaled ground motions. Among these 6958 scaled ground motions, 20 ground motions with minimum ${}^{j}SE^i$ are selected as a representation of the hazard for the given UHS. This ensures that a ground motion with two different scaling factors is not selected for the same hazard level. Furthermore, once a ground motion is selected for a higher hazard level (higher return period), it is withdrawn from the selection set of 6958 ground motions. In this way, no ground

motion is selected more than once across all scaling factors and hazard levels. This process is explained in Figure 4.1 and is repeated for all 5 sites and for the 4 bridges to select 20 ground motions for each of the 15 hazard levels. Hence, for each site and each bridge, 20×15=300 unique recorded scaled ground motions are selected. It should be noted here that even though the recorded ground motions corresponding to the same hazard level possess similar Average S_a, their $S_a(T^*)$ may still differ from each other. Hence this method is similar to the Cloud method of performing IDA rather than typical strip IDA. The number of ground motions that are classified as pulse-like (Shahi and Baker, 2014) ground motions among the 300 motions for each bridge is given in Table 4.1.

Table 4.1 - Number of pulse-like ground motions among the selected 300 ground motions

Site	Bridge A	Bridge B	Bridge C	Bridge F
LADT	101	111	97	107
SBSM	144	139	131	135
CCP	108	103	97	100
WNGC	126	124	126	137
STNI	126	113	133	140

4.3   Engineering Demand Parameter (EDP) of the bridge structures

The behavior of ordinary bridge structures is mainly deduced by examining the maximum Column Drift Ratio (CDR) of the central bent throughout the time-history of ground motion. Hence, the primary Engineering Demand Parameter (EDP) that bridge engineers rely their design on is the Column Drift Ratio (CDR) as it correlates well with the bridge damage and other EDPs (such as Deck Rotation Angle, Based Shear etc.). Due to the significant differences in the dynamic characteristics of bridge structures in the two orthogonal directions, in this study, the bi-directional components of the ground motions are used to conduct Non-Linear Time-History Analysis (NLTHA) of the bridge structures in incremental rotations from 0º to 180º (excluding 180º for straight bridges) with 15º increments. Hence, to be consistent with the Caltrans practice and at the same time not to over or under- estimate the EDPs, the primary EDP considered in this research is the median value of the maximum CDR obtained after analyzing the bridge structure using the two components of ground motions at the 13 intercept angles. This EDP is termed as $Rot50CDR$ and

is expressed in Equation 4.1. It should be noted that the EDP of Column Drift Ratio (CDR) is termed in the form of *RotppEDP*, where *Rot* indicates the rotation of ground motion components, *pp* indicates the percentile value used for the measure (e.g. "00", "50" and "100" correspond to minimum, median and maximum values, respectively; the median value is used in this study), and *EDP* indicates that the measure is an Engineering Demand Parameter (i.e., Column Drift Ratio *CDR*). The idea of *Rot50CDR* is borrowed from the state-of-practice ground motion Intensity Measure (IM), *RotD50* spectral acceleration (S_a); expressed as *RotDpp*, where *Rot* indicates the rotation of the two orthogonal components of the ground motions, *D* indicates the period dependency and *pp* corresponds to the percentile value (mainly limited to the 50 percentiles, i.e., the median value). Conclusively, *RotD50* is a measure of IM obtained after rotating the two ground motions components on a Single-Degree-of-Freedom (SDOF), *Rot50CDR* is a measure of the EDP (Column Drift Ratio *CDR*) obtained after rotating the two components of ground motions through 180 degrees on the Multi-Degree-of-Freedom (MDOF) bridge models. To avoid any confusion and for the sake of brevity, in this study, the *RotD50* spectral acceleration at bridge's first mode period, which is used as the primary IM of ground motions, is termed as $S_a(T)$. Hence, each ground motion is associated with one value of $S_a(T)$ (as an IM) and one value of *Rot50CDR* (as an EDP).

$$Rot50CDR = median \begin{Bmatrix} CDR_{0°} \\ CDR_{15°} \\ \vdots \\ CDR_{180°} \end{Bmatrix} \tag{4.1}$$

4.4   *IDA-Based* EDP Hazard Curves
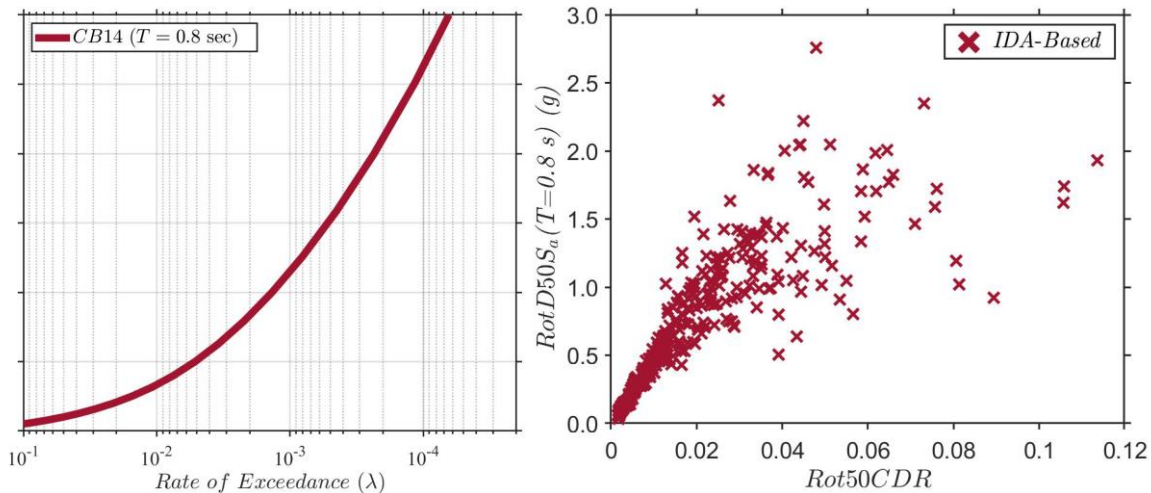
Figure 4.2 - IDA (CB14) data for Bridge B (T*=0.8s) at the LADT site

63
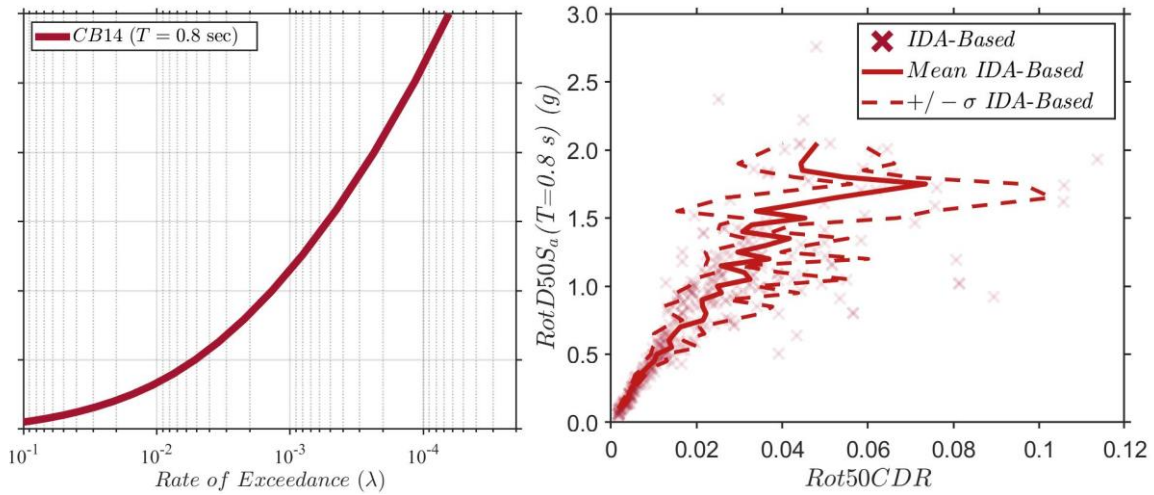
Figure 4.3 - IDA (CB14) curves for Bridge B (T*=0.8s) at the LADT site

300 recorded ground motions are selected for each bridge and site combination and used to conduct NLTHA. Hence, a total of 78,000 NLTHA (= 5 sites × 300 ground motions × 4 bridges × 13 intercept angles) are conducted for IDA, i.e., 3900 NLTHA for each bridge for every site. An example for LADT site for Bridge B is shown in Figure 4.2. The EDP-IM data is used to obtain the IDA curves by computing the mean and standard deviation of the EDP-IM data in lognormal domain. The IDA curves of the EDP-IM data are shown in the Figure 4.3. The IDA curves are then integrated over the CB14 S_a(*T*=0.8 sec) hazard curve using Equation 4.2 to obtain the EDP hazard curve. The obtained *IDA-Based* EDP hazard curves for Bridges A, B, C and F are presented in Figure 4.4. Similar to IM hazard curves, the EDP hazard curves can be directly utilized to obtain EDPs corresponding to a hazard level. Similar EDP hazard curves are obtained for other four sites and shown in Appendix A. In general it is observed that the EDP hazard curves of Bridge F are lower than other three bridges. This can be due to the longer period of the Bridge F as compared to other bridges. For a 1000-year return period the corresponding values of *Rot50CDR* for Bridges A, B, C, and F are computed to be 2.66%, 2.71%, 2.56%, and 1.41%, respectively, for LADT site.

$$\lambda_{EDP} = \sum G(EDP|IM)\Delta\lambda_{IM} \tag{4.2}$$
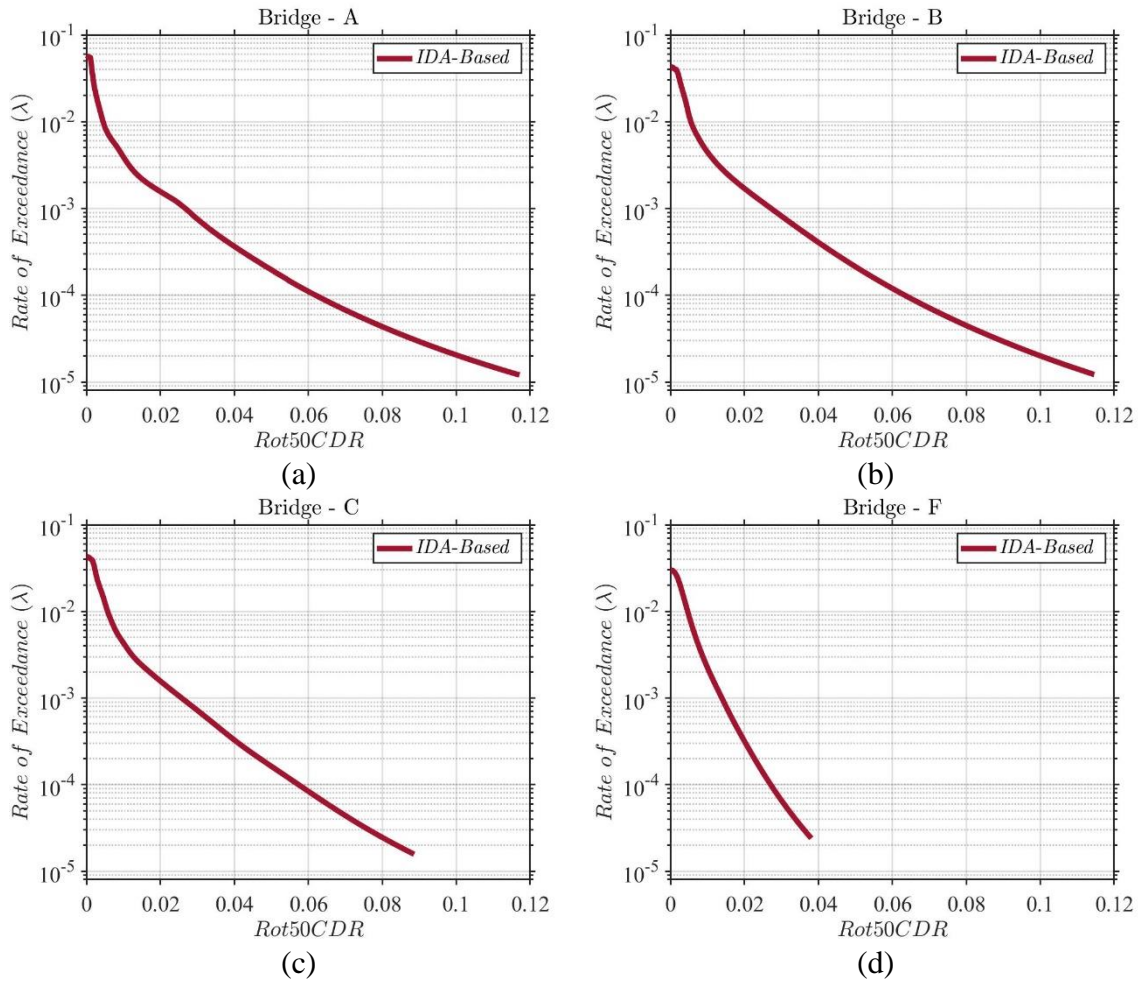
Figure 4.4 – *IDA-Based* hazard curves at LADT site for Bridges: (a) A, (b) B, (c) C, and (d) F

Chapter Citation:

1) Fayaz J., Rezaeian S., and Zareian F. (2021a). Evaluation of simulated ground motions using probabilistic seismic demand analysis: CyberShake (ver. 15.12) simulations for Ordinary Standard Bridges. *Soil Dynamics and Earthquake Engineering.* Volume 141, 106533.

CHAPTER 5

SEISMIC DEMAND ANALYSIS USING SITE-BASED SIMULATED GROUND MOTIONS

5.1 Introduction

This chapter briefly describes the broadband *Site-Based* parameterized stochastic DRD model to generate synthetic ground motions given information about the earthquake source, the site, and the source-to-site geometry. Compared to *Physics-Based* simulation models, *Site-Based* stochastic models are simpler, computationally efficient, and require fewer input parameters and that are easily accessible to engineers. This chapter describes the framework of utilizing the *Site-Based* DRD simulation tool to generate catalogs of synthetic ground motions representing a time-span of 100,000 years for the five sites in Southern California. This leads to around 10,500 ground motions per site. The simulated ground motions are then used to conduct NLTHA of the four Caltrans ordinary standard bridge structures to obtain their *Rot50CDR* EDPs. Therefore, a total of approximately 2,730,000 NLTHA (= 10,500 ground motions × 5 sites × 4 bridges × 13 intercept angles) are conducted in this chapter. The obtained EDPs are then used to develop EDP hazard curves which can then be used to estimate the *Rot50CDR* corresponding to a target hazard-level. The obtained EDP hazard curves are compared against the *IDA-Based* hazard curves and site-specific relations are proposed that correlate the ratio between the *Site-Based* EDPs and conventional *IDA-Based* EDPs with the hazard level, shallow site condition, and site basin depth. Furthermore, the results of the analyses are used along with statistical methods of *Hypothesis T-Test* and *KL-Divergence* to obtain the optimal number of ground motions and ground motion intercept angles that can statistically replicate the results of simulated ground motions that naturally possess the IM of target hazard level. This can assist engineers to utilize this study as a reference to ideally utilize the DRD model for seismic demand analysis.

5.2 DRD Ground Motion Simulation Model

This chapter uses broadband *Site-Based* parameterized stochastic models to generate synthetic ground motions given information about the earthquake source, the site, and the source-to-site

geometry hereafter referred to as Event Parameters. In this study, the far-field model of Rezaeian and Der Kiureghian (2012) is used to simulate ground motions for sites with $R_{RUP} > 30$ km, and the near-fault model of Dabaghi and Der Kiureghian (2018) is used for near-fault sites with $R_{RUP} \leq 30$ km because it accounts for the rupture directivity effect. These two models are combinedly denoted as the DRD model.

Both models employ a modulated and filtered white-noise (MFW) process with time-varying filter parameters proposed by Rezaeian and Der Kiureghian (2008) and can represent the characteristics of recorded ground motions, including temporal and spectral non-stationarity and inherent variability. The near-fault model accounts for the occurrence of the forward rupture directivity effect in the form of a velocity pulse and produces pulse-like and non-pulse-like motions in accordance with their observed proportions among recorded motions. The models are formulated in terms of a relatively small number of physically meaningful model parameters that describe the ground motion amplitude, duration, and frequency content. Both models can generate horizontal orthogonal pairs of synthetic ground motion time series for given Event Parameters. The Event Parameters that are required as input by the DRD model are the type of faulting F (= 0 for strike-slip faults, = 1 for reverse and reverse-oblique faults), the moment magnitude M_w, the closest distance R_{RUP} from the site to the fault rupture plane, and the shear-wave velocity V_{s30} of the top 30 m of soil at the site, in addition to, only for the near-fault model, the depth Z_{TOR} to the top of the rupture plane, and directivity parameters $s_{or}d$ and $\theta_{or}\phi$ (see Dabaghi and Der Kiureghian 2018, for the definition of the directivity parameters). These directivity parameters affect the probability that a ground motion is pulse-like and, if so, the amplitude and period of the pulse.

5.3 Site-Specific Simulations

In this study, site-specific synthetic ground motion catalogs representing a time span of 100,000 years are simulated for the five sites located in Southern California. The sites selected are a subset of the sites considered in the CyberShake ver. *15.12* (Graves *et al.* 2011); their coordinates and soil properties are listed in Table 5.1. The simulated catalogs represent the seismic hazard at each site. The evaluation of the seismic hazard at a site requires a seismic source model, which describes the geometry and magnitude of possible earthquake ruptures in a region of interest and their associated probabilities of occurrence over a specified time. The seismic source model used in this study is

based on the mean Uniform California Earthquake Rupture Forecast, Version 2 (UCERF2) single branch model (Field *et al.* 2009). This version is selected because all the necessary information about the ruptures is available and can easily be extracted from the database of CyberShake study *15.12*. Note that CyberShake study *15.12* doesn't directly used UCERF 2, while performs some modifications and puts additional constraints. These modifications and constraints are thus, also applied in this study. These modifications include setting the minimum magnitude of considered earthquakes to 6, excluding background seismicity, and adjusting rupture areas for consistency with the simulation model (Graves *et al.* 2011). Moreover, for a specific site, the ruptures that are within 200 km of the site are considered in the hazard analysis. The ruptures are assumed to follow independent Poisson distributions with annual probabilities of occurrence provided in CyberShake *15.12*. For each rupture, CyberShake also introduces a suite of variations in the hypocenter location and slip distribution thereby accounting for the natural variability in the rupture characteristics. This process results in an average of 415,000 rupture variations for each site.

Table 5.1 – Details of the sites and their *Site-Based* simulations

Site	LADT	WNGC	SBSM	STNI	CCP
Latitude	34.052	34.042	34.065	33.931	34.055
Longitude	-118.257	-118.065	-117.292	-118.179	-118.413
V_{s30} for UHT (m/s)	360	259	259	259	360
No. of relevant ruptures	7,019	7,076	7,076	7,001	6,939
No. of relevant rupture variations	476,920	478,210	478,210	475,910	475,065
Total No. of GMs in 100,000-year catalogs	10,406	10,663	10,663	10,102	10,046
No. of Pulse-Like GMs in 100,000-year catalogs	296	245	678	164	282

A comprehensive list of all the possible earthquake rupture sources in the region of interest (about 15,000 ruptures) is obtained from the CyberShake platform. For each rupture, the platform provides the fault name, the rupture geometry, the corresponding earthquake magnitude M_w, and the annual probability of occurrence. The CyberShake platform also provides a set of variations in the hypocenter location and the slip-distribution. This process results in a total of more than 800,000 different rupture variations in the region of study. For each site, Table 5.1 presents the number of relevant ruptures (those within 200 km) and rupture variations extracted from CyberShake. The source model is defined in terms of the geometry and magnitude of all the relevant ruptures, their annual probability of

occurrence, and the variations in the hypocenter location. This information is sufficient to obtain all necessary Event Parameters for simulating synthetic ground motions using the DRD model.

Azar *et al.* (2019) described a simulation-based approach for performing probabilistic seismic hazard analysis (PSHA). The same methodology is used here to develop synthetic ground motion catalogs, calculate hazard curves, and obtain deaggregation results for the seven sites of interest. First, the UCERF2 seismic source model and Monte Carlo simulation are used to develop a synthetic catalog of earthquake scenarios (or events) over a period of 100,000 years, by randomly sampling rupture variations according to their annual probability of occurrence. Next, for each of the seven sites and each earthquake scenario, the corresponding Event Parameters F, M_w, Z_{TOR}, R_{RUP}, V_{s30}, $s_{or}d$, and $\theta_{or}\phi$ are obtained. Then the DRD ground motion model is used to generate at each site of interest one synthetic pair of horizontal ground motion time series for each scenario (or event) in the catalog. Because the DRD model was not fitted to earthquakes from normal faults and because the same directivity parameters d and ϕ are used for all non-strike-slip faults, $F = 1$ is used for ground motions from normal and normal-oblique faults, *i.e.*, they are assumed to be similar to ground motions from reverse and reverse-oblique faults. Note that ground motions from normal and normal-oblique faults account for only 1% of the ground motions used in this study, and therefore this assumption is not expected to significantly affect the results. This procedure results in one catalog of 100,000 years at each of the five sites. Each catalog represents a possible realization of the ground motions that may occur at the site over a duration of 100,000 years. Table 5.1 presents for these sites the total number of ground motions in the simulated catalogs and the number of ground motions that are classified as pulse-like using Shahi and Baker (2014).

The ground motions are firstly validated through IM hazard curves and deaggregation results at each site. The simulations are used to compute hazard curves and conduct deaggregation and the results are compared with those obtained using the U.S. Geological Survey (USGS) Unified Hazard Tool (UHT) (2008), which is widely used by engineers for seismic design or assessment studies at a site with known location and V_{s30}. The 2008 dynamic version of the USGS UHT is used in this study because it uses UCERF2 as the seismic source model for California (Petersen *et al.* 2008). The 2008 USGS UHT uses as ground motion model the 2008 NGA GMPMs with equal weights (Petersen *et al.* 2008). Theses GMPE models include terms for basin depth. The IM for which the 2008 USGS UHT hazard curves are calculated is the *RotD50* horizontal component of spectral acceleration. Note that

the 2008 USGS UHT currently provides deaggregation and hazard curve results only for PGA and spectral periods of 0.2, 1, and 2 s, and only for V_{s30} values of 180, 259, 360, 537, 760 and 1150 m/s. For each site, the V_{s30} value nearest the one extracted from CyberShake is used in the UHT and is listed in Table 5.1.
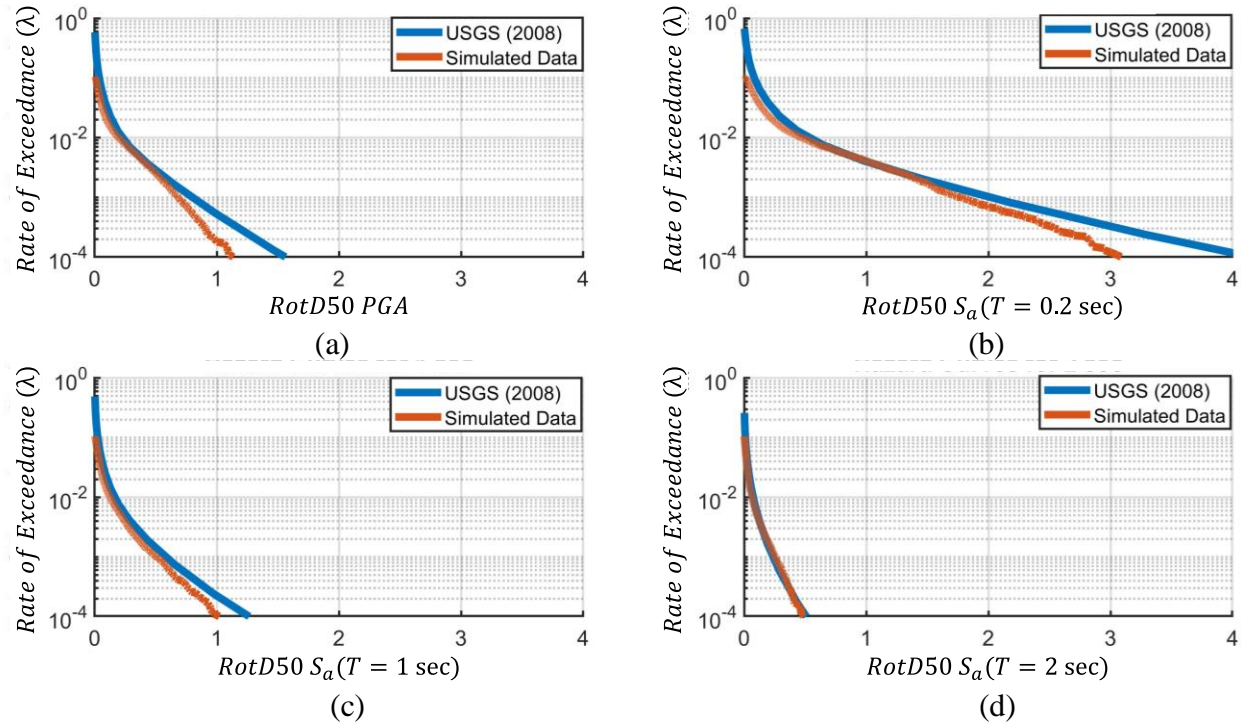
Figure 5.1 – IM hazard curves comparison for (a) PGA level, (b) $S_a(T$=0.2 s), (b) $S_a(T$=1 s), and (d) $S_a(T$=2 s)

Figure 5.1 compares the hazard curves obtained using the simulated 100,000-year catalog at LADT with the hazard curves obtained from the 2008 USGS UHT. Comparisons are made for: (a) PGA level, (b) S_a at $T = 0.2$ s, (b) S_a at $T = 1$ s, and (d) S_a at $T =2$ s. The $RotD50\ S_a(T)$ values obtained from the catalogs of 10,500 simulated ground motions for 100,000 years' time span are used directly to develop *Site-Based* IM hazard curves. This is accomplished by sorting the obtained ~10,500 $RotD50\ S_a(T)$ in a descending for each site and dividing their order number by the time-span of 100,000 years to attain the average annual rate of exceedance of IMs (λ). Figure 5.1a shows that the PGA hazard curves are similar for both methods. However, some differences exist for the $S_a(T)$ hazard curves. At $T = 0.2$ s, the DRD simulation-based approach results in higher mean annual frequency values of $S_a(T = 0.2)$ compared to the USGS UHT; see Figure 5.1b. This is not the case at $T = 1$ s, where the mean annual frequency values of $S_a(T = 1.0)$ from the USGS UHT are higher;

see Figure 5.1c. This difference could be due to the DRD model predicting larger spectral ordinates for $T = 0.2$ s and smaller spectral ordinates at $T = 1$ s compared with the 2008 NGA GMPEs. Deaggregation of the hazard at the high hazard levels are needed to determine the most contributing scenarios at that level and compare the ground motion levels predicted by the DRD model and the GMPEs. At $T = 2$ s, the results from both models are similar; see Figure 5.1d. Figure 5.2 shows the deaggregation into the contributing sources of the $S_a(T)$ hazard curves at $T = 1$ s and for a 5% probability of exceedance in 50 years from both the DRD simulation-based methodology and the USGS UHT. The contributing sources and their ordering are also generally consistent between the two methods; the most hazard contributing sources are Elysian Park (Upper) and Puente Hills (LA).

Any differences in the hazard curves and deaggregation results between the two methodologies can be attributed to several factors. First, although both methods use UCERF2 as their seismic source model, the simulation-based approach excludes background seismicity and ruptures at distances greater than 200 km. Second, and most importantly, differences exist in the ground motion models used by the two methods (2008 NGA GMPEs versus DRD model). For example, two out of the three 2008 NGA GMPEs used in the UHT include basin effects while the DRD model does not; and the DRD model includes the directivity effect while the 2008 NGA GMPEs lack this capability. Despite these differences, the simulated catalogs are deemed representative of the true hazard at the site for the analysis and discussion that follow.

(a)   (b)
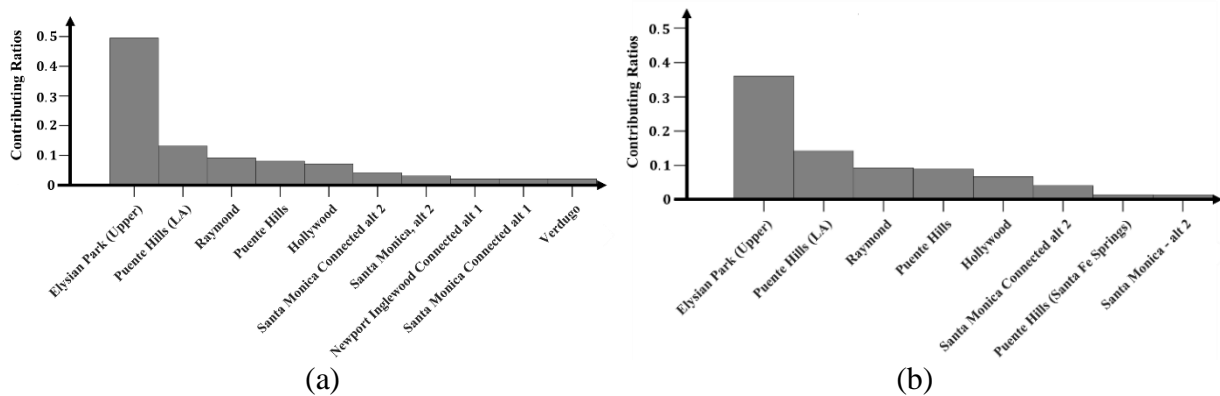
Figure 5.2 - Contributing sources for deaggregation of S_a at 1 s using: (a) DRD Simulations, (b) USGS UHT

5.4 *Site-Based* EDP Hazard Curves
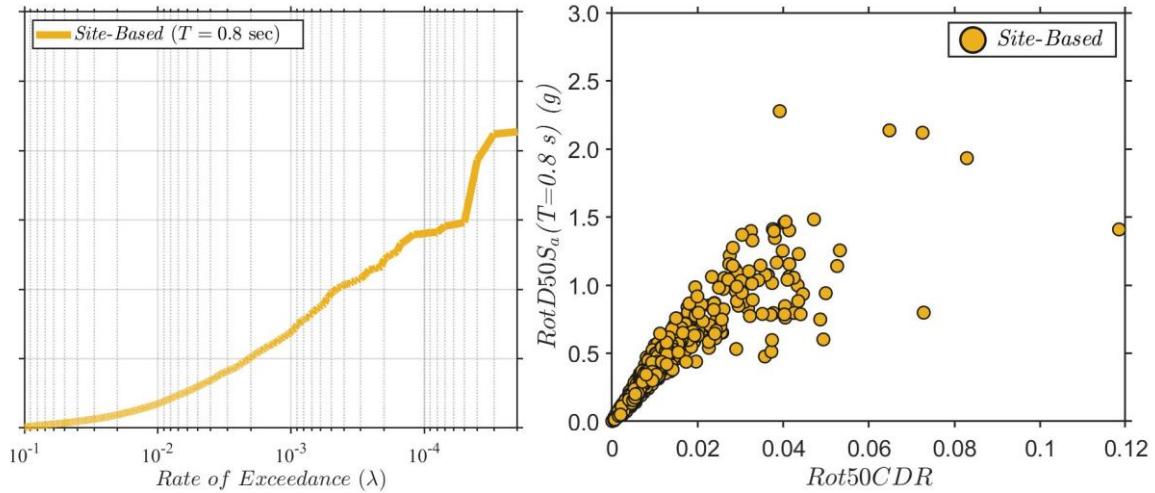
Figure 5.3 – *Site-Based* data for Bridge B (T*=0.8s) at the LADT site

The simulated ~10,5000 ground motions for each site are used to conduct NLTHA of the four bridge structures. Hence, a total of ~2,730,000 NLTHA (= 10,500 ground motions × 5 sites × 4 bridges × 13 intercept angles) are conducted, i.e., 136,500 NLTHA for each bridge for every site. The *Rot50CDR* values obtained from the catalogs of 10,500 simulated ground motions for 100,000 years' time span are used directly to develop *Site-Based* EDP hazard curves. This is accomplished by sorting the obtained ~10,500 *Rot50CDR* in a descending for each site and dividing their order number by the time-span of 100,000 years to attain the average annual rate of exceedance of EDPs (λ). An example for the EDP-IM data and IM hazard curve at LADT site for Bridge B is shown in Figure 5.3. It was observed from Fayaz *et al.* (2020b) that the difference between EDP hazard curves obtained through the procedure of numeric counting and EDP hazard curves obtained from integrating the EDP-IM cloud of the simulated data over the IM hazard curve are highly similar. Hence the bias of integration is ignored and the numerical EDP hazard curves are used for this study. Unlike the conventional IDA method, the simulated ground motion catalogs not only represent the S_a of the 100,000 years but also include the natural variability in the other ground motion characteristics, such as Arias Intensity, duration, frequency content, and other intensity measures. All these ground motion characteristics can lead to variability in the response of the structures. The obtained *Site-Based* EDP hazard curves for Bridges A, B, C and F are presented in Figure 5.4. Similar to IM hazard curves, the EDP hazard curves can be directly utilized to obtain EDPs corresponding to a hazard level. Similar EDP hazard curves are obtained for other four sites

and presented in Appendix A. In general it is observed that the EDP hazard curves of Bridge F are lower than other three bridges. This can be due to the longer period of the Bridge F as compared to other bridges which causes lower demands on the bridge structure leading lower EDPs. For a 1000-year return period the corresponding values of *Rot50CDR* for Bridges A, B, C, and F are computed to be 1.78%, 1.99%, 1.83%, and 1.27%, respectively, for LADT site.
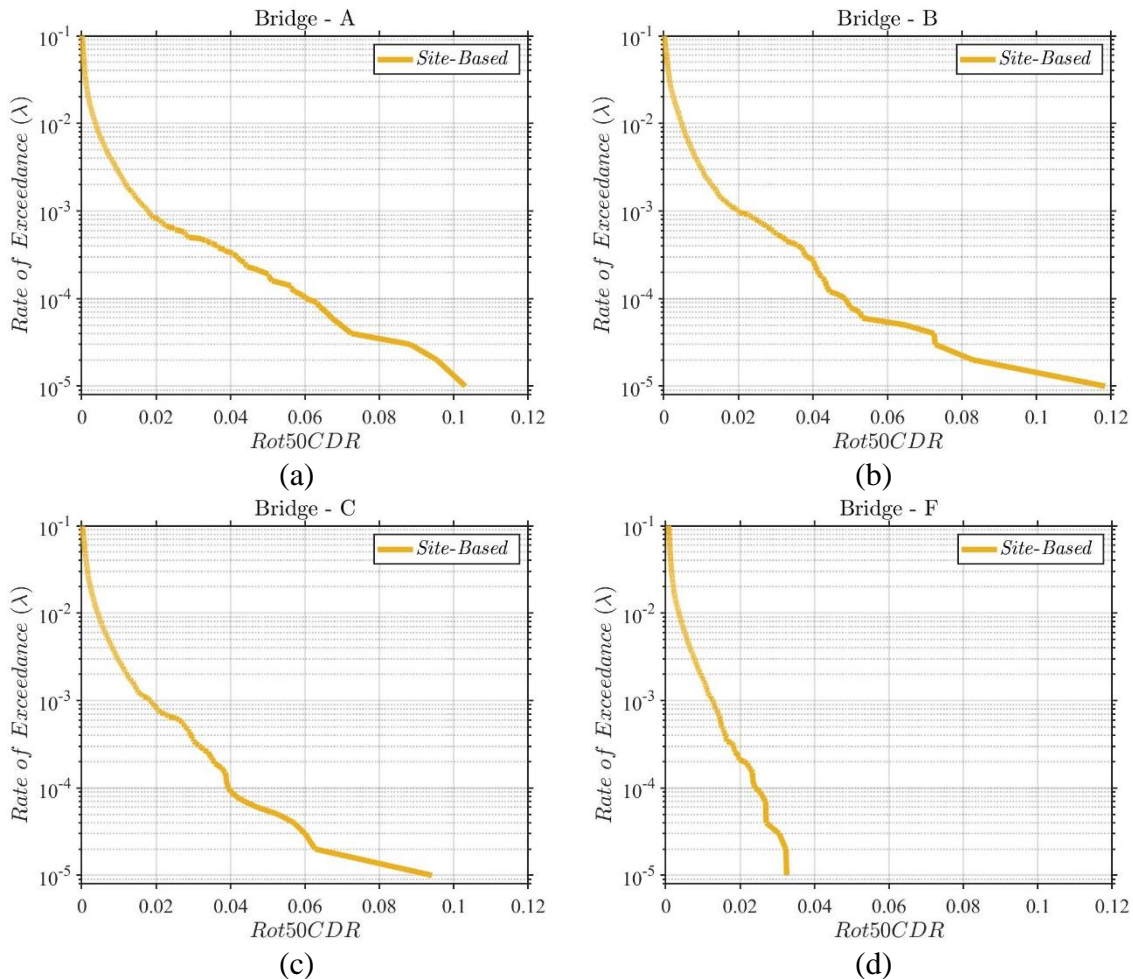
Figure 5.4 – *Site-Based* hazard curves at LADT site for Bridges: (a) A, (b) B, (c) C, and (d) F

5.5 Comparison of *Site-Based* and *IDA-Based* EDP Hazard Curves

The developed *Site-Based* EDP hazard curves are compared against the hazard curves developed using IDA in the Chapter 4, for all four bridges and five sites. An example of the comparison is provided in Figure 5.5 which compares the *Site-Based* EDP hazard curve against the *IDA-Based* EDP hazard curve for the four bridge structures for LADT site. Similar plots are developed for visual comparison of the EDP hazard curves. From these plots it is observed that the EDP hazard
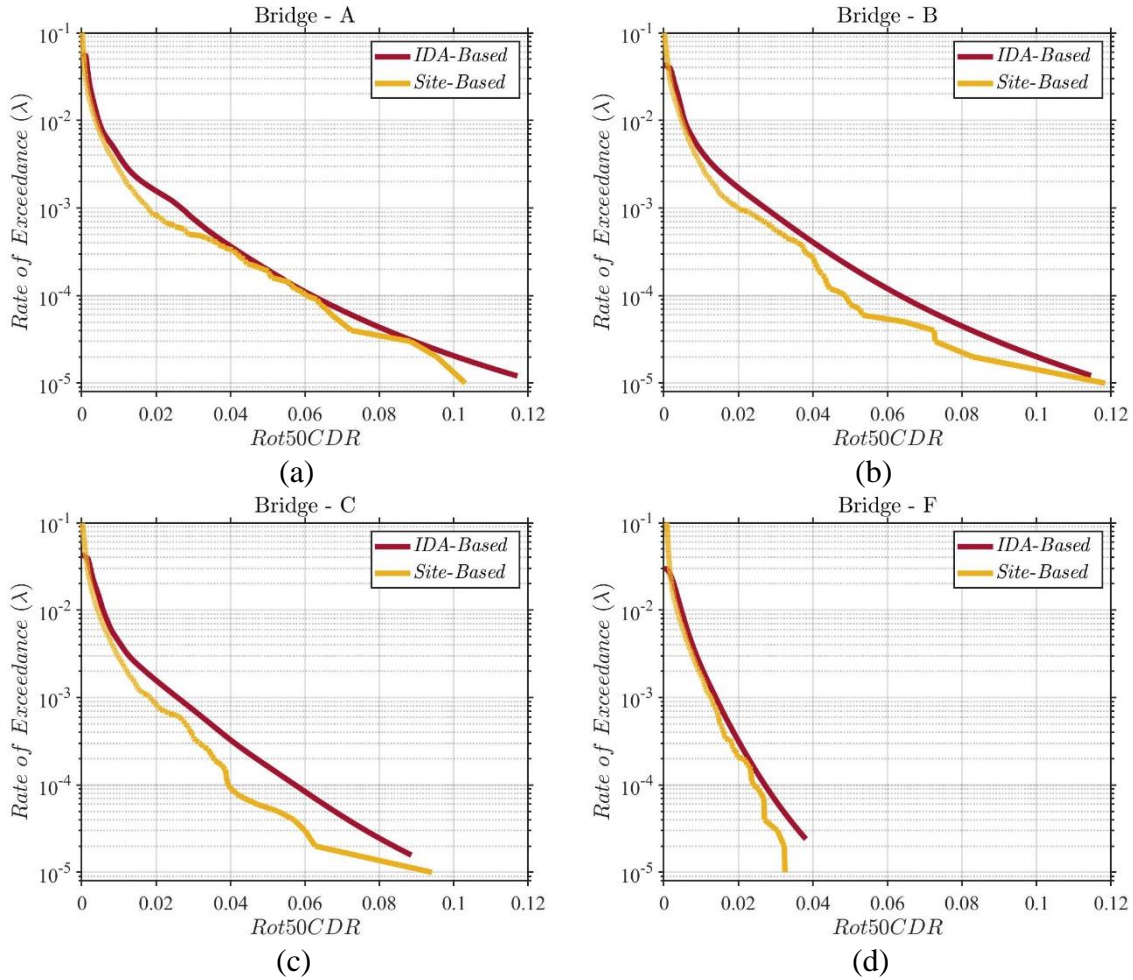
Figure 5.5 – Comparison of *Site-Based* and *IDA-Based* hazard curves at LADT site for Bridges: (a) A, (b) B, (c) C, and (d) F

curves tend to be generally identical for all bridge structures. However, it is observed that in all cases that the *Site-Based* EDP hazard curves tend be lower than the *IDA-Based* EDP hazard curves. This shows that use of ground motion recordings from different regional seismicity can lead to increased estimations in the EDP hazard curves. This is mainly due to the higher variability in the EDPs in EDP-IM data caused by recorded ground motions, especially for larger IM levels. This can be observed by comparing Figures 5.3 and 4.2 especially for ground motions with $S_a > 1.5g$. The integration process accumulates this variability, causing the *IDA-Based* EDP hazards curves to grow faster as compared to *Site-Based* EDP hazard curves. To quantify the differences between the two EDP hazard curves, ratios between the two types of EDPs are computed for various hazard levels for all four bridges and five sites. The EDP ratios are presented in Figure 5.6 which shows that the ratio increases from ~0.5 to ~1 with an increase in hazard levels. The ratios are observed to be closest to ~1 for Bridge F which means that the two types of EDPs are identical. The ratios closest to ~1 are

74

observed for return periods above 2000 years which means that the *Site-Based* hazard curves tend to be similar to *IDA-Based* hazard curves. Apart from the regional seismicity, the differences in the EDP hazard curves can be adhered to the fact that the *Site-Based* ground motions do not only represent the hazard levels in terms of S_a values but also contains the appropriate representation of other IMs. This phenomenon is not present in the scaling and selection method of recorded ground motions for IDA.
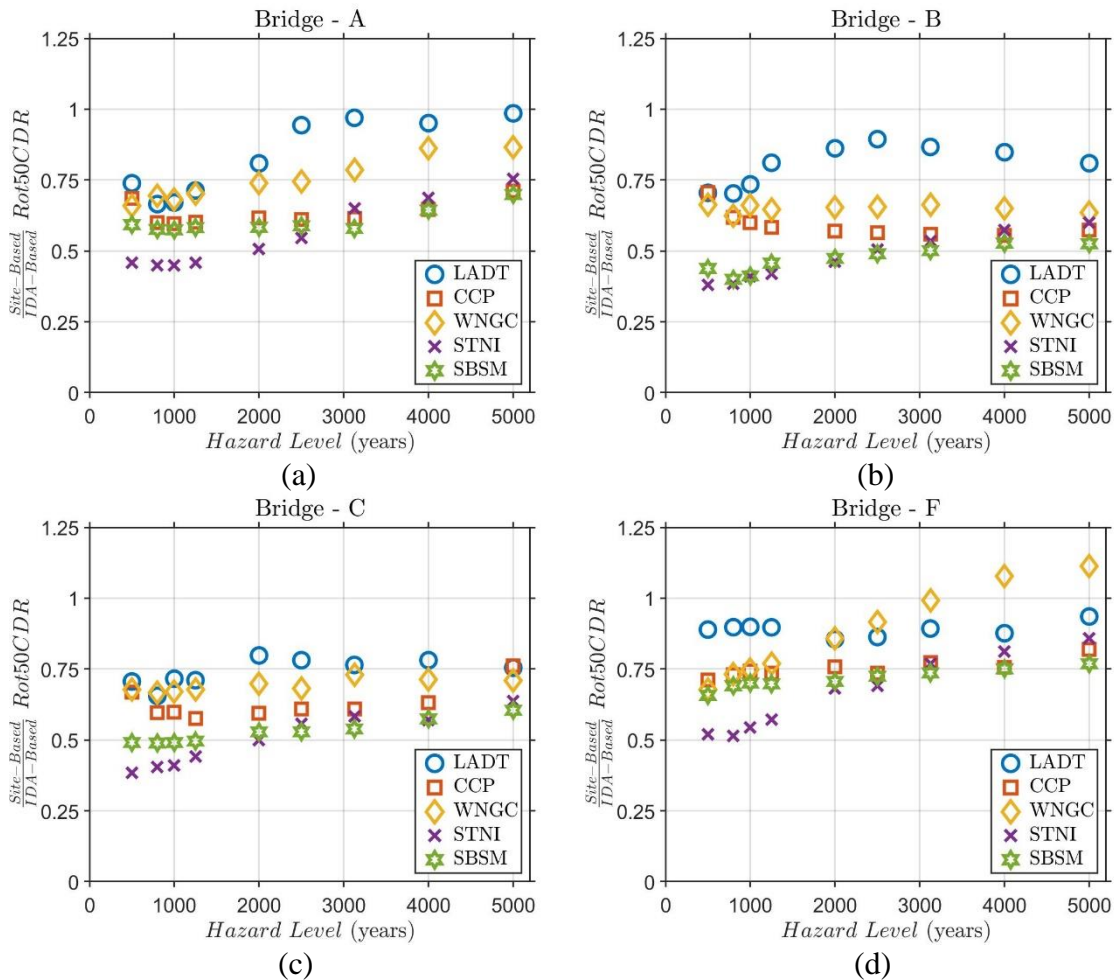
Figure 5.6 – Ratios between *Site-Based* and *IDA-Based* hazard curves for Bridges: (a) A, (b) B, (c) C, and (d) F

To assist with improving ground motion simulation methods using this research (or similar research where probabilistic seismic demand analysis is used for validation of a ground motion simulation method), dependencies of the difference in the response of each type of bridge (A, B, C, and F) on return period and site parameters are provided. This is done by combining the results from all sites and conducting mixed-effects regression analysis with the target variable being the

ratio of *Rot50CDR*s between *Site-Based* EDP hazard curves and *IDA-Based* EDP hazard curves. The predictor variables include Return Period (*RP*), site Shear-Wave Velocity (V_{s30}) and site Basin-Depth ($Z_{2.5}$). This is expressed in Equation 5.1 and is fitted to data independently for each bridge. In Equation 5.1, ε_{ij} represents the within-site variability for the ith hazard level and jth site with zero mean and variance of $\phi_a{}^2$, and σ_i represents the between-site variability for the ith hazard level with zero mean and variance of $\tau_a{}^2$. The coefficients and the goodness-of-fit measure (R^2_{adj}) for each bridge are given in Table 5.2.

$$\ln\left(\frac{\text{SiteBased } Rot50CDR}{\text{IDABased } Rot50CDR}\right) = a_0 + a_1(\ln(RP)) + a_2(\ln(V_{s30})) + a_3(\ln(Z_{2.5})) + \varepsilon_{ij}(0, \phi_a^2) + \sigma_i(0, \tau_a^2) \quad (5.1)$$

It can be observed from Table 5.2 that for all four bridges the value of ϕ_a is higher than τ_a, which means that the site-to-site variability is highly explanatory in the computation of the EDP ratios. Hence, the use of mixed-effects regression is appropriate for describing the EDP ratios. Also, the goodness-of-fit measure R^2_{adj} is observed to be consistently above 0.75, which means that the regression equations can estimate the EDP ratios for the four bridge structures with a high level of accuracy. The coefficient *a₁* is observed to be positive for all bridges demonstrating that as the return period increases, the *Site-Based Rot50CDR* tends increase faster as compared to *IDA-Based Rot50CDR*. The effect of *V*s30 (*i.e.* *a₂*) is observed to be identical between single-column bridges (*i.e.* A and F) and multi-column bridges (*i.e.* B and C). Furthermore, the coefficient *a₃*, which accounts for the effect of basin depth, is negative for all bridges. This goes with the intuition as the *Site-Based* simulated ground motions do not incorporate any basin effects in the simulations. Hence the impact of basin depth is identical to all bridge structures. In general, for the bridge analysis, it is concluded that there is a need for detailed site-specific analysis to understand the fundamental underlying reasons for the differences between these EDP hazard curves. The relations provided in Table 5.2 can assist engineers in the short-term to validate their methods of bridge analysis and can be used in scaling of the EDPs as per the design site and bridge type. It should be noted that the regression equations proposed in this study are based on the DRD simulations of the five southern California sites with soft soils; hence they can be biased towards these conditions. However, these equations can provide initial estimates of the scaling factors and can be easily updated with more data from different site conditions.

Table 5.2 - Fitted coefficients of the *Site-Based* mixed-effects regressions

Bridge	a_0	a_1	a_2	a_3	ϕ_a	τ_a	R^2_{adj}
A	-0.034	0.124	-0.192	-0.268	0.115	0.089	0.79
B	-2.439	0.069	0.264	-0.141	0.185	0.089	0.85
C	-1.468	0.091	0.081	-0.177	0.133	0.074	0.84
F	0.243	0.115	-0.198	-0.224	0.187	0.084	0.73

5.6 Statistical Analysis to Obtain Sample Number of Ground Motions and Intercept Angles for Proper Seismic Demand Estimation

The bi-directional synthetic ground motions simulated for the five sites (and additional two sites PAS and STG) are used to conduct Non-Linear Time-History Analysis (NLTHA) of the four bridge structures. In this section the EDP is denoted as $Rot50CDR^i_m$ (expressed in Equation 5.2) where i denotes the ground motion and m denotes the number of equally spaced intercept angles. For example, $m = 6$ implies that ground motion components are rotated in equally spaced 6 angles from $0°$ to $180°$, which means ground motions are applied at $0°$, $30°$, $60°$, $90°$, $120°$, and $150°$ with respect to the bridge`s longitudinal axis. The site-specific catalog simulations are conducted for $m = 13$ (*i.e.,* intercept increment angle of 10 degrees).

$$Rot50CDR^i_m \; = \; median\left\{CDR^i_{0\times\frac{180°}{m}} \quad CDR^i_{1\times\frac{180°}{m}} \quad \cdots \quad CDR^i_{m\times\frac{180°}{m}}\right\}^T \qquad (5.2)$$
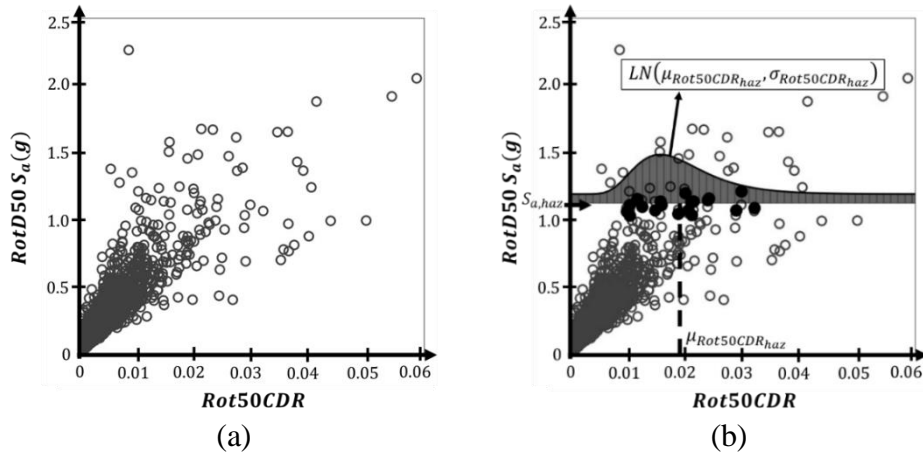

(a)   (b)

Figure 5.7 - (a) *RotD50 S_a* (IM) vs. *Rot50CDR* (EDP) for Bridge A at the CCP site (b) Selected points associated with $S_{a,haz}$

An example of *Rot50CDR* vs $S_a(T)$ for the catalog of simulated ground motions representing 100,000 years' time-span at the CCP site for Bridge A is given in Figure 5.7a. Similar plots for other sites are

achieved by analyzing each bridge under their respective catalogs of simulated ground motions. AS mentioned earlier, the catalogs of the simulated ground motions do not only represent the S_a levels of the 100,000 years but also the other ground motion characteristics such as frequency content, time history evolution, duration, etc. Hence the EDPs obtained from these ground motions can be considered a good representation of the EDP hazard at the site. Since performing this type of simulation in design office can be quite cumbersome and would require substantial computational resources, this section aims to identify sample number (n) of hazard-targeted simulated ground motions which when applied in m uniformly spaced intercept angles, lead to sample EDPs that are statistically equivalent to the EDPs obtained using the whole catalog of site-specific ground motions at the target hazard level IM. The term 'hazard-targeted' in this context means ground motions that naturally possess the $S_a(T)$ of the target hazard level. The $S_a(T)$ associated with the target hazard level at the first mode period T is termed as $S_{a,haz}(T)$ where the hazard level in the case of bridge designs is 5% in 50 years (975 years return period). The *Rot50CDR* EDPs corresponding to the $S_{a,haz}(T)$, obtained after conducting NLTHA using the catalog of ground motions, are termed as *Rot50CDR*_{haz} and are assumed to follow a lognormal distribution with a mean $\mu_{Rot50CDR_{haz}}$ and standard deviation $\sigma_{Rot50CDR_{haz}}$. To obtain the statistics of *Rot50CDR*_{haz}, the *Rot50CDR* values associated with ground motions having $S_a(T)$ in the interval $S_{a,haz}(T) \pm 0.05g$ are used to compute the distribution parameters. An example of the selected values of *Rot50CDR* associated with $S_{a,haz}(T)$ for the CCP site for Bridge A is given in Figure 5.7b (*i.e.*, shown with solid black dots). For each bridge, for all sites, various values of the sample number of ground motions (n) and intercept angles (m) are used in different combinations and, the obtained sample distribution parameters are then tested for statistical equivalency against a lognormal distribution with a mean $\mu_{Rot50CDR_{haz}}$ and standard deviation $\sigma_{Rot50CDR_{haz}}$. The n number of sample *Site-Based* ground motions are simulated using the algorithm described in Fayaz *et al.* (2021b) such that they naturally possess $S_a(T)$ equal to the $S_{a,haz}(T)$, however, differ in their other ground motion characteristics. Finally, the number of hazard-targeted ground motions (n) and intercept angles (m) that satisfy the statistical equivalency for at least five out of seven sites is proposed for the four types of bridge structures.

Six values of n including 7, 9, 11, 13, 15 and 17 are tested along with four values of m which include, 3 (increments of 60º), 4 (increments of 45º), 6 (increments of 30º), and 12 (increments of 15º). For each of the 24 (6 values of n × 4 values of m) combinations of n and m, ten trials of ground motion

simulation and NLTHA are conducted to make the conclusions statistically robust. It is important to note that irrespective of the value of *m*, the EDP associated with a ground motion is always *Rot50CDR, i.e.,* the median value of the *m* CDRs obtained from *m* rotations of the two components of ground motion. This means that for each ith ground motion the $Rot50CDR_m^i$ is obtained using Equation 5.2. This is done for *n* ground motions leading to a vector of $\boldsymbol{Rot50CDR_{n,m}}$ as expressed in Equation 5.3, where the subscript contains *n* and *m* representing the number of ground motions (hence the number of *Rot50CDR* EDPs in the vector) and the number of intercept angles that the *Rot50CDR* EDPs are obtained from.

Furthermore, for each combination of *n* and *m*, ten sets of *n*×1 vectors of $\boldsymbol{Rot50CDR_{n,m}}$ are obtained, and their average statistics are tested against the lognormal population distribution of $Rot50CDR_{haz}$. The test is conducted in two-fold; firstly, a *Hypothesis T-Test* is conducted to test the match between average sample statistics and the population mean $\mu_{Rot50CDR_{haz}}$, and secondly, the whole sample distributions are tested against the population distribution for entropy loss using *Kullback-Leiber* (*KL*) divergence. While the first test only compares the central value of responses of the two distributions, the second compares the probability distributions for information loss.

$$\boldsymbol{Rot50CDR_{n,m}} = \left\{Rot50CDR_m^i\right\} \; where \; \begin{cases} i \in \{1, 2, \dots, n\} \\ n \in \{7, \ 9, 11, 13, 15, 17\} \\ m \in \{3, 4, 6, 12\} \end{cases} \qquad (5.3)$$

5.6.1   Hypothesis T-Test

Most of the simplified statistical tools are mainly based on correctly estimating the central value of the true distribution using the central value of the sample. Hence in this study, the first test is conducted to determine the minimum value of $n \times m$ that on average for ten trials, achieve a sample statistic that is statistically equivalent to the true mean $\mu_{Rot50CDR_{haz}}$. As mentioned above, by taking the median value of *CDR*s due to the *m* intercept angles, the *n* ground motions result in *n* number of *Rot50CDR* EDPs for each trial. Since the aim here is to match only the true mean $\mu_{Rot50CDR_{haz}}$, different types of statistics of the *n* number of sample *Rot50CDR* EDPs are tested to achieve the $\mu_{Rot50CDR_{haz}}$. The statistics that are tested include: *mean*, *median*, 3rd *largest*, *median*+1, and *median*-1 of the *n* number of sample *Rot50CDR* EDPs. For example, if *n* = 9, the 9 values of *Rot50CDR* EDPs are sorted in ascending order; *median* denotes the 5th value, *median*+1 represents the 6th value, *median*-1 represents

the 4th value and *3rd largest* corresponds to the 7th value in the order. The statistics of the *n* number of *Rot50CDR* EDPs arising from *n* ground motions with *m* number of equally spaced intercept angles are generally represented as $Rot50CDR_{n,m}^{SI}$ where *SI* represents the statistical indicator (*mean*, *median*, *3rd largest*, *median*+1, and *median*-1), *n* is the number of ground motions used in obtaining the estimate, and *m* is the number of ground motion intercept angles. This is briefly tabulated in Table 5.3.

Table 5.3 - Description of Statistical Indicators (*SI*) of *Rot50CDR*

$Rot50CDR_{n,m}^{SI}$	Description
$Rot50CDR_{n,m}^{Mean}$	Mean of *n* values *Rot50CDR* (from *m* intercept angles)
$Rot50CDR_{n,m}^{Median}$	Median of *n* values *Rot50CDR* (from *m* intercept angles)
$Rot50CDR_{n,m}^{Median+1}$	Next higher value to the median of *n* values *Rot50CDR* (from *m* intercept angles)
$Rot50CDR_{n,m}^{Median-1}$	Next lower value to the median of *n* values *Rot50CDR* (from *m* intercept angles)
$Rot50CDR_{n,m}^{3rd}$	3rd highest value among the *n* values *Rot50CDR* (from *m* intercept angles)

The mean and standard deviation of the $Rot50CDR_{n,m}^{SI}$ for the 10 trials are computed and termed as $\overline{Rot50CDR_{n,m}^{SI}}$ and $s_{n,m}^{SI}$. The statistical equivalence of $\overline{Rot50CDR_{n,m}^{SI}}$, for each statistical indication (*SI*), to the $\mu_{Rot50CDR_{haz}}$ is tested through *Hypothesis T-Test*. The *T-Test* involves specifying a Null Hypothesis (H_0), which is statistically tested by calculating the *T-score* of the H_0 using the *T-distribution* and comparing its probability against a specified significance level (α). If the probability of a *T-score* (*i.e., p-value*) is less than the specified significance α, then the Null Hypothesis (H_0) is rejected. The Null Hypothesis (H_0) specified in this research is that $\overline{Rot50CDR_{n,m}^{SI}}$ is equal to $\mu_{Rot50CDR_{haz}}$ (*i.e.* $H_0 \sim \overline{Rot50CDR_{n,m}^{SI}} - \mu_{Rot50CDR_{haz}} = 0$). The *T-score* is calculated using Equation 5.4, where *N* represents the number of samples (=10), and to balance the *Type-I* and *Type-II* errors, a α of 0.05 (5%) is used. A schematic of this test is provided in Figure 5.8.

$$t-score = \frac{\overline{Rot50CDR_{n,m}^{SI}} - \mu_{Rot50CDR_{haz}}}{s_{n,m}^{SI}/\sqrt{N}} \tag{5.4}$$
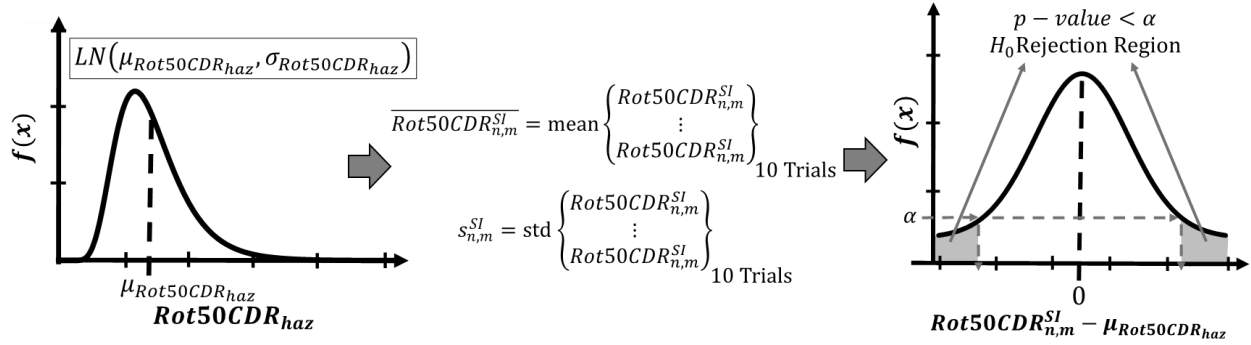
Figure 5.8 - Illustration of *Hypothesis T-Test*

The values of *n* and *m* that satisfy the requirements of the *T-Test* and lead to the least number of simulations (*i.e.*, least $n \times m$) are chosen as the optimal values of *n* and *m*. Based on the results of all simulations it was noticed that *median*+1 statistic deemed suitable for all bridges to represent their $\mu_{Rot50CDR_{haz}}$ for each site. Furthermore, for the two-spanned bridges (*i.e.* Bridge A and Bridge B), $n = 9$ and $m = 6$ (which means intercept angles from 0º to 150º with 30º increment) are selected as the optimal values, while for Bridge C, $n = 11$ and $m = 6$ (which means intercept angles from 0º to 150º with 30º increment) are sufficient to statistically represent the $\mu_{Rot50CDR_{haz}}$. Furthermore, for Bridge F, $n = 11$ and $m = 13$ (which means intercept angles from 0º to 180º with 15º increment). Since Bridge F is curved, the response of 180º is not equal to the response of 0º; hence 180º is included. An example of the *T-Test* for the CCP site for Bridge A for the proposed statistic of *median+1* is provided in Figure 5.9. Figure 5.9a shows the variation *p-value* with respect to the changing values of *n* and *m*. As can be seen from the figure, $n = 9$ and $m = 6$, which means $n \times m = 54$ simulations are the least number of simulations required to obtain a statistically equivalent EDP to the $\mu_{Rot50CDR_{haz}}$. Figure 5.9b further portrays the variation of the values of *Rot50CDR* for the 9 ground motions for the 10 trials of simulations with respect to $\mu_{Rot50CDR_{haz}}$. The dark filled circles in the plot show the *Rot50CDR* for the selected statistic of *median+1*. As can be seen from the figure, the *median*+1 (6th largest among 9) *Rot50CDR* for the 10 trials lies quite close to the $\mu_{Rot50CDR_{haz}}$ which was statistically confirmed by the hypothesis test. While the results are presented only for the CCP site, the results for other sites were observed to be similar, and the final proposed values of the number of hazard-targeted ground motions (*n*) and intercept angles (*m*) along with the corresponding intercept angle increments are tabulated in Table 5.4.
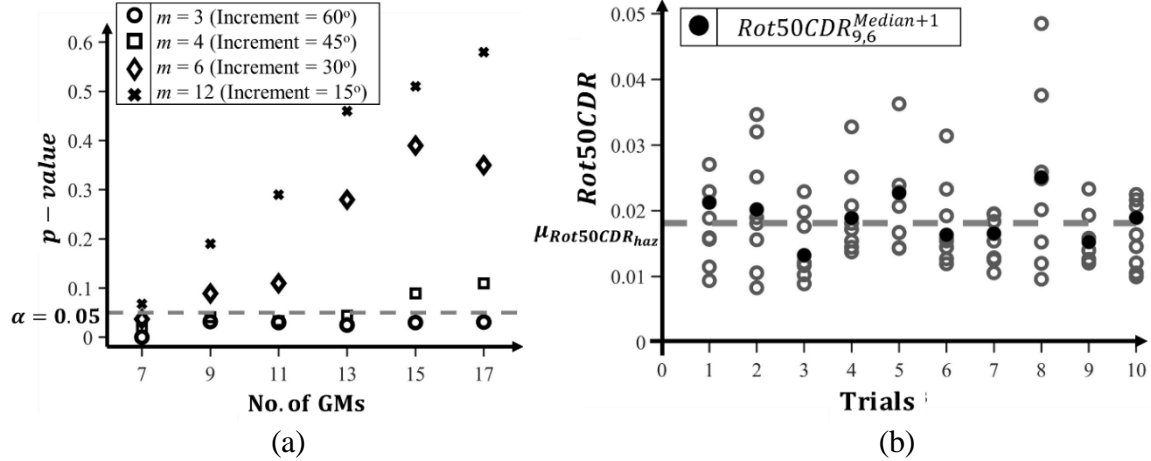
Figure 5.9 - (a) Results of hypothesis tests for all $m \times n$ for Bridge A at CCP site, (b) $Rot50CDR$s for $n = 9$ GMs and $m = 6$ angles (30° increment) for ten trials

Table 5.4 - Proposed number of hazard-targeted simulations to statistically estimate $\mu_{Rot50CDR_{haz}}$

Bridge	Number of Ground Motions (n)	Number of Equally Spaced Intercept Angles (m)	Intercept Angle Increment
A	9	6	30° (0° to 150°)
B	9	6	30° (0° to 150°)
C	11	6	30° (0° to 150°)
F	11	13	15° (0° to 180°)

5.6.2 Kullback-Leibler (KL) Divergence Test

Apart from the central values of EDP distribution, other statistical measures of the distribution may have value for engineers and researchers. One application of such understanding is on the development of Load and Resistance Factor Design (LRFD) type equations for proportioning structural components (Fayaz and Zareian 2019). This section shows how to identify n and m that lead to a sample distribution of $Rot50CDR$ that match the population distribution of $Rot50CDR_{haz}$; *Kullback-Leibler* (*KL*) *Divergence* is used to measure such equivalency. *Kullback-Leibler* (*KL*) *Divergence* is a measure to determine the information entropy loss between two distributions. Greater randomness implies higher entropy, and greater predictability implies lower entropy. *KL Divergence* (D_{KL}) measures the similarity between two probability distribution by aiming to identify the divergence of a probability distribution given a baseline distribution. That is, for a target distribution, P, we compare a competing distribution, Q, by computing the expected value of the log-odds of the two distributions using Equation 5.5. The smallest possible value for D_{KL} is zero (which means that the distributions are equal), and the maximum value is infinity. We obtain infinity when

P is defined in a region where Q can never exist. Therefore, it is common to assume both distributions exist on the same support. The target distribution P, in this study, is the lognormal distribution of $Rot50CDR_{haz}$ with population mean $\mu_{Rot50CDR_{haz}}$ and deviation $\sigma_{Rot50CDR_{haz}}$.

$$D_{KL}(P||Q) = \sum_i P(i)log\frac{P(i)}{Q(i)}$$

(5.5)

Table 5.5 - Number of simulations corresponding to their p and q

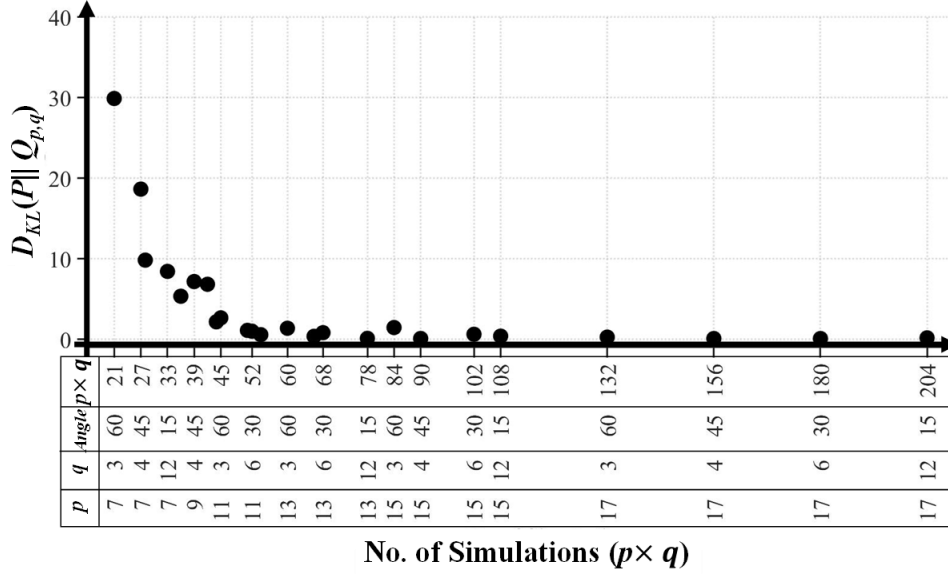| Number of Ground Motions (p)	Number of Equally Spaced Intercept Angles (q)	Intercept Angle Increment	Number of Simulations ($=p \times q$)
7	3	60°	21
	4	45°	28
	6	30°	42
	12	15°	84
9	3	60°	27
	4	45°	36
	6	30°	54
	12	15°	108
11	3	60°	33
	4	45°	44
	6	30°	66
	12	15°	132
13	3	60°	39
	4	45°	52
	6	30°	78
	12	15°	156
15	3	60°	45
	4	45°	60
	6	30°	90
	12	15°	180
17	3	60°	51
	4	45°	68
	6	30°	102
	12	15°	204

Figure 5.10 - Results of $D_{KL}(P// Q_{p,q})$ for Bridge A at CCP site

In this section, an attempt is made to obtain an optimal lognormal sample distribution $Q_{p,q}$ with mean using p number of ground motions and q equally spaced intercept angles from $0°$ to $180°$. The p number of *Rot50CDR* EDPs obtained from conducting NLTHA using the sample p number ground motions with q number of equally spaced intercept angles, are assumed to follow a lognormal distribution with sample mean $\overline{Rot50CDR}_{p,q}$, and sample standard deviation $S_{p,q}$. Similar to the previous section, 24 combinations of p and q are used to compute sample distribution values. This is repeated for ten trials which leads to ten values of $\overline{Rot50CDR}_{p,q}$ and $S_{p,q}$ for each combination of p and q. The average of these ten values of $\overline{Rot50CDR}_{p,q}$ and $S_{p,q}$ are calculated and termed as $\overline{Rot50CDR}_{p,q_{avg}}$ and $S_{p,q_{avg}}$. For all 24 combinations of p and q, $\overline{Rot50CDR}_{p,q_{avg}}$ and $S_{p,q_{avg}}$ are computed and then used to compute lognormal sample distribution $Q_{p,q}$, which is compared against the true distribution P by calculating $D_{KL}(P// Q_{p,q})$ using Equation 5.6. The two distributions are compared between ±3 standard deviations (covering 99% of distributions). The results of this are shown in Figure 5.10, where $D_{KL}(P//Q_{p,q})$ is plotted against the number of simulations (*i.e.*, $p \times q$) that represent 24 combinations of p and q for Bridge A for the CCP site. Table 5.5 relates the number of simulations with their corresponding p and q. The $Q_{p,q}$ after which the rate of decrease in $D_{KL}(P//Q_{p,q})$ is deemed low for all the sites, that $Q_{p,q}$ is selected as the optimal sample distribution with the p number of ground motions and q equally spaced intercept angles. This means that the values of p and q which lead to minimal information loss while requiring the least number of simulations are selected as the optimal number of ground motions and equally spaced intercept angles to estimate the population distribution P. For

Bridge A, simulation of $p = 13$ and $q = 6$ (intercept angle increment $= 30°$) is selected as optimal sample distribution $Q_{p,q}$ to represent the population distribution P for the site. The comparison of the selected $Q_{p,q}$ against the population distribution P is shown in Figure 5.11. Similar plots were generated for all bridges and all sites, and the proposed values of p and q that satisfy the statistical equivalency for at least five sites are presented in Table 5.6.

Table 5.6 - Proposed number of hazard-targeted simulations to statistically estimate
$$P \sim LN(\mu_{Rot50CDR_{haz}}, \sigma_{Rot50CDR_{haz}})$$

Bridge	Number of Ground Motions (p)	Number of Equally Spaced Intercept Angles (q)	Intercept Angle Increment
A	13	6	30° (0° to 150°)
B	13	6	30° (0° to 150°)
C	15	6	30° (0° to 150°)
F	17	7	30° (0° to 180°)
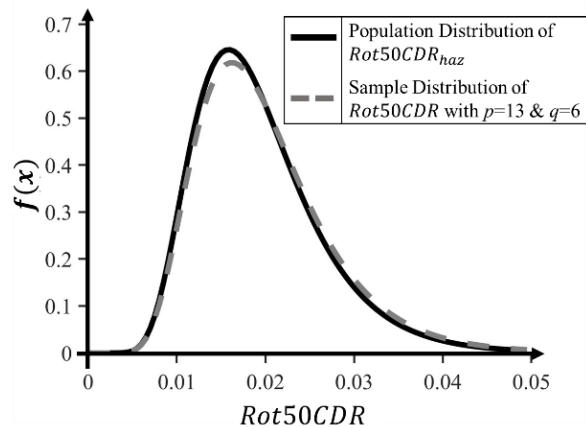
Figure 5.11 - Comparison of population distribution (P) of $Rot50CDR_{haz}$ vs. sample distribution ($Q_{p,q}$) of $Rot50CDR$ with $p=$ 13 GMs and $q=$ 6 angles (30° increment angle)

5.7 Conclusions

This chapter uses the DRD *Site-Based* ground motion simulation tool to conduct NLTHA of the four ordinary bridge structures for the five sites in southern California. Catalogs of around 10,500 ground motions, corresponding to 100,000 years' time-span, are simulated for each site based on the UCERF2 rupture forecast model. The simulated ground motions not only represent the distribution of a single IM for the 100,000 years but also represent the variability in and the correlation between other ground motion characteristics, such as duration, frequency content, and other intensity measures.

This chapter provides bridge design engineers with framework of using *Site-Based* simulated ground motions for obtaining the EDP hazard curves. The *Site-Based* EDP hazard curves are developed for the four bridge structures and for the five sites. The developed EDP hazard curves are then compared against the *IDA-Based* EDP hazard curves obtained in Chapter 4 of this report. The EDP hazard curves from the two frameworks tend be quite identical and lead to similar *Rot50CDR* values. The differences between the two types of EDP hazard curves are computed in terms of ratio between the *Rot50CDR*s at various hazard levels for the five sites for each bridge structure. The EDP ratios are then used to develop a mixed-effects model that relate the EDP ratio with the return period, site shear-wave velocity and site basin depth. The mixed effects models are developed independently for each bridge structure with the random intercept denoting the variability between sites. These ratios can be utilized by the engineers to convert *IDA-Based* EDPs to *Site-Based* EDPs. Given that the study is mainly focused for engineering practice, performance is associated with EDPs; damage and loss assessment and performance evaluation are out of the scope of this study. Some other limitations of this study include that the ground motions are only simulated for a time span of 100,000 years, basin effects and soil-structure interaction effects are ignored, and only five sites and four bridges are considered. However, the selected sites and bridges are considered to be good representatives of the state-of-practice. Also note that some differences are observed in the ground motion levels predicted by the DRD model and the 2008 NGA GMPEs for periods around 1 s. These differences may affect the resulting IM-hazard curves at periods around 1 s but are not expected to significantly affect the EDP-IM relations and the results of this study. Moreover, engineers will not use simulated motions to calculate IM hazard curves but will use hazard studies commonly adopted for their site or region of interest (such as the USGS UHT) to obtain the IM of the design hazard level and the corresponding hazard-targeted simulated ground motions. Therefore, the differences in ground motion models are not expected to affect their results significantly.

Using the EDP-IM data and IM-hazard curves obtained from the simulated catalogs, two types of statistical tests are conducted to arrive at the optimal number of hazard-targeted ground motions and equally spaced intercept angles that are required to match the EDPs corresponding to the IM of the target hazard level. *Hypothesis T-Test* is conducted to match only the point estimates of the expected value of EDPs, while *KL Divergence* is used to match the distribution of EDPs at the target IM hazard level. It is derived that by using statistical indicators of *median*+1, Bridges A, B, C, and F require 9, 9, 11, and 11 number of ground motions with rotation increments of 30º, 30º, 30º, and 15º, respectively.

To match the whole distribution, using *KL Divergence*, it is concluded that Bridges A, B, C, and F require 13, 13, 15, and 17 number of ground motions, all with rotation increments of 30°. To properly estimate the seismic demands of ordinary bridge structures, the engineers can use Table 5.4 and Table 5.6 for determining the number of analyses to conduct for the respective type of ordinary bridge structure.

Chapter Citation:

1) Fayaz J., Dabaghi M., and Zareian F. (2020b). Utilization of Site-Based Simulated Ground Motions for Hazard-Targeted Seismic Demand Estimation: application for Ordinary Bridges in Southern California. *Journal of Bridge Engineering*, Vol. 25, Issue 11.

2) 43) Fayaz J., Azar S., Dabaghi M., and Zareian F. (2021b). An Efficient Algorithm to Simulate Hazard-Targeted Site-Based Synthetic Ground Motions. Earthquake Spectra. Vol 37, Issue 2.

3) Fayaz J., and Zareian F. (2021c). An Efficient Algorithm to Simulate Site-Based Ground Motions that match a Target Spectrum. *Earthquake Engineering and Structural Dynamics*. (2nd round of review).

CHAPTER 6

SEISMIC DEMAND ANALYSIS USING PHYSICS-BASED SIMULATED GROUND MOTIONS

6.1 Introduction

One of the primary reasons behind development of simulated ground motions is to reach a technological setting where seismic loading, including simulations, can be tailored for the engineering of a specific structure and its surrounding environment. In this chapter *Physics-Based* ground motions simulated for the CyberShake (ver. *15.12*) study are utilized for developing EDP hazard curve. Catalogs of simulated ground motions representing a 200,000-year history are selected from the Southern California Earthquake Center (SCEC) CyberShake (ver. *15.12*) database for the five sites in Southern California (~20,500 unscaled ground motions per site). The selected ground motions are then used for Non-Linear Time History Analysis (NLTHA) of the four Ordinary Standard Bridge structures to obtain their *Rot50CDR* EDPs. Therefore, a total of approximately 5,330,000 NLTHA (= 20,500 ground motions × 5 sites × 4 bridges × 13 intercept angles) are conducted in this chapter. For each site and bridge structure, this data is used to obtain *Physics-Based Rot50CDR* hazard curves. These are compared against EDP hazard curves that are developed using conventional IDA method in Chapter 4. The two EDP hazard curves are compared at various return periods in terms of EDP ratios between the two types of EDPs (*Physics-Based* and *IDA-Based*) and site-specific relations are proposed that correlate the ratios with the hazard level, site condition, and site basin depth. It is observed that CyberShake (ver. *15.12*) yields similar EDP values to empirical data for shorter return periods. For longer return periods, however, EDPs from *Physics-Based* analysis tend to be lower than the EDPs obtained from *IDA-Based* analysis for short-period bridges while the case is opposite for long-period bridges.

6.2 CyberShake Simulations

This chapter uses the ground motions simulated for the CyberShake *15.12* (Graves *et al.*, 2011). CyberShake is a computational platform developed by the Southern California Earthquake Center (SCEC) that is based on integration of a collection of scientific software and middleware that

performs 3D *Physics-Based* simulations. For CyberShake *15.12* study, deterministic *Physics-Based* (Source-Based) models are utilized in simulations for frequencies up to 1 Hz; the results are then augmented with high-frequency (1-10 Hz) stochastically simulated seismograms produced using the Graves and Pitarka (2015) module from the SCEC Broadband platform. The study simulated ground motions for 337 sites on a closely-spaced grid in the Southern California region. Among these, a subset of 5 representative sites (explained in Chapter 3) is selected for this study. The Earthquake Rupture Forecast used in CyberShake *15.12* study is based on the UCERF2 single branch model (Field *et al*. 2009). However, it does not use UCERF 2 directly and performs some modifications and incorporates additional constraints. These modifications include setting the minimum magnitude of considered earthquakes to 6, excluding background seismicity, and adjusting rupture areas for consistency with the simulation model (Graves *et al*. 2011). For each site, CyberShake-UCERF2 provides a list of potential ruptures with their annual probabilities of occurrences and also introduces a suite of variations in the hypocenter location and slip distribution to account for the natural variability in rupture characteristics. This process results in an average of 415,000 rupture variations for each site. Each rupture variation is associated with a ground motion waveform simulated for CyberShake *15.12* study. The ruptures are assumed to follow independent Poisson distribution and all the hypocentral variations are assumed to be equally likely for each rupture.

Table 6.1 – Details of the sites and their *Physics-Based* simulations

Site	LADT	WNGC	SBSM	STNI	CCP
Latitude	34.052	34.042	34.065	33.931	34.055
Longitude	-118.257	-118.065	-117.292	-118.179	-118.413
No. of relevant ruptures	7,019	7,076	7,076	7,001	6,939
No. of relevant rupture variations	476,920	478,210	478,210	475,910	475,065
Total No. of GMs in 200,000-year catalogs	20,984	21,359	22,848	20,415	19,822
No. of Pulse-Like GMs in 200,000-year catalogs	783	1167	1721	1014	965

For the selected subset of five sites, Monte Carlo simulations are used to obtain a catalog of ground motion simulations representing a 200,000-year history within 200 km of the site. This is done by randomly sampling rupture variations for each site according to their annual probability of

occurrences and then obtaining their corresponding ground motions. The obtained catalog represents a realization of ~20,500 ground motions that may occur at the site over 200,000 years. Table 6.1 includes the site information and the number of events (or ground motions, GMs) in the simulated catalog for each site. Table 6.1 also provides information about the number of ground motions classified as pulse-like among the simulated catalogs using the algorithm proposed by Shahi and Baker (2014).

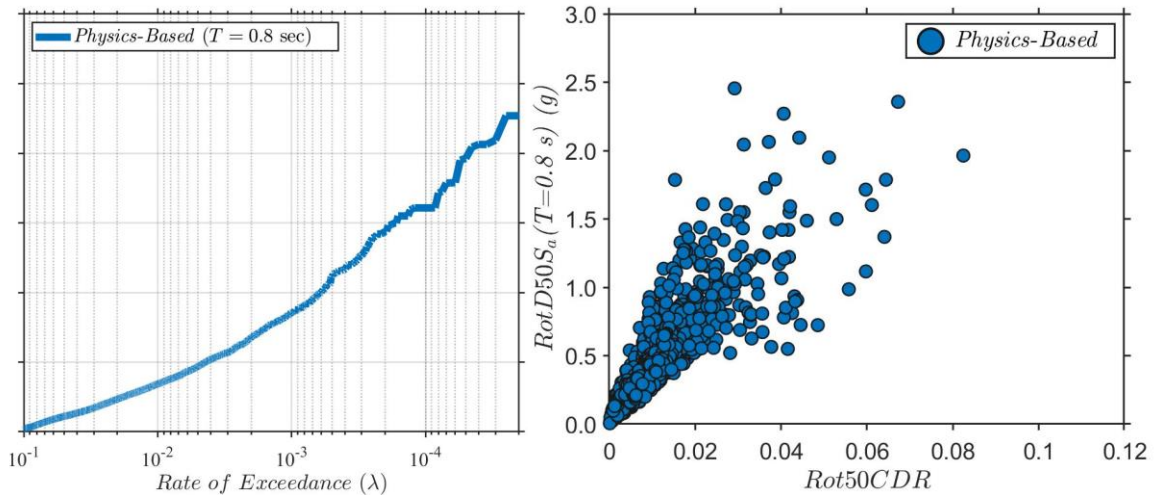## 6.3 *Physics-Based* **EDP Hazard Curves**

Figure 6.1 – *Physics-Based* data for Bridge B (T*=0.8s) at the LADT site

The simulated ground motions for two orthogonal horizontal components for each site are used to perform NLTHA of the four bridge structures to obtain their *Rot50CDR* EDPs. Hence this chapter is based on the results of 5 sites × 20,500 ground motions × 4 bridges × 13 intercept angles = 5,330,000 NLTHA (266,500 NLTHA for each of the four bridges for every site representing 200,000 years). The *Rot50CDR* and $RotD50\ S_a(T)$ values obtained from the catalogs of simulated ground motions are used directly to develop EDP and IM hazard curves, respectively. This is accomplished by sorting the obtained ~20,500 *Rot50CDR* and $S_a(T)$ in a descending for each site and dividing their order number by the time-span of 200,000 years to attain the average annual rate of exceedance (λ) of EDPs and IMs. An example of the IM hazard curves and the EDP-IM data using *Physics-Based* simulations of CyberShake is presented for Bridge B for the LADT site in Figure 6.1. It was observed from Fayaz *et al.* (2021a) that the difference between EDP hazard

curves obtained through the procedure of numeric counting and EDP hazard curves obtained from integrating the EDP-IM cloud of the simulated data over the IM hazard curve are highly similar.
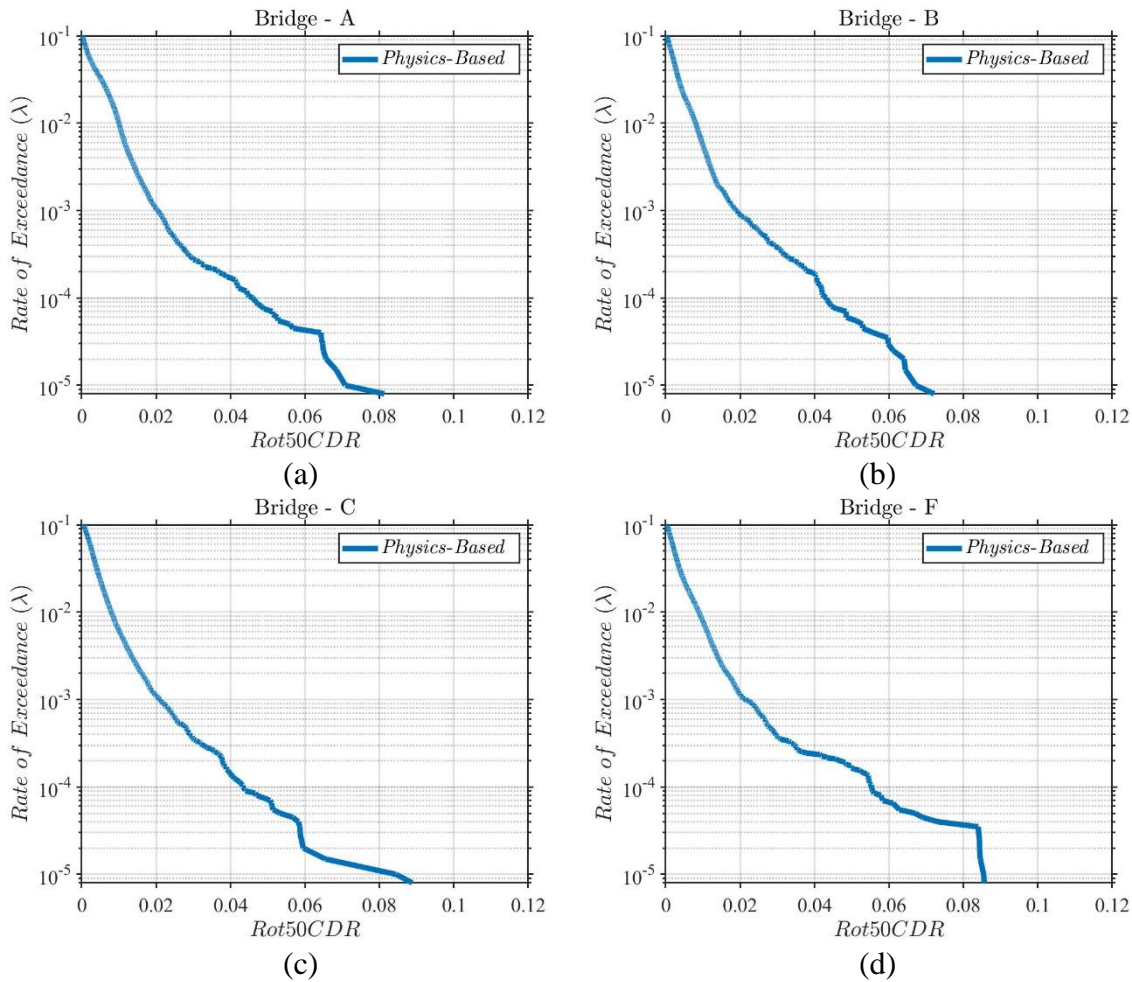

Figure 6.2 – *Physics-Based* hazard curves at LADT site for Bridges: (a) A, (b) B, (c) C, and (d) F

Hence the bias of integration is ignored and the numerical EDP hazard curves are used for this study. Unlike the conventional IDA method, the simulated ground motion catalogs not only represent the S_a of the 200,000 years but also include the natural variability in the other ground motion characteristics, such as Arias Intensity, duration, frequency content, and other intensity measures. All these ground motion characteristics can lead to variability in the response of the structures. The obtained *Physics-Based* EDP hazard curves for Bridges A, B, C and F are presented in Figure 6.2. Similar to IM hazard curves, the EDP hazard curves can be directly utilized to obtain EDPs corresponding to a hazard level. Similar EDP hazard curves are obtained for other four sites and are presented in Appendix A. In general it is observed that the EDP hazard curves of Bridge

F are higher than other three bridges. This can be due to the longer period of the Bridge F as compared to other bridges and the fact that the period lies at the intersection of the stochastic and deterministic parts of the CyberShake *15.12* simulations. This means that the the validation of the CyberShake ground motions may require further validations for this transition period zone and the validations efforts must go beyond IM level and should be done using EDPs of real structures (such as Fayaz *et al.* 2020c). For a 1000-year return period the corresponding values of *Rot50CDR* for Bridges A, B, C, and F are computed to be 2.04%, 1.92%, 2.09%, and 2.14%, respectively, for LADT site.

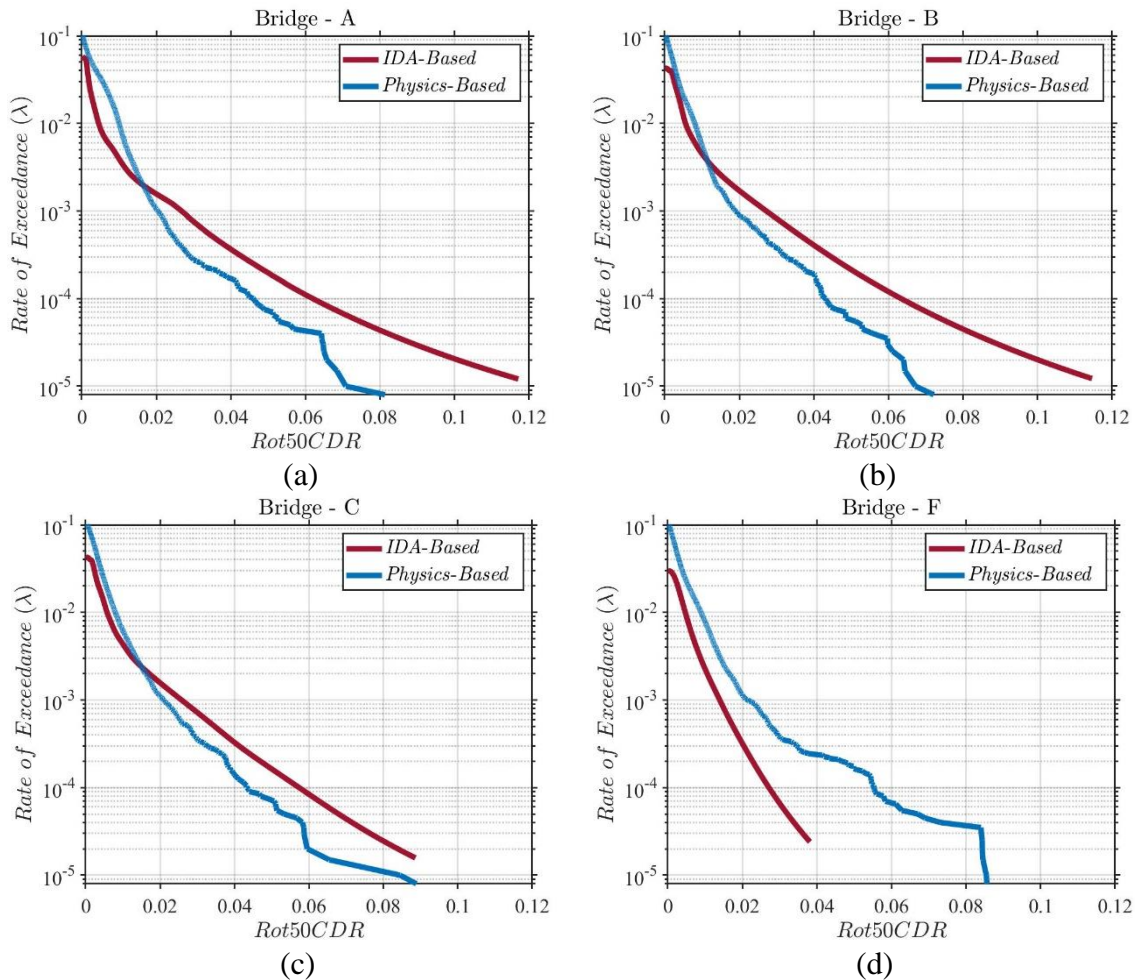## 6.4 Comparison of *Physics-Based* and EDP Hazard Curves

Figure 6.3 – Comparison of *Physics-Based* and *IDA-Based* hazard curves at LADT site for Bridges: (a) A, (b) B, (c) C, and (d) F

The developed *Physics-Based* EDP hazard curves are compared against the hazard curves developed using IDA in the Chapter 4, for all four bridges and five sites. An example of the comparison is

provided in Figure 6.3 which compares the *Physics-Based* EDP hazard curves against the *IDA-Based* EDP hazard curves for the four bridge structures for LADT site. Similar plots are developed for visual comparison of the EDP hazard curves. From these plots it is observed that, in general, for bridges with shorter periods (A, B, and C) that *IDA-Based* EDP hazard curves lie in the vicinity of *Physics-Based* EDP hazard curves for smaller return periods; whereas for longer return periods it consistently leads to higher values of *Rot50CDR* as compared to *Physics-Based* EDP hazard curves. For Bridge F, it is observed that *Physics-Based* EDP hazard curves leads to higher *Rot50CDR* as compared to *IDA-Based* EDP hazard curves for all hazard levels. Comparing Figure 6.1 with Figure 4.2, it can be observed that the *Physics-Based* EDP-IM tends to show a lower median response for all IM levels and a lower variability for large IM levels as compared to the *IDA-Based* results. Similar patterns were observed for other bridges and sites, except for SBSM and STNI, where the variability in EDP-IM data from *Physics-Based* data was significantly higher than *IDA-Based*. This can be attributed to the fact that the simulated ground motions of these sites include the effects of a deep basin and directivity pulses in terms of intensity, frequency, and duration characteristics. These characteristics can uniquely affect the response of the bridge structures, causing the variability in EDPs to increase. This peculiar feature is not incoported in the selection and scaling processes of the conventional methods of Incremental Dynamic Analysis (IDA). Also among the bridge responses, the reposne of Bridge F is obsereved to be different to the other three bridge structures. As mentioned earlier, CyberShake *15.12* simulations are hybrid of deterministic and stochastic simulations; the higher frequency content (>1 Hz) is simulated with stochastic approaches and is added to the *Physics-Based* deterministic estimates of the ground motion time series corresponding to lower frequency content (≤1 Hz). Hence, the ground motions affect the response of bridge structures differently for bridges with shorter periods (Bridges A, B, and C) as compared to bridges with longer periods (Bridge F) in contrast to the recorded ground motions. Since the recorded ground motions have not been selected to explicitly account for the basin and directivity effects, the *IDA-Based* results tend to be similar for Bridges A, B and C, and F. This is consistent with the findings of Bijelić *et al*. (2019), who used building structures to compare simulation-based EDP hazard curves with conventional methods. Also the as explained in the Fayaz *et al.* (2020c), noticeable differences are obserevd in the IM hazard curves of CyberShake *15.12* study and CB14. The integration process of *IDA-Based* EDP hazard curves futher accumlates these diffferences leading to higher diffference from the Phyics-Based EDP

hazard curves. Fayaz *et al.* (2021a) futher showed that the primary differences between the two types of EDP hazard curves are due to the differences between their EDP-IM data and recommended that the engineering community should validate simulated ground motions not only based on IM levels but also based on more in-depth comparisons made on the EDP-IM level.
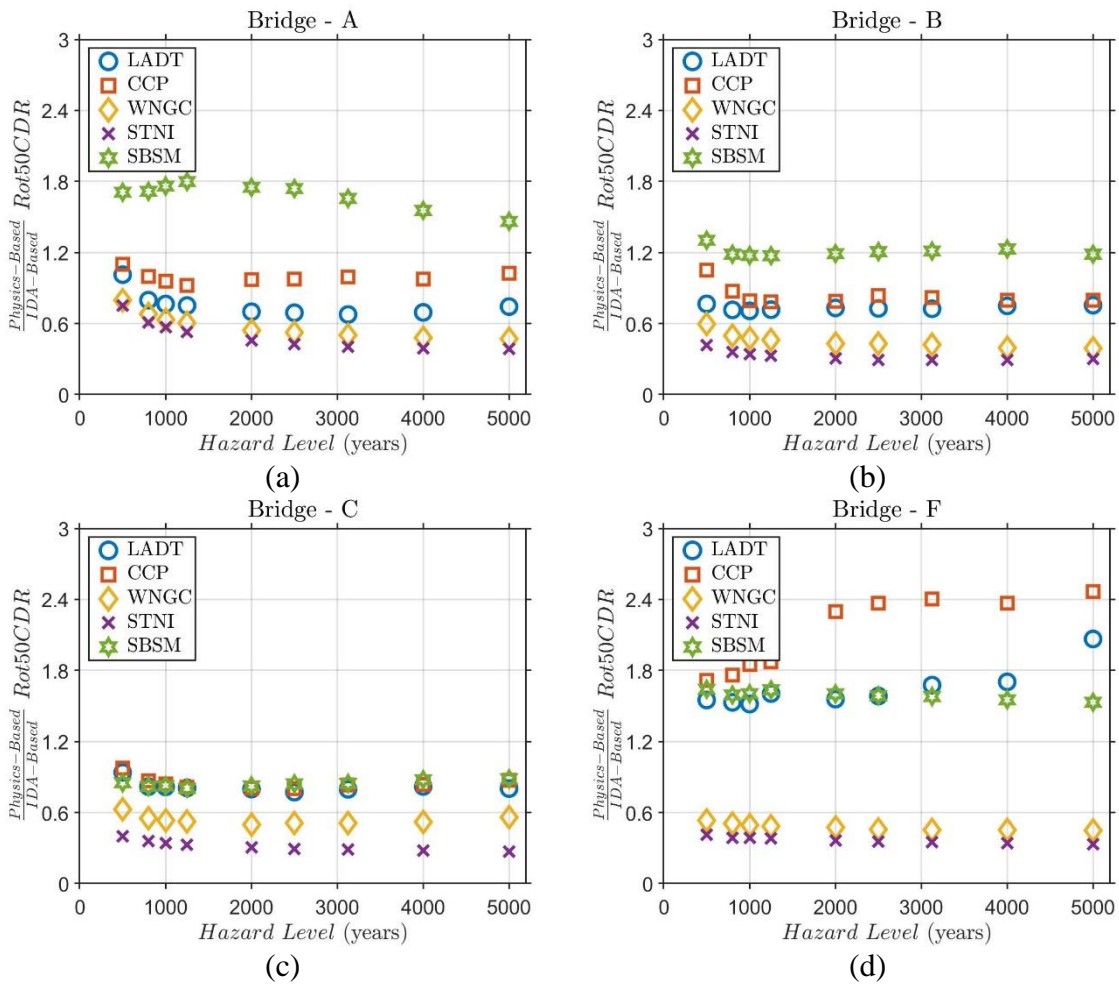

Figure 6.4 – Ratios between *Physics-Based* and *IDA-Based* hazard curves for Bridges: (a) A, (b) B, (c) C, and (d) F

To further understand the reasons for, and parameters that effect, ratios between EDPs obtained from *Physics-Based* and *IDA-Based* hazard curves are computed for various hazard levels. The ratios are used to construct functional forms to facilitate the conversion of EDPs from one form of representing ground motion hazard to another given the hazard level (in years), and site characteristics V_{s30} (in m/s) and $Z_{2.5}$ (in m). Figure 6.4 presents the ratios of *Physics-Based* and *IDA-Based* EDPs for various hazard levels. As a general observation, it can be seen that the ratios for Bridges A, B*,* and C tend to vary slightly with the return period, whereas for Bridge F, an

increasing trend is observed. The ratios are observed to be between 0.6 to 1 for Bridges A, B, and C, and mainly greater than 1.5 for Bridge F for most of the sites. This means, in general, there is a ~40% reduction in *Rot50CDR* arising from *Physics-Based* analysis of CyberShake *15.12* simulations as compared to IDA using recorded ground motions for short period bridges. For the bridge with a longer period, Bridge F, there is a ~60% to ~70% increase in *Rot50CDR* arising from *Physics-Based* analysis of CyberShake *15.12* simulations compared to IDA using recorded ground motions. The site-to-site variability in these ratios are observed to be higher than the ones observed in Chapter 5 (using *Site-Based* ground motions). This is due to the fact that the *Physics-Based* ground motions are able to more accurately model the 3-D surface of the basin and include the site and directivity effects for the propogation of ground motion waveforms.

To assist with improving ground motion simulation methods using this research (or similar ones where probabilistic seismic demand analysis is used for validation of a ground motion simulation method), functional forms for the difference in the response of each type of bridge (A, B, C, and F) in terms of return period and site parameters are provided. This is done by combining the results from all sites and conducting mixed-effects regression analysis with the target variable being the ratio of *Rot50CDR*s between *Physics-Based* EDP hazard curves and *IDA-Based* EDP hazard curves. The predictor variables include Return Period (*RP*), site Shear-Wave Velocity (V_{s30}) and site Basin-Depth ($Z_{2.5}$). This is expressed in Equation 6.1 and is fitted to data independently for each bridge. In Equation 5.1, ε_{ij} represents the within-site variability for the ith hazard level and j^{th} site with zero mean and variance of $\phi_b{}^2$, and σ_i represents the between-site variability for the ith hazard level with zero mean and variance of $\tau_b{}^2$. The coefficients and the goodness-of-fit measure (R^2_{adj}) for each bridge are given in Table 6.2.

$$\ln\left(\frac{\text{PhysicsBased } Rot50CDR}{\text{IDABased } Rot50CDR}\right) = b_0 + b_1(\ln(RP)) + b_2(\ln(V_{s30})) + b_3(\ln(Z_{2.5})) + \varepsilon_{ij}(0, \phi_b^2) + \sigma_i(0, \tau_b^2) \qquad (6.1)$$

It can be observed from Table 6.2 that for all four bridges the value of ϕ is higher than τ, which means that the site to site variability is highly explanatory in the computation of the EDP ratios Hence, the use of mixed-effects regression is appropriate for describing the EDP ratios. Also, goodness-of-fit measure R^2_{adj} is observed to be consistently above 0.95, which means that the regression equations can estimate the EDP ratios for the four bridge structures with a high level of accuracy. The coefficient b_1 is observed to be negative for all Bridges A, B, and C, demonstrating

that as the return period increases, *IDA-Based* EDPs tends to be larger than *Physics-Based* EDPs; while the case is opposite for Bridge F. This is due to the higher variability in EDPs caused by recorded ground motions escpecially for larger IM levels. Due to the higher variability in EDPs in EDP-IM data, the integration process accumulates this variability causing the EDP hazard curve *IDA-Based* to grow faster as compared to *Physics-Based*.The value of b_2 is positive for all four bridges, with the highest value being noted for Bridge F. This means with an increase in V_{s30} of the site, *Physics-Based* EDPs tends to increase as compared to *IDA-Based* EDPs. This effect is observed to be highly dominant for Bridge F that has a relatively longer period than other OSBs. Furthermore, the coefficient b_3 is negative for shorter period Bridges A, B and C, while it is significantly positive for Bridge F. This trend shows that with an increase in basin depth $Z_{2.5}$, *Physics-Based* tends to be lower than *IDA-Based* for shorter period bridges (A, B and C). However, for a long period Bridge F, it is observed that a deeper basin tends to increase the EDP ratio and leads to higher *Physics-Based* as compared to *IDA-Based*. This is postulated to be due the fact that the simulated ground motions, specifically for the sites with significant basin effects (SBSM and STNI), possess high basin and directivity effects, which alters the lower frequency content of the ground motions. This alteration in lower frequencies tends to increase the variability in EDP-IM of Bridge F (possessing longer natural period), which leads to an increase in the *Physics-Based* EDP hazard curve. However, the impact of these basin effects is not observed to be highly prominent in the EDP-IM relationship for shorter period Bridges A, B, and C. The stochastic part (frequency content > 1 Hz) of the CyberShake simulated ground motions is modeled using a plane-layer velocity structure; hence the basin effects are only captured approximately via the V_{s30} amplification factors. These features may be responsible for significantly different behavior observed for the regressions of Bridges A, B, and C versus Bridge F. In general, for the bridge analysis, it is concluded that there is a need for detailed site-specific analysis because basin effects and site amplification can lead to significant changes in the response of bridge structures. The relations provided in Table 6.2 can assist engineers in validating their methods of bridge analysis and can be used in scaling the EDPs as per the design site. It should be noted that the regression equations proposed in this study are based on the Cybershake *15.12* simulations of five southern California sites with soft soils; hence they can be biased towards these conditions. However, these equations can provide initial estimates of the scaling factors and can be easily updated with more data from different site conditions.

Table 6.2 – Fitted coefficients of the *Physics-Based* mixed-effects regressions

Bridge	b_0	b_1	b_2	b_3	ϕ_b	τ_b	R^2_{adj}
A	-11.169	-0.142	2.124	-0.269	0.261	0.094	0.96
B	-15.781	-0.076	2.799	-0.291	0.162	0.069	0.97
C	-14.371	-0.055	2.516	-0.241	0.139	0.063	0.97
F	-40.165	0.017	6.844	0.427	0.207	0.093	0.98

6.5 Conclusions

This chapter demonstrates a framework for utilizing *Physics-Based* (in particular CyberShake *15.12* study) simulated ground motions for development of EDP hazard curves. Catalogs of simulated ground motions representing a time-span of 200,000 years for five sites in Southern California with a diverse site and local seismicity conditions are obtained from CyberShake (i.e., 20,000 ground motions from events with $M_w > 6.0$ occurring within 200 km of each site) and applied to four OSB structures. NLTHA is conducted to calculate the bridge column drift ratio (CDR); the effect of uncertainty in the ground motion incident angle is incorporated by rotating the applied time-series at 10° increments. The results show that CyberShake (ver. *15.12*) simulation study is relatively acceptable for seismic performance assessment of OSBs with short periods. For the OSBs with long period ($T = 1.1$ sec) EDPs obtained from *Physics-Based* analysis tend to be higher than the EDPs obtained from utilizing recorded ground motions and performing IDA. Further observations suggest that the difference between EDP-IM data is the primary source of EDP hazard curve variations. It is recommended that validation efforts should go beyond comparisons of IM levels and also include EDP-IM level validation. Furthermore, to account for the site-specific differences in the EDP hazard curves obtained from *Physics-Based* analysis, predictive equations are provided using regression. The proposed equations relate the ratio between the EDPs obtained from *Physics-Based* analysis and *IDA-Based* analysis with the return period, and site characteristics V_{s30} and $Z_{2.5}$. These relations can be used by the engineers to scale the EDPs obtained from conventional IDA methods to the EDPs obtained using NLTHA under a large set of CyberShake simulated ground motions. Lastly, further analysis is recommended to compare the EDP-IM obtained from CyberShake with those of recorded motions, especially for long period and transition period (between stochastic and deterministic) bridge structures. Note

that the results and analysis presented in this study only pertain to CyberShake *15.12* study and other CyberShake studies coupled with different selection procedures for recorded ground motions may lead to different conclusions. However, this framework can be easily extended to other simulation studies and other types of structures.

Chapter Citation:

1) Fayaz J., Azar S., Dabaghi M., and Zareian F. (2020c). Methodology for Validation of Simulated Ground Motions for Seismic Response Assessment: Application to CyberShake Source-Based Ground Motions. *Bulletin of the Seismological Society of America*. Vol. 111, No. 1, Pages 226-241.

2) Fayaz J., Rezaeian S., and Zareian F. (2021a). Evaluation of simulated ground motions using probabilistic seismic demand analysis: CyberShake (ver. 15.12) simulations for Ordinary Standard Bridges. *Soil Dynamics and Earthquake Engineering*. Volume 141, 106533.

REFERENCES

1) AASHTO (2011). AASHTO Guide Specifications for LRFD Seismic Bridge Design, American Association of State Highway and Transportation Officials, 2nd Edition with 2014 Interim Revisions.

2) Abrahamson N. (2006). Seismic hazard assessment: Problems with current practice and future developments. *First European Conference on Earthquake Engineering and Seismology*, Geneva, Switzerland.

3) Abrahamson N. A., and Youngs R. R., (1992). A stable algorithm for regression analyses using the random effects model. *Bull. Seismol. Soc. Am.* 82, 505–510.

4) Abrahamson N. A., and Bommer J. J. (2005). Probability and uncertainty in seismic hazard analysis. *Earthquake Spectra*, 21(2), 603–607.

5) Abrahamson N. A., Kuehn N. M., Walling M., and Landwehr N. (2019). Probabilistic seismic hazard analysis in California using nonergodic ground motion models. *Bulletin of the Seismological Society of America*, 109(4), 1235–1249.

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

7) Afshari K., and Stewart J.P. (2014). Physically parameterized prediction equations for significant duration in active crustal regions. *Earthquake Spectra*, 32(4):2057-2081.

8) Azar, S., Dabaghi, M., and Rezaeian, S., (2019). Probabilistic seismic hazard analysis using stochastic simulated ground motions. Conference Proceedings, ICASP13, South Korea

9) Baker W. (2011). Conditional Mean Spectrum: Tool for Ground-Motion Selection. *Journal of Structural Engineering*, 137(3): 322-331.

10) Bazzurro, P., and N. Luco (2004). Parameterization of non-stationary acceleration time histories, *PEER Report*, University of California, Berkeley, California.

11) Bijelić, N., Lin T., and Deierlein G.G. (2019). Evaluation of Building Collapse Risk and Drift Demands by Nonlinear Structural Analyses Using Conventional Hazard Analysis versus Direct Simulation with Cybershake Seismograms. *Bulletin of the Seismological Society of America*, 109(5): 1812–1828

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

13) Bozorgnia Y., Hachem M., Campbell K.W. (2010). Ground Motion Prediction Equation ("Attenuation Relationship") for Inelastic Response Spectra. *Earthquake Spectra*, Volume 26, No. 1.

14) Bozorgzadeh A., Megally S., Restrepo J., and Ashford S. A. (2006). Capacity evaluation of exterior sacrificial shear keys of bridge abutments, *Bridge Engineering* 555–565

15) Brown C. D. and Davis H. T. (2006). Receiver operating characteristic curves and related decision measures: a tutorial. *Chemometrics and Intelligent Laboratory Systems.* 80: 24–38.

16) Caltrans (2013). Memo to Designers. California Dept. of Transportation, Sacramento, CA.

17) Caltrans (2013). Seismic Design Criteria. Version 1.7, California Department of Transportation.

18) Caltrans (2019). Seismic Design Criteria. Version 2.0, California Department of Transportation.

19) Campbell K.W., and Bozorgnia, Y. (2014). Campbell-Bozorgnia NGA-West2 horizontal ground motion model for active tectonic domains. *Earthquake Spectra*, 30(3), 1087-1115.

20) Campbell K.W., and Bozorgnia Y. (2019). Ground motion models for the horizontal components of Arias intensity (AI) and cumulative absolute velocity (CAV) using the NGA-West2 Database. *Earthquake Spectra*, 35(3):1289-1310.

21) Chiou B.S., and Youngs R.R. (2014). Update of the Chious and Youngs NGA Model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, 30(3):1117-1153.

22) Choi E. (2002). Seismic Analysis and Retrofit of Mid-America Bridges. *Ph.D. Dissertation*, Department of Civil and Environmental Engineering, Georgia Institute of Technology: Atlanta, GA, 381 pp.

23) Dabaghi M., and Der Kiureghian A., (2018). Simulation of orthogonal horizontal components of near-fault ground motion for specified earthquake source and site characteristics. *Earthquake Engineering and Structural Dynamics*, 47, 1369–1393.

24) Dabaghi M., Der Kiureghian A., Rezaeian S., and Luco N., (2013). Seismic hazard analysis using simulated ground motions. *11th International Conference on Structural Safety and Reliability* (ICOSSAR), Columbia University New York, NY.

25) Dhanya J., and Raghukanth S.T.G. (2017). Ground motion prediction model using artificial neural network. *Pure Appl Geophys*, 175:1035-1064

26) Douglas J (2019). Ground Motion Prediction Equations 1964-2019. http://www.GMPM.org.uk/GMPMreport2014.pdf. Accessed December 19th, 2019.

27) Douglas J., and Aochi H. (2008). A survey of techniques for predicting earthquake ground motions for engineering purposes. *Surv. Geophys.* 29, no. 3, 187–220.

28) Dutta A. and Mander J.B., (1998). Seismic Fragility Analysis of Highway Bridges. *INCEDE-MCEER Center-to-Center Workshop on Earthquake Engineering Frontiers in Transportation Systems*.

29) Du W., and Wang G. (2013). Intra-event spatial correlations for cumulative absolute velocity, Arias intensity, and spectral accelerations based on regional site conditions. *Bulletin of the Seismological Society of America* 103(2A), 1117-1129.

30) Du W., and Wang G. (2018). Ground motion selection for seismic slope displacement analysis using a generalized intensity measure distribution method. *Earthquake Engineering and Structural Dynamics*, 47(5), 1352-1359.

31) Duncan J.M., and Mokwa R.L. (2001). Passive earth pressures: theories and tests. *Journal of Geotechnical and Geoenvironmental Engineering*, 127(3), 248-257.

32) Earth Mechanics, INC. (2005). Field investigation report for abutment backfill characterization. *Report No. SSRP-05/02*, University of California, San Diego, CA.

33) Eads L., Miranda E., and Lignos D. G. (2015). Average spectral acceleration as an intensity measure for collapse risk assessment. *Earthquake Engineering Structural Dynamics*, 44(12): 2057-2073.

34) Eads L., Miranda E., and Lignos D. (2016). Spectral shape metrics and structural collapse potential. *Earthquake Engineering and Structural Dynamics*, 45, 1643-1659.

35) Ebrahimian H., and Jalayer F. (2020). Selection of seismic intensity measures for prescribed limit states using alternative nonlinear dynamic analysis methods. *Earthquake Engineering & Structural Dynamics*.

36) Fang H.Y. (1999). "Foundation Engineering Handbook," Second Edition, Van Nostrand Reinhold, New York.

37) Fayaz, J. and Zareian, F (2019). Reliability Analysis of Steel SMRF and SCBF Structures Considering the Vertical Component of Near-Fault Ground Motions. *ASCE- Journal of St. Eng.*, Vol. 145, Issue 7

38) Fayaz J., Riquelme M., and Zareian F. (2020a). Sensitivity of The Response of Box-Girder Seat-Type Bridges to the Duration of Ground Motions arising from Crustal and Subduction Earthquakes. *Engineering Structures*, Vol. 219, 110845.

39) Fayaz J., Dabaghi M., and Zareian F. (2020b). Utilization of Site-Based Simulated Ground Motions for Hazard-Targeted Seismic Demand Estimation: application for Ordinary Bridges in Southern California. *Journal of Bridge Engineering*, Vol. 25, Issue 11.

40) Fayaz J., Azar S., Dabaghi M., and Zareian F. (2020c). Methodology for Validation of Simulated Ground Motions for Seismic Response Assessment: Application to Cybershake Source-Based Ground Motions. *Bulletin of the Seismological Society of America*. Vol. 111, No. 1.

41) Fayaz J., Xiang Y., and Zareian F. (2020d). Generalized Ground Motion Prediction Model (GGMPM) Using Hybrid Neural Networks. *Earthquake Engineering and Structural Dynamics*. Volume 50, Issue 6.

42) Fayaz J., Rezaeian S., and Zareian F. (2021a). Evaluation of simulated ground motions using probabilistic seismic demand analysis: CyberShake (ver. 15.12) simulations for Ordinary Standard Bridges. *Soil Dynamics and Earthquake Engineering*. Volume 141, 106533.

43) Fayaz J., Azar S., Dabaghi M., and Zareian F. (2021b). An Efficient Algorithm to Simulate Hazard-Targeted Site-Based Synthetic Ground Motions. *Earthquake Spectra*. Vol 37, Issue 2.

44) Fayaz J., and Zareian F. (2021c). An Efficient Algorithm to Simulate Site-Based Ground Motions that match a Target Spectrum. *Earthquake Engineering and Structural Dynamics*. (2nd round of review).

45) Field E. H., Dawson T. E., Felzer K. R., Frankel A. D., Gupta V., Jordan T. H., Parsons T., Petersen M. D., Stein R. S., Weldon II R. J., and Wills C. J. (2009). Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2). *Bulletin of the Seismological Society of America*, 99(4): 2053-2107.

46) Field E.H, Arrowsmith R.J., Biasi G.P., Bird P., Dawson T.E., Felzer K.R., Jackson D.D., Johnson K.M., Jordan T.H., Madden C., Michael A.J., Milner K.R., Page M.T., Parsons T., Powers P.M., Shaw B.E., Thatcher W.R., Weldon R.J. II, Zeng Y. (2014). Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3) – The Time-Independent Model. *Bulletin of the Seismological Society of America*, 104 (3): 1122–1180.

47) Field E. H., Jordan T. H., and Cornell C. A. (2003). OpenSHA: A Developing Community-Modeling Environment for Seismic Hazard Analysis. *Seismological Research Letters*, 74(4): 406-419.

48) Garren S. T. (1998). Maximum likelihood estimation of the correlation coefficient in a bivariate normal model with missing data. *Statistics & Probability Letters*. 38 (3): 281–288

49) Gerstenberger M. C., Marzocchi W., Allen T., et al. (2020), Probabilistic Seismic Hazard Analysis at Regional and National Scales: State of the Art and Future Challenges. *Reviews of Geophysics*, Volume 58, Issue 2.

50) Graves R., Jordan T. H., Callaghan S., Deelman E., Field E., Juve G., Kesselman C., Maechling P., Mehta G., Milner K., Okaya D., Small P., and Vahi K. (2011). Cybershake: A Physics-Based Seismic Hazard Model for Southern California. *Pure and Applied Geophysics*, 168(3): 367-381.

51) Graves R., and Pitarka A. (2015). Refinements to the Graves and Pitarka (2010) Broadband Ground-Motion Simulation Method. *Seismological Research Letters*, 86(1): 75-80.

52) Hansen N. (2006). The CMA evolution strategy: a comparing review. *Towards a new evolutionary computation. Advances on estimation of distribution algorithms.* IcoS Institute of Computational Science, ETH, Zürich

53) Hochreiter S., and Schmidhuber J. (1996). LSTM can solve hard long-time lag problems. *Proceedings of the 9th International Conference on Neural Information Precossing Systems*, December 1996, 473-479

54) Hosmer D. W., and Lemeshow S. (2000). Applied Logistic Regression, 2nd ed. Wiley. ISBN 978-0-471-35632-5

55) Huang, C, Galasso, C. (2019) Ground-motion intensity measure correlations observed in Italian strong-motion records. *Earthquake Engng Struct Dyn*, 48, 1634– 1660.

56) Huang, C, Tarbali, K, Galasso, C. (2020) Correlation properties of integral ground-motion intensity measures from Italian strong-motion records. *Earthquake Engng Struct Dyn*, 1– 18.

57) Jalayer F., Beck J.L., and Zareian F. (2012). Analyzing the Sufficiency of Alternative Scalar and Vector Intensity Measures of Ground Shaking Based on Information Theory, *Journal of Engineering Mechanics*, Vol. 138, No. 3.

58) Kaviani P., Zareian F. and Taciroglu E. (2012). Seismic Behavior of Reinforced Concrete Bridges with Skew-Angled Seat-Type Abutments. *Engineering Structures*, 45, 137–150.

59) Khalili-Tehrani P., Taciroglu E., and Shamsabadi A. (2010) Backbone curves for passive lateral response of walls with homogeneous backfills. *Proc. of 2009 Soil-Foundation-Structure Interaction. Workshop (Orense, Chouw, and Pender, eds.)*, University of Auckland, New Zealand, 2, 149-154, 2010

60) Kiani J., and Pezeshk S. (2017) Sensitivity analysis of the seismic demands of RC moment resisting frames to different aspects of ground motions. *Earthq Eng Struct Dyn*, 46(15):2739-2755.

61) Kingma D.P., and Ba J. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations*. San Diego, 2015.

62) Kottari A. (2016). Design and capacity assessment of external shear keys in bridge abutments. *Ph.D. Dissertation*, Department of Structural Engineering, University of California San Diego: San Diego, CA.

63) Kotsoglou A., and Pantazopoulou S. (2010). Response simulation and seismic assessment of highway overcrossing, *Earthquake Engineering and Sructural Dynamics*, **3**(9).

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

65) Lemaître G., Nogueira F., and Aridas C.K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research, Volume: 18, Issue: 1.

66) Lin T., and Baker J.W. (2015). Conditional Spectra. In: Beer M., Kougioumtzoglou I., Patelli E., Au IK. (eds) *Encyclopedia of Earthquake Engineering*, 461-472. Springer, Berlin, Heidelberg.

67) Liu J.M., Wang T., Wu S.R., and Gao M.T. (2016). New empirical relationships between Arias intensity and peak ground acceleration. *Bulletin of Seismological Society of America,* 106(5), 2168-2176.

68) Luco N., and Cornell C.A. (2007). Structure-Specific Scalar Intensity Measures for Near-Source and Ordinary Earthquake Ground Motions. *Earthquake Spectra*, Vol 23, Issue 2.

69) Luco, N. (2002). Probabilistic seismic demand analysis, SMRF connection fractures, and near-source effects, *Ph.D. Thesis*, Department of Civil and Environmental Engineering, Stanford University, CA.

70) Mackie K.R., and Stojadinovic B. (2003). Seismic Demands for Performance-Based Design of Bridges. *Report No. 2003/16, PEER*, Berkeley CA.

71) McKenna F., Scott M. H., and Fenves G. L. (2010). Nonlinear finite element analysis software architecture using object composition. *Journal of Computing in Civil Engineering*, 24(1): 95-107.

72) Omrani R., Mobasher B., Sheikhakbari S., Zareian F., and Taciroglu E. (2017). Variability in the predicted seismic performance of a typical seat type California bridge due to epistemic uncertainties in its abutment backfill and shear-key models. *Engineering Structures* 148 718–738.

73) Petersen, M. D., Frankel A. D., Harmsen S. C., Mueller C. S., Haller K. M., Wheeler R. L., and Perkins D. M., (2008). Doc. for the 2008 update of the united states national seismic hazard maps, *2331-1258*

74) Prechelt, L. (2002). Early Stopping – But When? *Neural Networks: Tricks of the Trade*. University Karlsruhe

75) Ramanathan, K. (2012). Next generation seismic fragility curves for California bridges incorporating the evolution in Seismic design philosophy. *Ph.D. Dissertation*, Department of Civil and Environmental Engineering, Georgia Institute of Technology: Atlanta, GA, 331 pp.

76) Rezaeian S., and Der Kiureghian A., (2012). Simulation of orthogonal horizontal ground motion components for specified earthquake and site characteristics. *Earthquake Engineering and Structural Dynamics,* 41(2), 335-353.

77) Rodríguez-Castellanos A., Ruiz S.E., Bojórquez E., and Reyes-Salazar A. (2020) Influence of spectral acceleration correlation models on conditional mean spectra and probabilistic seismic hazard analysis. *Earthquake Engng Struct Dyn*, 1– 20.

78) Scharge I. (1981). Anchoring of Bearing by Friction. *Publication SP- American Concrete Institute*, p197-215

79) Shamsabadi A., Ashour M., and Norris G. (2005). Bridge Abutment Nonlinear Force-Displacement Capacity Prediction for Seismic Design. *Geotechnical and Geoenvironmental Engineering*. Vol. 131, No. 2, pp.1-9, 2005

80) Shamsabadi A. and Kapuskar M. (2006). Nonlinear seismic soil-abutment-structure interaction analysis of skewed bridges. *Proceedings of the 5th National Seismic Conference on Bridges and Highways*, San Francisco, CA, 12 pp.

81) Shamsabadi A., Rollins K.M., and Kapuskar M. (2007). Nonlinear soil–abutment–bridge structure interaction for seismic performance-based design. *Journal of Geotechnical and Geoenvironmental Engineering*, 133(6): 707-720.

82) Shamsabadi A., Khalili-Tehrani P., Stewart J.P., and Taciroglu E. (2010). Validated simulation models for lateral response of bridge abutments with typical backfills. *Journal of Bridge Engineering*, 15(3), 302-311.

83) Shahi S.K., and Baker J.W. (2014). An efficient algorithm to identify strong velocity pulses in multi-component ground motions. *Bulletin of the Seismological Society of America*, 104(5), 2456–2466

84) Srivastava M.S., and Yanagihara H. (2010). Testing the equality of several covariance matrices with fewer observations than dimension. *Journal of Multivariate Analysis*. 101(6), 1319-1329.

85) Tezcan J, and Cheng Q. (2012). Support vector regression for estimating earthquake response spectra. *Bull Earthq Eng*, 10(4):1205-1219.

86) Tothong P., and Cornell C.A. (2006). An Empirical Ground-Motion Attenuation Relation for Inelastic Spectral Displacement. *Bulletin of the Seismological Society of America*, Vol. 96, No. 6.

87) Timothy D. A., Robert B. D., Jonathan P. S., Emel S., Walter J. S., Brian S., Chiou 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): 989-1005.

88) U.S. Geological Survey (USGS) (2019) Unified Hazard Tool (dynamic conterminous US 2008 edition). Available at: https://earthquake.usgs.gov/hazards/interactive/index.php (accessed 3 April 2019)

89) Xu Y, Tang X. S., Wang J. P., and Kuo-Chen H. (2016). Copula-based joint probability function for PGA and CAV: A case\ study from Taiwan, *Earthquake Engineering and Structural Dynamics*, 45, 2123–2136.

90) Yeow T.Z., Orumiyehei A., Sullivan T.J., MacRae G.A., Clifton G.C., and Elwood K.J. (2018). Seismic performance of steel friction connections considering direct-repair costs. *Bulletin of Earthquake Engineering*, 16, 5963-5993.

91) Yoon Y. H., Ataya S., Mahan M., Malek A., Saiidi M. S., and Zokaie T. (2019). Probabilistic Damage Control Application: Implementation of Performance-Based Earthquake Engineering in Seismic Design of Highway Bridge Columns. *Journal of Bridge Engineering*, 24(7): 04019068.

92) Zhu R, Zeng D, and Kosorok MR. (2015). Reinforcement Learning Trees. *Journal of the American Statistical Association*. 110 (512): 1770–1784.
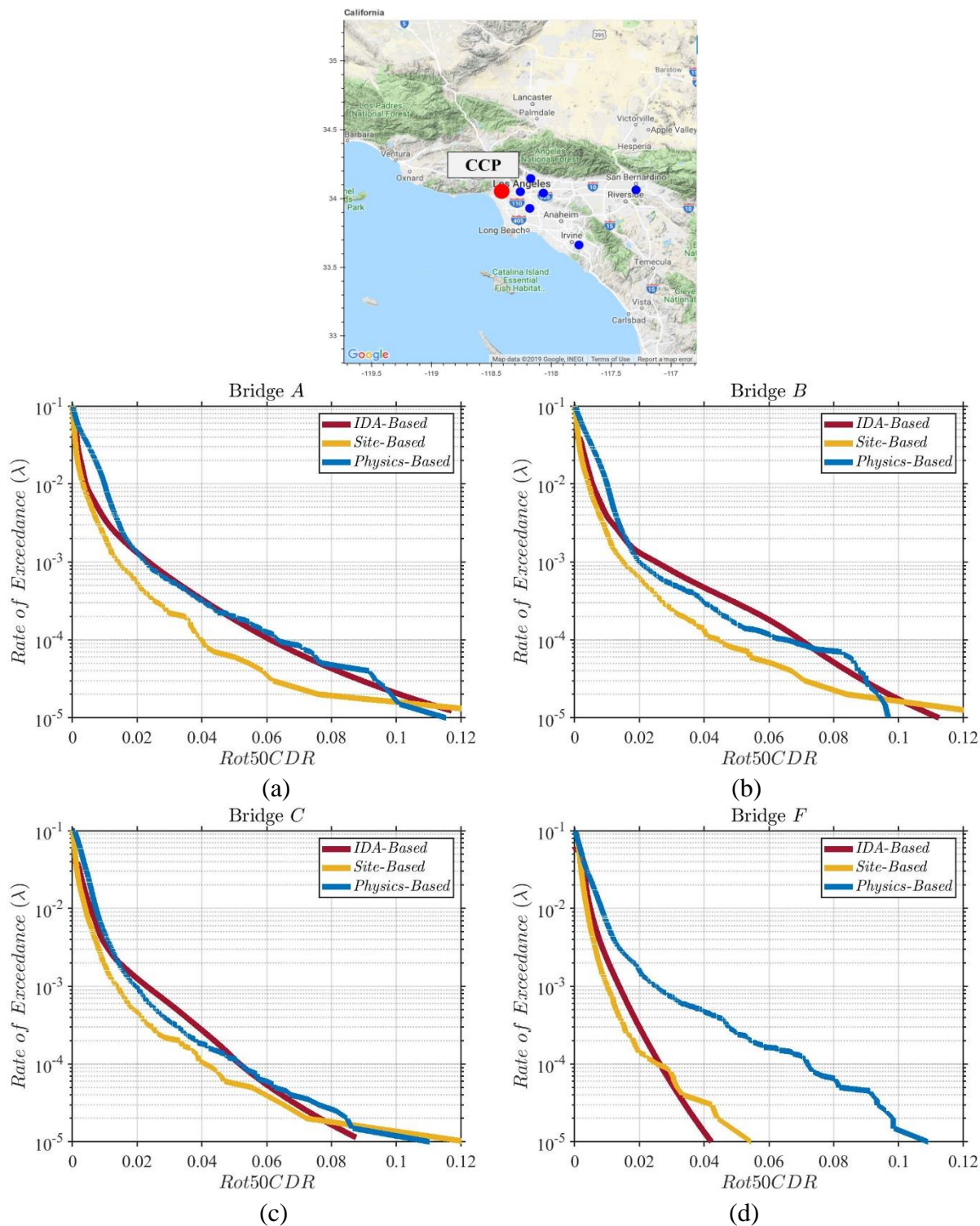
APPENDIX A

$Rot50CDR$ EDP HAZARD CURVES

Figure A.1 – Comparison of EDP hazard curves at CCP site for: (a) A, (b) B, (c) C, (d) F
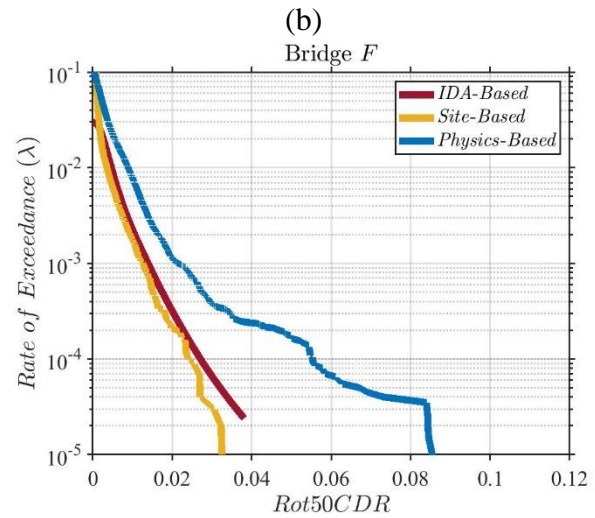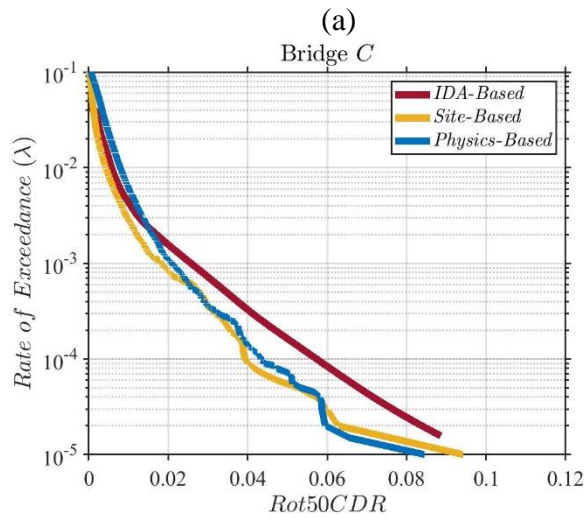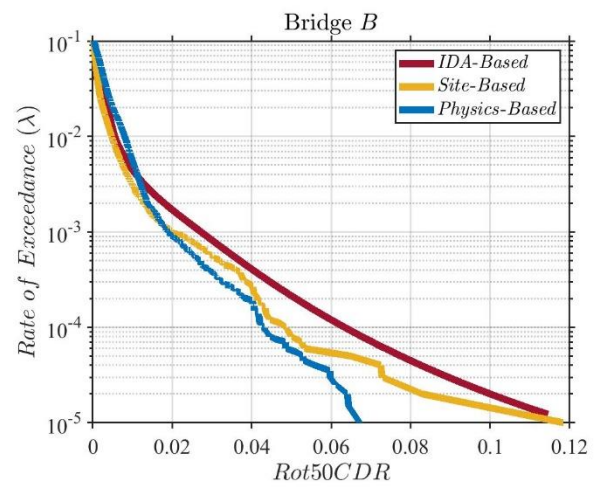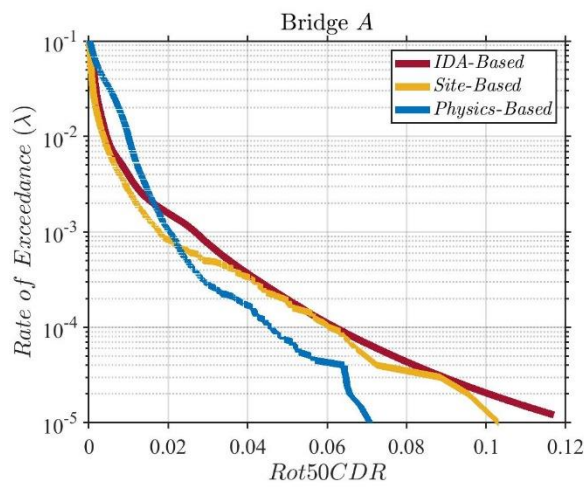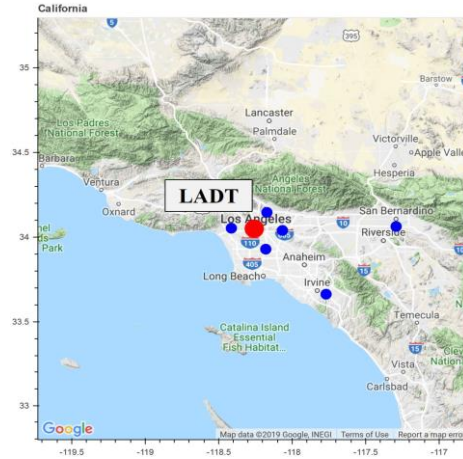
Figure A.2 – Comparison of EDP hazard curves at LADT site for: (a) A, (b) B, (c) C, (d) F
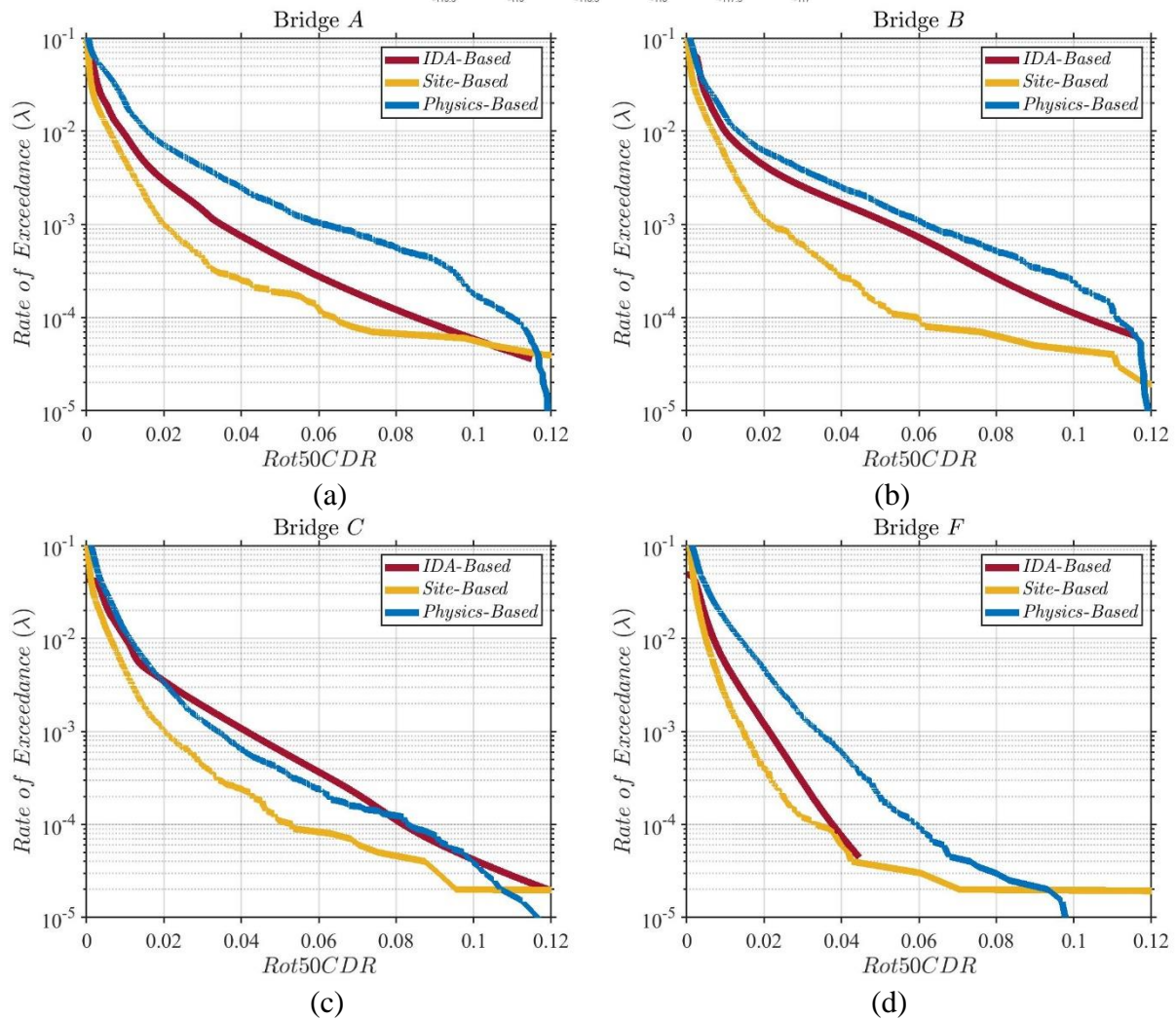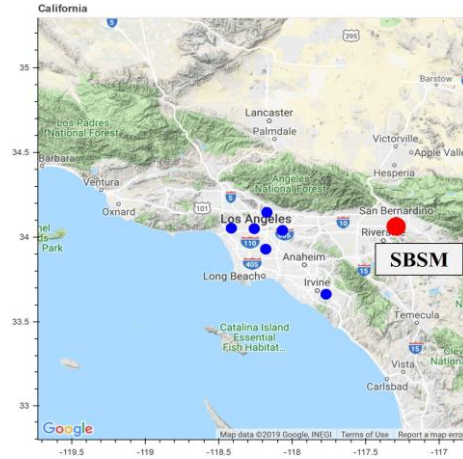
Figure A.3 – Comparison of EDP hazard curves at SBSM site for: (a) A, (b) B, (c) C, (d) F
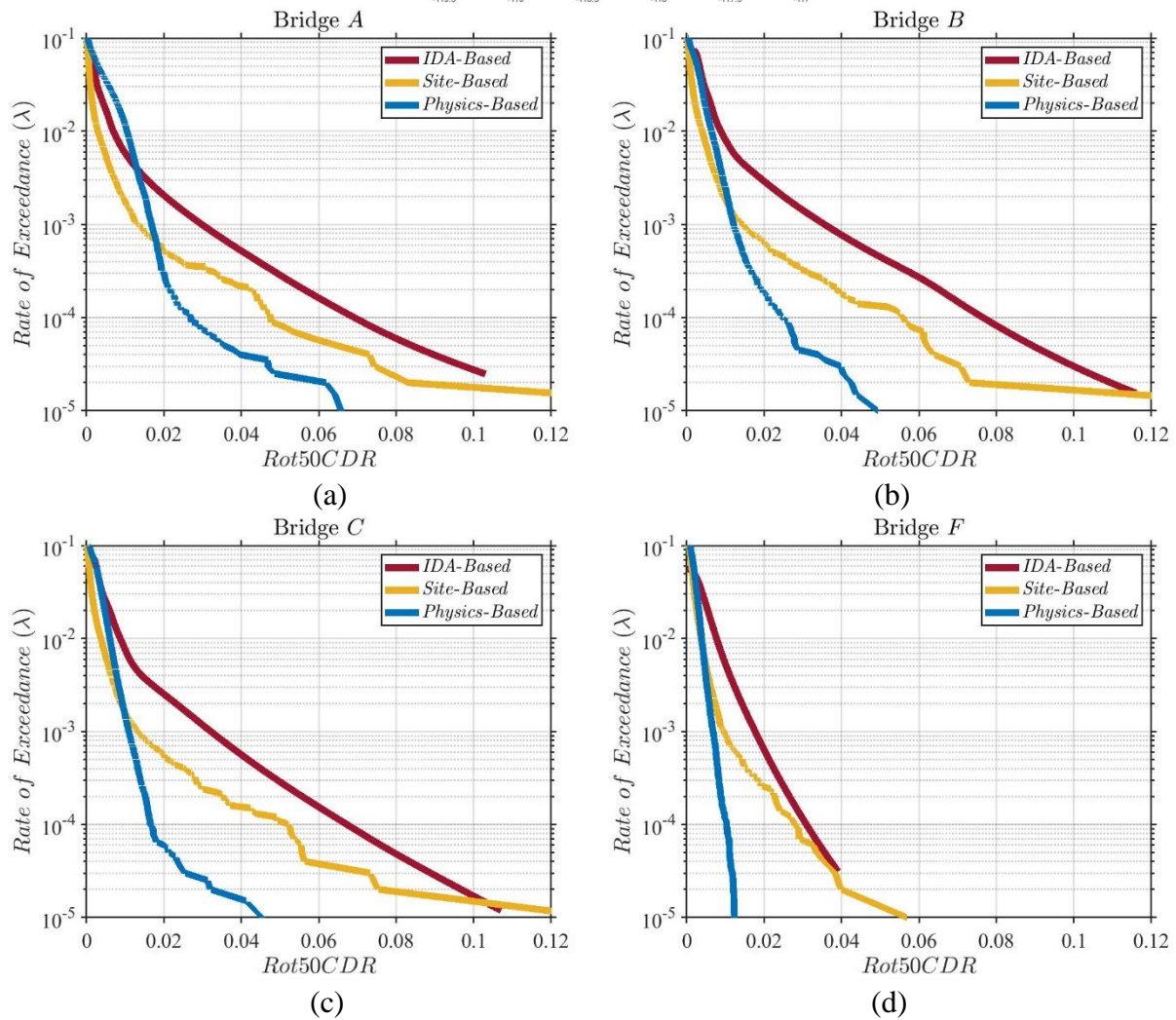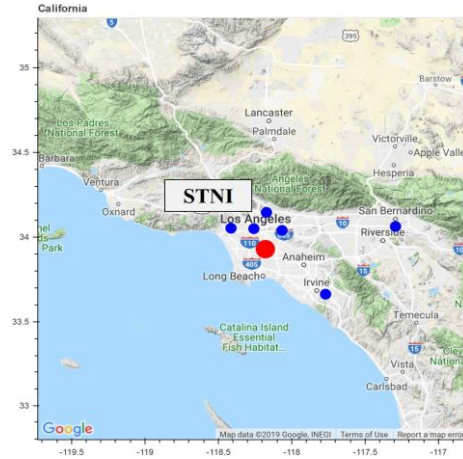
Figure A.4 – Comparison of EDP hazard curves at STNI site for: (a) A, (b) B, (c) C, (d) F
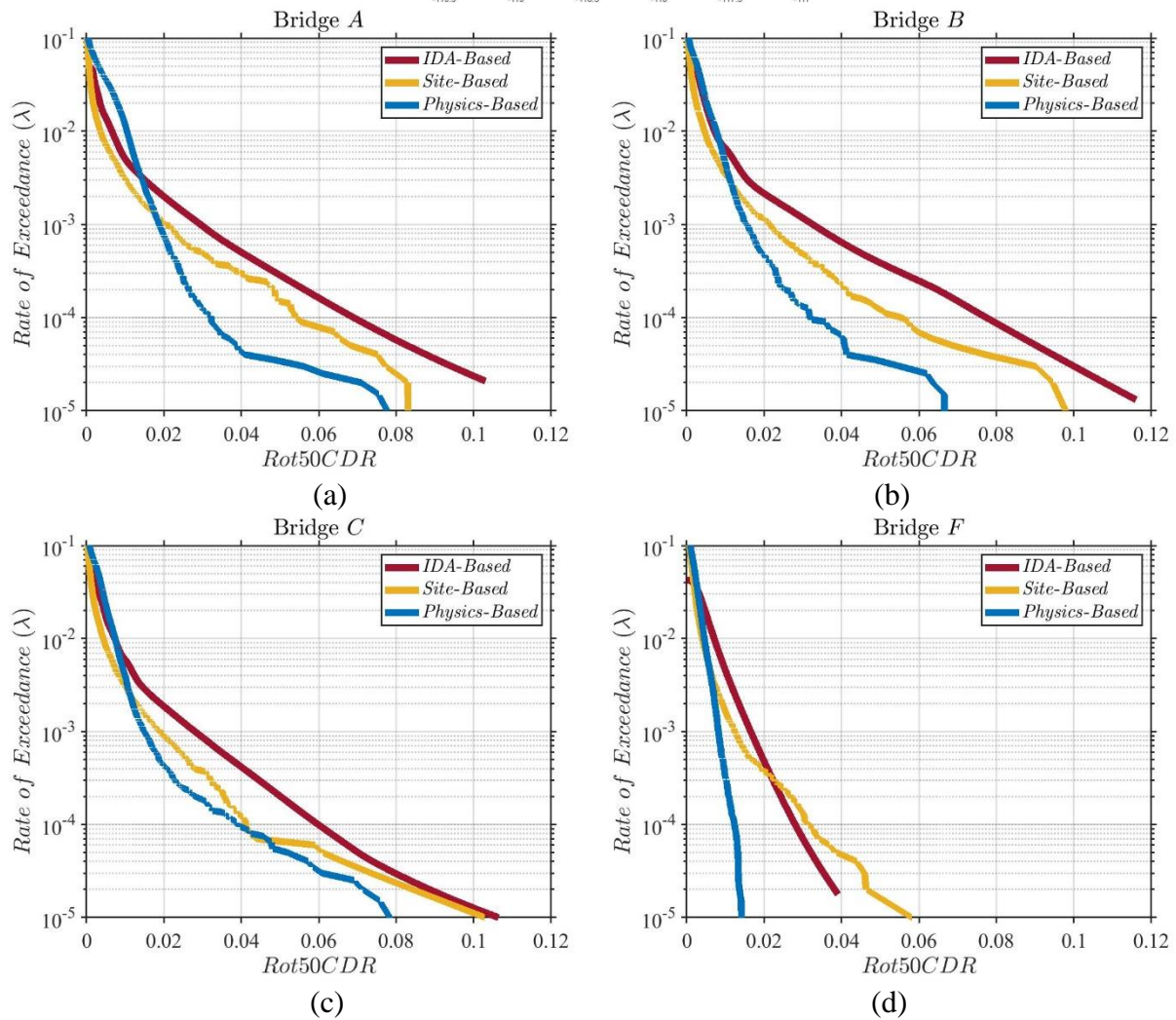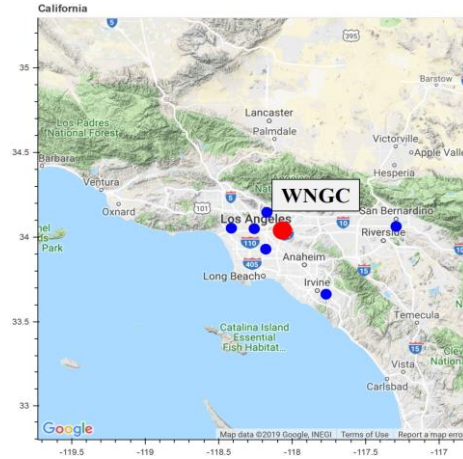
(a)

(b)

(c)

(d)

Figure A.5 – Comparison of EDP hazard curves at WNGC site for: (a) A, (b) B, (c) C, (d) F